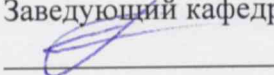


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДЕНО
Решением кафедры и допущено
«11» октября 2021 г., протокол № 3

Заведующий кафедрой
 А.А. Романов

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ОСНОВНОЙ ПРОФЕССИОНАЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ
ВЫСШЕГО ОБРАЗОВАНИЯ – ПРОГРАММЫ МАГИСТРАТУРЫ**

Направление подготовки
09.04.03 «Прикладная информатика»

Программа подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
Очная

Ульяновск 2021г.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

В.И. МЕНЬЩИКОВА

МЕТОДОЛОГИЯ НАУЧНОГО ПОЗНАНИЯ
методические рекомендации по выполнению
практических работ по дисциплине
«Методология научного познания»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

МЕНЬЩИКОВА В.И.

Методология научного познания : методические рекомендации по выполнению практических работ / В.И. Меньщикова. – Ульяновск : УлГТУ, 2021. – 34 с.

Методические рекомендации по выполнению практических работ адресованы студентам для выполнения и оформления практических работ по дисциплине «Методология научного познания». В данных методических рекомендациях предоставлены варианты заданий, рекомендации и требования к практическим работам. Методические рекомендации разработаны в соответствии с рабочей программой дисциплины «Методология научного познания». Методические рекомендации предназначены для студентов, обучающихся по направлению 09.04.03 «Прикладная информатика» профиль «Искусственный интеллект и бизнес-аналитика».

Работа подготовлена на кафедре «Информационные системы».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по практическим работам	6
Зачет	7
1. Базовые понятия теории и методологии научного исследования	9
Вопросы для собеседования	11
2. Система методов научных исследований в предметной области	12
Вопросы для собеседования	19
3. Система методов прикладных исследований в предметной области	20
Вопросы для собеседования	25
Практическая работа №1	26
Контрольные вопросы	27
Практическая работа №2	28
Контрольные вопросы	29
Практическая работа №3 и 4	30
Контрольные вопросы	32
Список используемой литературы	33

ВВЕДЕНИЕ

Целью освоения дисциплины (модуля) «Методология научного познания» является формирование у обучающихся комплексных знаний, умений и практических навыков в области организации научной деятельности.

Задачами освоения дисциплины (модуля) являются формирование у обучающихся:

- раскрыть базовые понятия теории и методологии научного исследования, систему методов научных и прикладных исследований, рассмотреть основные структурные компоненты научных и прикладных исследований,

- освоить способы и методы постановки проблемы новизны исследований, оформления, подготовки к публикации и внедрения результатов исследования.

В результате изучения дисциплины «Методология научного познания» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне.

Тематический план дисциплины:

Раздел 1. Введение в дисциплину

Тема 1. Базовые понятия теории и методологии научного исследования

Современные трактовки методологии научного исследования. Исследование как форма развития научного знания. Место и роль методологии в системе научного познания. Понятие метода научного исследования. Интегрирующая роль метода в научном познавательном процессе. Функции методологии науки как составной части научного исследования. Понятие методики научного исследования. Роль методики в организации научных и прикладных исследований.

Тема 2. Система методов научных исследований в предметной области

Система методов научного исследования. Понятия метода, принципа, способа познания. Проблема классификации методов. Общенаучные принципы и методы научного познания. Общенаучные подходы в исследовании. Субстратный подход. Структурный подход. Функциональный подход. Системный подход. Алгоритмический подход. Вероятностный подход. Информационный подход. Анализ и синтез. Абстрагирование и конкретизация. Дедукция и индукция. Методы научной дедукции. Аналогия. Требования к научной аналогии. Моделирование. Исторический и логический методы. Методы эмпирического исследования. Наблюдение. Измерение. Сравнение. Эксперимент. Методы теоретического исследования. Классификация. Обобщение и ограничение. Формализация. Аксиоматический метод. Система форм познания в научном исследовании. Понятие научного факта. Проблема. Требования к постановке проблем. Гипотеза. Требования к выдвижению гипотез. Научное доказательство. Опровержение. Теория. Обоснование истинности научного знания.

Тема 3. Система методов прикладных исследований в предметной области

Специальные методы исследования Математические методы. Метод формализации. Сетевые модели. Метод «затраты — выпуск». Социологические методы. Социально-психологические методы. Тестирование. Выборка. Стандартизированное (формальное) и свободное интервью. Панельное, клиническое, фокусированное интервью. Метод экспертных оценок. Статистическая сводка. Группировка. Корреляционный анализ. Принцип ковариации. Абсолютные величины. Относительные величины. Показатели интенсивности, экстенсивности, динамики, соотношения. Цепные и базисные темпы роста. Абсолютный прирост. Темп прироста. Средние величины. Средняя арифметическая, хронологическая, геометрическая. Балансовый метод. Нормативный метод. Графический метод. Размах вариации (диапазон колебаний). Дисперсия. Среднеквадратическое (стандартное) отклонение. Коэффициент вариации. Детерминированный факторный анализ. Аддитивные, мультипликативные, кратные, смешанные детерминированные модели. Метод элиминирования. Способ цепных подстановок.

Раздел 2. Методология научного исследования

Тема 4. Основные структурные компоненты научных и прикладных исследований в предметной области.

Научное исследование как вид деятельности. Структурные характеристики деятельностного цикла. Субъект, цель, объект, средства, условия, комплекс действий, результат, оценка результата — их проявление в научном исследовании. Практическая и теоретическая актуальность научного исследования. Оценка степени научной разработанности проблемы. Формулировка темы исследования. Признаки корректности формулировки темы: семантическая корректность, прагматическая корректность. Формулировка цели научного исследования как прогнозирование основных результатов исследования. Задачи научного исследования как формулировки частных вопросов, решение которых обеспечивает достижение основного результата исследования. Понятие объекта и предмета научного исследования. Их соотношение и взаимные переходы. Эмпирическая и теоретическая база исследования. Интегральный метод исследования. Логика и структура научного исследования.

Тема 5. Проблема новизны научных и прикладных исследований в предметной области.

Понятие и признаки новизны научного исследования. Новизна эмпирических исследований: определение новых неизученных областей; выявление новых проблем; получение новых (не зафиксированных ранее) фактов; введение новых фактов в научный оборот; обработка известных фактов новыми методами; выявление новых видов корреляции между фактами; формулирование неизвестных ранее эмпирических закономерностей; разработка новых методов и методик осуществления эмпирических исследований. Новизна теоретических исследований: новизна вводимых понятий, или трактовки существующего понятийного аппарата; новизна поставленной теоретической проблемы; новизна гипотезы;

новизна теоретических положений внутри действующей парадигмы; аргументированная новизна межпарадигмальной теории; разработка новых методов и методик осуществления теоретических исследований.

Тема 6. Оформление, публикация и внедрение результатов исследования

Структура и содержание отчёта о научно-исследовательской работе. Учебно-исследовательские и научно-исследовательские работы. Рукопись научного исследования: содержание и структура. Требования к оформлению. Научный отчёт. Система стандартов по информации, библиотечному и издательскому делу (СИБИД). ГОСТ 7.32-2001. ГОСТ 7.73-96. ГОСТ 7.60-2003. Средства и формы публичного представления результатов научной работы. Научный стиль. Логичность. Ясность. Последовательность. Публикации: доклады, статьи, тезисы, сообщения и др. Официальное рецензирование результатов научного исследования. Демонстрационный материал и техника. Оформление на конкурсы, выставки, конференции. Планирование внедрения: формы, этапы и документальное оформление. Оценка эффективности научных исследований. Методы оценки результатов исследований (теоретических и прикладных). Виды эффектов от НИР (научно-технических, социальный, экономический эффект) – критерии и методы расчёта. Патентно-информационное обеспечение исследований и разработок.

Собеседование по практическим работам

Собеседование по выполнению практических работ осуществляется с целью проверки освоения студентом всех запланированных по дисциплине «Методология научного познания» компетенций. Каждое практическое занятие студент выполняет некоторое задание по конкретной теме. Шкала оценивания имеет вид (таблица 1).

Таблица 1

Шкала оценивания

Критерии оценки качества решения задачи	Балл
Обучающийся демонстрирует неполные знания теоретического материала, но не может сопоставить его с практическим материалом, не может определить правильный алгоритм выполнения задания, допускает значительные неточности, не может провести анализ результатов с учетом помощи преподавателя, а также сформулировать выводы	1
Обучающийся демонстрирует неполные знания теоретического и практического материала, не может определить правильный алгоритм выполнения задания, допускает значительные неточности, не может провести анализ результатов с учетом помощи преподавателя, а также сформулировать выводы	2
Обучающийся демонстрирует неполные знания теоретического и практического материала, определяет правильный алгоритм решения	3

задачи или выполнения задания, допускает значительные неточности, проводит анализ результатов при помощи преподавателя, выводы формулирует также при помощи преподавателя	
Обучающийся демонстрирует знания теоретического и практического материала, определяет правильный алгоритм выполнения практического задания, допускает незначительные неточности, проводит неполный анализ результатов	4
Обучающийся демонстрирует знания теоретического и практического материала, определяет правильный алгоритм выполнения задания, проводит анализ полученных результатов, формулирует правильные выводы	5

Зачет

Зачет по дисциплине проводится в устной форме с ответом на сопутствующие теоретические вопросы.

Шкала оценивания имеет вид (таблица 2)

Таблица 2

Шкала оценивания

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Дан полный, развернутый ответ на поставленный вопрос, показана совокупность осознанных знаний по дисциплине, доказательно раскрыты основные положения вопросов; в ответе прослеживается четкая структура, логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений. Знания по предмету демонстрируются на фоне понимания его в системе данной науки и междисциплинарных связей. Ответ изложен литературным языком. Могут быть допущены недочеты в определении понятий, исправленные студентом самостоятельно в процессе ответа	Зачтено
Дан полный, развернутый ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные связи. Ответ четко структурирован, логичен, изложен литературным языком. Могут быть допущены некоторые неточности или незначительные ошибки, исправленные студентом с помощью преподавателя	Зачтено
Дан недостаточно полный и недостаточно развернутый	Зачтено

<p>ответ. Логика и последовательность изложения имеют нарушения. Допущены ошибки в раскрытии понятий, употреблении терминов. Обучающийся не способен самостоятельно выделить существенные и несущественные признаки и причинно-следственные связи. В ответе отсутствуют выводы. Умение раскрыть значение обобщенных знаний не показано. Речевое оформление требует поправок, коррекции</p>	
<p>1) Ответ представляет собой разрозненные знания с существенными ошибками по вопросу. Присутствуют фрагментарность, нелогичность изложения. Обучающийся не осознает связь обсуждаемого вопроса с другими объектами дисциплины. Отсутствуют выводы, конкретизация и доказательность изложения. Речь неграмотная, терминология не используется. Дополнительные и уточняющие вопросы преподавателя не приводят к коррекции ответа студента.</p> <p>2) Ответ на вопрос полностью отсутствует.</p> <p>3) Отказ от ответа</p>	<p>Не зачтено</p>

1. Базовые понятия теории и методологии научного исследования

Наука — сфера человеческой деятельности, направленной на выработку и теоретическую систематизацию объективных знаний о действительности. Основой этой деятельности является сбор фактов, их постоянное обновление и систематизация, критический анализ и, на этой базе, синтез новых знаний или обобщений, которые не только описывают наблюдаемые природные или общественные явления, но и позволяют построить причинно-следственные связи и, как следствие, — прогнозировать. Те теории и гипотезы, которые подтверждаются фактами или опытами, формулируются в виде законов природы или общества.

Наука в широком смысле включает в себя все условия и компоненты соответствующей деятельности:

- разделение и кооперацию научного труда;
- научные учреждения, экспериментальное и лабораторное оборудование;
- методы научно-исследовательской работы;
- понятийный и категориальный аппарат;
- систему научной информации;
- всю сумму накопленных ранее научных знаний.

Совокупность занимающихся наукой людей составляет научное сообщество. Научное сообщество представляет собой сложную самоорганизующуюся систему, в которой действуют и государственные институты, и общественные организации, и неформальные группы.

Отличительной чертой этого сообщества является повышенная степень признания авторитета, достигнутого научными успехами, и сниженный уровень признания авторитета властного, что порой приводит к конфликту государства и научного сообщества. Также следует отметить более высокую, чем в других социальных сферах, эффективность неформальных групп и особенно отдельных личностей.

Важнейшими функциями научного сообщества являются признание или отрицание новых идей и теорий, обеспечивающее развитие научного знания, а также поддержка системы образования и подготовки новых научных кадров.

Изучать в научном смысле — это значит быть научно объективным. Нельзя отбрасывать факты в сторону только потому, что их трудно объяснить или найти им практическое применение. Дело в том, что сущность нового в науке не всегда видна самому исследователю. Новые научные факты и даже открытия из-за того, что их значение плохо раскрыто, могут долгое время оставаться в резерве науки и не использоваться на практике.

Развитие идеи до стадии решения задачи обычно совершается как плановый процесс научного исследования. Науке известны и случайные открытия, но только плановое, хорошо оснащенное современными средствами научное исследование надежно позволяет вскрыть и глубоко познать объективные закономерности в природе. В дальнейшем процесс целевой и

общеидейной обработки первоначального замысла продолжается, вносятся уточнения, изменения, дополнения, развивается намеченная схема исследования.

Научное исследование – это целенаправленное познание, результаты которого выступают в виде системы понятий, законов и теорий.

Характеризуя научное исследование, обычно указывают на следующие его отличительные признаки:

- это обязательно целенаправленный процесс, достижение осознанно поставленной цели, четко сформулированных задач;
- это процесс, направленный на поиск нового, на творчество, на открытие неизвестного, на выдвижение оригинальных идей, на новое освещение рассматриваемых вопросов;
- оно характеризуется систематичностью: здесь упорядочены, приведены в систему и сам процесс исследования, и его результаты;
- ему присуща строгая доказательность, последовательное обоснование сделанных обобщений и выводов.

Объектом научно-теоретического исследования выступает не просто отдельное явление, конкретная ситуация, а целый класс сходных явлений и ситуаций, их совокупность.

Цель, непосредственные задачи научно-теоретического исследования состоят в том, чтобы найти общее у ряда единичных явлений, вскрыть законы, по которым возникают, функционируют, развиваются такого рода явления, т.е. проникнуть в их глубинную сущность.

Основные средства научно-теоретического исследования:

- совокупность научных методов, всесторонне обоснованных и сведенных в единую систему;
- совокупность понятий, строго определенных терминов, связанных между собою и образующих характерный язык науки.

Результаты научных исследований воплощаются в научных трудах (статьях, монографиях, учебниках, диссертациях и т. д.) и лишь затем, после их всесторонней оценки используются в практике, учитываются в процессе практического познания и в снятом, обобщенном виде включаются в руководящие документы.

Основными характеристиками исследования являются: объект и предмет исследования, тема, цель и задачи исследования, научная идея или концепция, теоретическая и практическая значимость, выводы исследования и сформулированные практические и теоретические рекомендации.

Тема исследования отражает определенный аспект проблемы, решение которой будет ставиться в качестве цели исследования, и содержит в себе направленность на принятие решения.

Цель научного исследования - определение конкретного объекта и всестороннее, достоверное изучение его структуры, характеристик, связей на основе разработанных в науке

принципов и методов познания, а также получение полезных для деятельности человека результатов, внедрение в производство с дальнейшим эффектом.

Носитель проблемной ситуации, на которую направлена исследовательская работа представляет собой объект исследования.

Предмет исследований - свойство, поведение, функционирование объекта, подлежащего непосредственному изучению. Объект и предмет исследования соотносятся между собой как общее и частное.

Концепция - руководящая идея или совокупность идей, предлагающих новую теоретическую базу для понимания и исследования явлений или процессов, происходящих в обществе, система взглядов на что-либо, основная мысль, когда определяются цели, задачи исследования и указываются пути его ведения.

Задачи исследования формулируются как конкретные этапы решения представленной проблемы и являются, по сути, содержательной детализацией поставленной цели.

Оценка исследования осуществляется как с точки зрения теоретической ценности (новизна исследования, актуальность, оригинальность), так и с позиций практической значимости.

Заключение формируется в конце научного исследования и содержит основные выводы, результаты по каждой поставленной задаче.

Вопросы для собеседования

1. Понятие методологии научного исследования. Чем обусловлена роль методологии в научном исследовании?
2. Понятие методики научного исследования. Имеет ли смысл различать методологию и методику?
3. Функции методологии науки. Не ограничивают ли методология и методика творчество исследователя?
4. Организация научно-экономического исследования. Как взаимосвязаны методология, методика и организация экономического исследования?
5. Содержание и структура методологической культуры исследователя.
6. Каковы наилучшие формы повышения методологической культуры исследователя?

2. Система методов научных исследований в предметной области

Метод (греч. – способ познания) – в самом широком смысле слова – «путь к чему-либо», способ деятельности субъекта в любой ее форме.

Понятие «методология» имеет два основных значения: система определенных способов и приемов, применяемых в той или иной сфере деятельности (в науке, политике, искусстве и т.п.); учение об этой системе, общая теория метода, теория в действии.

Методология – учение о правилах мышления при создании теории науки.

На первых порах методология вытекала из знаний, предписанных геометрией как наукой, где содержались нормативные указания по изучению реального мира.

Затем методология выступила как комплекс правил по изучению мироздания и перешла в сферу философии. Платон и Аристотель смотрели на методологию как на логическую универсальную систему, как на орудие истинного познания.

Современное понимание методологии основывается на представлении о роли и многообразии функций, которые выполняет сама наука для познания и регулирования социальных и природных процессов и явлений, формирования общественного сознания и мировоззрения, повышения эффективности человеческой деятельности. Оно важно для рационального использования науки, производства, разработки стратегии и тактики развития экономики и культуры, социального прогресса и всестороннего развития личности.

Определение методологии должно опираться на следующие признаки этого понятия.
Методология:

- а) определяет способы получения научных знаний, которые отражают постоянно меняющуюся педагогическую действительность;
- б) направляет, предопределяет основной путь, с помощью которого достигается определенная научно- исследовательская цель;
- в) обеспечивает всесторонность получения информации об изучаемом процессе или явлении;
- г) помогает введению новой информации в фонд теории научного познания;
- д) обеспечивает уточнение, обогащение, систематизацию терминов и понятий в науке;
- е) создает систему научной информации, опирающуюся на объективные факты и логико-аналитический инструмент научного познания.

Эти признаки понятия «методология», определяющие ее функции в науке, позволяют сделать следующий вывод: методология – это концептуальное изложение цели, содержания, методов исследования, которое обеспечивает получение максимально объективной, точной, систематизированной информации о происходящих процессах и явлениях.

Исходя из этого выделяют:

- методы-операции;
- методы-действия.

Такой подход не противоречит определению метода, которое дает энциклопедический словарь:

- во-первых, метод как способ достижения какой-либо цели, решения конкретной задачи – метод-действие;
- во-вторых, метод как совокупность приемов или операций практического или теоретического освоения действительности – метод-операция.

Таким образом, рассматривать методы исследования можно в следующей группировке:

Теоретические методы:

- методы – познавательные действия: диалектика; научные теории, проверенные практикой; доказательство; метод анализа систем знаний; дедуктивный (аксиоматический); индуктивно-дедуктивный; выявление и разрешение противоречий; постановка проблемы; построение гипотезы;
- методы-операции: анализ, синтез, сравнение, абстрагирование, конкретизация, обобщение, формализация, индукция, дедукция, идеализация, аналогия, моделирование, мысленный эксперимент, воображение.

Эмпирические методы:

- методы отслеживания объекта (обследование, мониторинг, изучение и обобщение опыта);
- методы преобразования объекта (опытная работа, эксперимент);
- методы исследования объекта во времени (ретроспектива, прогнозирование);
- методы-операции: изучение литературы, документов и результатов деятельности (наблюдение, измерение, опрос (устный и письменный), экспертные оценки, тестирование).

Научные методы также классифицируют и по следующим признакам: в зависимости от содержания изучаемых объектов различают методы естествознания и методы социально-гуманитарного исследования.

Методы исследования классифицируют и по отраслям науки: математические; биологические; медицинские; социально-экономические; правовые и т.д.

Методами метатеоретического уровня являются диалектический; метафизический; герменевтический и др. Некоторые ученые к этому уровню относят метод системного анализа, а другие его включают в число общелогических методов.

В зависимости от сферы применения и степени общности различают методы:

1. всеобщие (философские), действующие во всех науках и на всех этапах познания;
2. общенаучные, которые могут применяться в гуманитарных, естественных и технических науках;
3. частные – для родственных наук;
4. специальные – для конкретной науки, области научного познания. Подобную классификацию методов можно встретить и в юридической литературе.

К числу важнейших гносеологических признаков научного метода (к какому бы типу он ни относился) относятся:

- объективность – опосредствованность достоверным знанием;
- общезначимость – всеобщий интерсубъективный характер научного метода в отличие от остающейся уделом ненауки персонифицированности, уникальности;
- воспроизводимость – инвариантность результатов для любого субъекта в любой сходной ситуации;
- целесообразность – определенность, заданность принципов интеллектуального движения, осмысленность реализации как отдельных шагов, так и систем операций в целом;
- необходимость – гарантированность результатов в отличие от ненаучной особенности случайного, непреднамеренного их достижения;
- эффективность – запланированность социальной ассимиляции, внедрения, потребления результатов, что не свойственно ненаучному познанию, базирующемуся на ситуативном, индивидуально конституированном способе получения и применения результатов.

От рассматриваемого понятия метода следует отграничивать понятия техники, процедуры и методики научного исследования.

Под техникой исследования понимают совокупность специальных приемов для использования того или иного метода, а под процедурой исследования – определенную последовательность действий, способ организации исследования.

Методика – это совокупность способов и приемов познания. Например, под методикой криминологических исследований понимают систему способов, приемов, средств сбора, обработки, анализа и оценки информации о преступности, её причинах и условиях, личности преступника и других криминологических явлениях.

Любой научный метод разрабатывается на основе определенной теории, которая тем самым выступает его необходимой предпосылкой.

Эффективность, сила того или иного метода обусловлена содержательностью, глубиной, фундаментальностью теории, которая «сжимается в метод».

В свою очередь «метод расширяется в систему», т.е. используется для дальнейшего развития науки, углубления и развертывания теоретического знания как системы, его материализации, объективизации в практике.

Тем самым теория и метод одновременно тождественны и различны. Их сходство состоит в том, что они взаимосвязаны, и в своем единстве отражают реальную действительность.

Будучи едиными в своем взаимодействии, теория и метод не отделены жестко друг от друга и в то же время не есть непосредственно одно и то же. Они взаимопереходят, взаимопреобразуются: теория, отражая действительность, преобразуется, трансформируется в метод посредством разработки, формулирования вытекающих из нее принципов, правил,

приемов и т.п., которые возвращаются в теорию (а через нее – в практику), ибо субъект применяет их в качестве регулятивов, предписаний, в ходе познания и изменения окружающего мира по его собственным законам.

Однако нельзя полностью отождествлять научную теорию и методы познания и утверждать, что всякая теория и есть вместе с тем метод познания и действия. Метод не тождествен прямо и непосредственно теории, а теория не является непосредственно методом, ибо не она есть метод познания, а необходимо вытекающие из нее методологические установки, требования, регулятивы.

Основные различия теории и метода состоят в следующем:

- теория – результат предыдущей деятельности, метод – исходный пункт и предпосылка последующей деятельности;
- главные функции теории – объяснение и предсказание (с целью отыскания истины, законов, причины и т. п.), метода – регуляция и ориентация деятельности;
- теория – система идеальных образов, отражающих сущность, закономерности объекта, метод – система регулятивов, правил, предписаний, выступающих в качестве орудия дальнейшего познания и изменения действительности;
- теория нацелена на решение проблемы, что собой представляет данный предмет, метод – на выявление способов и механизмов его исследования и преобразования.

Следует иметь в виду, что в современной науке понятие «предмет познания» употребляется в двух основных значениях.

Во-первых, как предметная область – стороны, свойства, отношения действительности, обладающие относительной завершенностью, целостностью и противостоящие субъекту в его деятельности (объект познания). Например, предметная область в зоологии – это множество животных. Различные науки об одном и том же объекте имеют различные предметы познания (например, анатомия изучает строение организмов, физиология – функции его органов и т.п.).

Предметы познания могут быть как материальными, так и идеальными (сам процесс познания, его формы, уровни и т.д., различного рода абстракции, духовная культура или такие состояния как «дух народа», «дух времени» и т.д.).

Во-вторых, как система законов, которым подчиняется данный объект. Нельзя «разводить» предмет и метод, видеть в последнем только внешнее средство по отношению к предмету, никак не зависимое от него и лишь «налагаемое» на предмет чисто внешним образом.

В настоящее время стало очевидным, что система методов, методология не может быть ограничена лишь сферой научного познания, она должна выходить за ее пределы и непременно включать в свою орбиту и сферу практики. При этом необходимо иметь в виду тесное взаимодействие этих двух сфер.

одержание изучаемых наукой объектов служит критерием для различия методов естествознания и методов социально-гуманитарных наук. К числу характерных признаков научного метода (к какому бы типу он ни относился) чаще всего относят: объективность, воспроизводимость, эвристичность, необходимость, конкретность и др.

В современной науке достаточно успешно «работает» многоуровневая концепция методологического знания. В этом плане все методы научного познания могут быть разделены на следующие основные группы (по степени общности и широте применения).

1. Философские методы, среди которых наиболее древними являются диалектический и метафизический. По существу каждая философская концепция имеет методологическую функцию, является своеобразным способом мыслительной деятельности. Поэтому философские методы не исчерпываются двумя названными. К их числу также относятся такие методы как аналитический (характерный для современной аналитической философии), интуитивный, феноменологический, герменевтический (понимание) и др.

2. Общенаучные подходы и методы исследования, которые получили широкое развитие и применение в науке. Они выступают в качестве своеобразной «промежуточной методологии» между философией и фундаментальными теоретико- методологическими положениями специальных наук.

К общенаучным понятиям чаще всего относят такие понятия, как «информация», «модель», «структура», «функция», «система», «элемент», «оптимальность», «вероятность» и др.

Характерными чертами общенаучных понятий являются, во-первых, «сплавленность» в их содержании отдельных свойств, признаков, понятий ряда частных наук и философских категорий.

На основе общенаучных понятий и концепций формулируются соответствующие методы и принципы познания, которые и обеспечивают связь и оптимальное взаимодействие философии со специально-научным знанием и его методами. К числу общенаучных принципов и подходов относятся системный и структурно-функциональный, кибернетический, вероятностный, моделирование, формализация и ряд других.

Особенно бурно в последнее время развивается такая общенаучная дисциплина как синергетика – теория самоорганизации и развития открытых целостных систем любой природы природных, социальных, когнитивных (познавательных).

Среди основных понятий синергетики такие понятия как «порядок», «хаос», «нелинейность», «неопределенность», «нестабильность», «диссипативные структуры», «бифуркация» и др.

3. Частно-научные методы – совокупность способов, принципов познания, исследовательских приемов и процедур, применяемых в той или иной науке, соответствующей данной основной форме движения материи. Это методы механики, физики, химии, биологии и социально-гуманитарных наук.

4. Формальные и содержательные методы – отображение содержательного знания в знаково-символическом виде, базируются на различие естественных и искусственных языков.

Дальнейшее углубление формализации связано с построением искусственных (формализованных) языков, предназначенных для более точного и строго выражения знания, чем естественный язык, с целью исключить возможность неоднозначного понимания – что характерно для естественного языка (язык математики, логики, химии и др.).

Именно использование специальной символики позволяет устранить многозначность слов обычного языка. В формализованных рассуждениях каждый символ строго однозначен.

Достоинство искусственных языков состоит, прежде всего в их точности, однозначности, а самое главное – в возможности представления обычного содержательного рассуждения посредством вычисления.

Значение формализации в научном познании состоит в следующем:

- она дает возможность анализировать, уточнять, определять и разъяснять (эксплицировать) понятия. Обыденные представления (выражаемые в разговорном языке), хотя и кажутся более ясными и очевидными с точки зрения здравого смысла, но оказываются неподходящими для научного познания в силу их неопределенности, неоднозначности и неточности;
- она приобретает особую роль при анализе доказательств. Представление доказательств в виде последовательности формул, получаемых из исходных с помощью точно указанных правил преобразования, придает ему необходимую строгость и точность;
- она служит основой для процессов алгоритмизации и программирования вычислительных устройств, а тем самым и компьютеризации не только научно-технического, но и других форм знания.

При формализации рассуждения об объектах переносятся в плоскость оперирования со знаками (формулами). Отношения знаков заменяют собой высказывания о свойствах и отношениях предметов.

Таким путем создается обобщенная знаковая модель не которой предметной области, позволяющая обнаружить структуру различных явлений и процессов при отвлечении от качественных, содержательных характеристик последних.

Главное в процессе формализации состоит в том, что над формулами искусственных языков можно производить операции, получать из них новые формулы и соотношения.

Тем самым операции с мыслями о предметах заменяются действиями со знаками и символами. Формализация в этом смысле представляет собой логический метод уточнения содержания мысли посредством уточнения ее логической формы.

5. Гипотетико-дедуктивный метод. Его сущность заключается в создании системы дедуктивно связанных между собой гипотез, из которой в конечном счете выводятся утверждения об эмпирических фактах.

Этот метод тем самым основан на выведении (дедукции) заключений из гипотез и других посылок, истинное значение которых неизвестно. Поэтому заключения тут носят вероятный характер. Такой характер заключения связан еще и с тем, что в формировании гипотезы участвует и догадка, и интуиция, и воображение, и индуктивное обобщение, не говоря уже об опыте, квалификации и таланте ученого. А все эти факторы почти не поддаются строго логическому анализу.

Исходные понятия: гипотеза (предположение) – положение, выдвигаемое в начале предварительного условного объяснения некоторого явления или группы явлений; предположение о существовании некоторого явления. Истинность такого допущения неопределенна, оно проблематично.

Дедукция (выведение): а) в самом общем смысле – это переход в процессе познания от общего к частному (единичному), выведение последнего из первого; б) в специальном смысле – процесс логического вывода, т.е. перехода по определенным правилам логики от некоторых данных предположений (посылок) к их следствиям (заключениям).

Общая структура гипотетико-дедуктивного метода (или метода гипотез):

- ознакомление с фактическим материалом, требующим теоретического объяснения, и попытка такового с помощью уже существующих теорий и законов. Если нет, то:
- выдвижение догадки (предположение) о причинах и закономерностях данных явлений с помощью многих логических приемов;
- оценка серьезности предположений и отбор из множества догадок наиболее вероятной.

При этом гипотеза проверяется на:

- логическую непротиворечивость;
- совместимость с фундаментальными теоретическими принципами данной науки (например, с законом сохранения и превращения энергий).

Разновидностью гипотетико-дедуктивного метода можно считать математическую гипотезу, где в качестве гипотез вы ступают некоторые уравнения, представляющие модификацию ранее известных и проверенных соотношений. Изменяя эти соотношения, составляют новое уравнение, выражающее гипотезу, которая относится к неисследованным явлениям.

Гипотетико-дедуктивный метод является не столько методом открытия, сколько способом построения и обоснования научного знания.

6. Дисциплинарные методы – система приемов, применяемых в той или иной научной дисциплине, входящей в какую-нибудь отрасль науки или возникшей на стыках наук. Каждая

фундаментальная наука представляет собой комплекс дисциплин, которые имеют свой специфический предмет и свои своеобразные методы исследования.

7. Методы междисциплинарного исследования – совокупность ряда синтетических, интегративных способов (возникших как результат сочетания элементов различных уровней методологии), нацеленных главным образом на стыки научных дисциплин. Широкое применение эти методы нашли в реализации комплексных научных программ.

Вопросы для собеседования

1. Понятия метода, принципа, способа познания.
2. Философские и общенаучные принципы и методы научного познания.
3. Общенаучные подходы в научном исследовании.
4. Общенаучные методы познания.
5. Методы эмпирического исследования.
6. Методы теоретического исследования.
7. Понятие научного факта.
8. Понятие и требования к научной гипотезе.
9. Научное доказательство и опровержение.
10. Понятие и виды теорий.

3. Система методов прикладных исследований в предметной области

Эмпирический уровень научного познания характеризуется непосредственным исследованием реально существующих, чувственно воспринимаемых объектов. На этом уровне осуществляется процесс накопления информации об исследуемых объектах (путем наблюдения, измерения, экспериментов) здесь происходит первичная систематизация полученных знаний (в виде таблиц, схем, графиков).

Теоретический уровень научного исследования осуществляется на рациональной (логической) ступени познания. На данном уровне происходит выявление наиболее глубоких, существенных сторон, связей, закономерностей, присущих изучаемым объектам и явлениям. Результатом теоретического познания становятся гипотезы, теории, законы.

Среди теоретических методов научных исследований выделяют формализацию как метод, основанный на выявлении и фиксации формальной структуры исследуемого процесса или явления, а также приписывающий содержательным элементам процесса (явления) некоторые абстрактные символы и значения²⁴. Результатом формализации является построение некоторой модели, позволяющей получить новое знание или информацию об исследуемом объекте. Процесс формализации может осуществляться различными способами. Первый способ представляет собой формализацию на базе естественного языка. Однако этот тип формализации не обеспечивает однозначного использования понятий и, соответственно, приводит к трудностям научной коммуникации. В этой связи полезным является использование логической формализации, выполняемой на основе правил формальной и математической логики. Основной частью логической формализации является ее символизация, т.е. разработка системы определенных условных обозначений в рамках создания искусственного языка науки.

Наиболее распространенным и эффективным типом формализации является математизация. Математизацию можно определить, как специфический метод формализации, основанный на применении процедур измерения, сравнения и счета.

Математизация в экономике и менеджменте развивалась по следующим направлениям:

- математическое моделирование — построение формально- количественных, математических моделей исследуемых процессов и явлений;
- эконометрика — решение социально-экономических задач с применением принципов статистического измерения и сбора данных, численного выражения исследуемой реальности с определением меры и границ этого выражения;
- математический эксперимент - метод многократного математического моделирования как эксперимент над моделями;
- неметрическая математизация экономического знания.

Математическое моделирование является одним из основных инструментов научных исследований в экономике, что привело к созданию значительного числа экономико-математических моделей. Математическая модель является приближенным представлением реальных объектов, процессов или систем, выраженным в математических терминах и сохраняющим существенные черты оригинала. Математические модели в количественной форме, с помощью логико-математических конструкций, описывают основные свойства объекта, процесса или системы, его параметры, внутренние и внешние связи. Можно выделить следующие модели.

Оптимизация с несколькими альтернативами предполагает создание модели, позволяющей найти решение из конечного числа альтернатив, представленных в виде таблицы или графа с их прогнозируемыми вкладами в достижение цели и вероятностями осуществления таких вкладов. Модели, реальные процессы и явления в которых описываются в виде явных функциональных зависимостей (уравнений линейных или нелинейных, дифференциальных или интегральных, систем этих уравнений), получили название аналитических моделей. Их применение в экономике и менеджменте ограничено сложностью исследуемых процессов и явлений, когда аналитические зависимости не всегда возможно построить. В этом случае исследователи прибегают к имитационному моделированию.

Имитационное моделирование представляет собой численный метод проведения вычислительных экспериментов средствами информационных технологий и систем с математическими моделями, имитирующими поведение реальных объектов, процессов, явлений во времени в течение заданного периода. Его реализация основывается на совокупности методов алгоритмизации функционирования объектов исследований, программной реализации алгоритмических описаний, организации, планирования и выполнения средствами информационных технологий экспериментов с математическими моделями, имитирующими функционирование исследуемые процессы и явления в течение заданного периода. Термин «имитационное моделирование» означает, что мы имеем дело с такими математическими моделями, с помощью которых нельзя заранее вычислить или предсказать поведение системы, а для предсказания поведения системы необходим вычислительный эксперимент (имитация) на математической модели при заданных исходных данных.

Эвристика — специальные методы решения задач (эвристические методы), которые обычно противопоставляются формальным методам решения, опирающимся на точные математические модели. Использование эвристических методов (эвристик) сокращает время решения задачи по сравнению с методом полного ненаправленного перебора возможных альтернатив; получаемые решения не являются, как правило, наилучшими, а относятся лишь к множеству допустимых решений.

Предсказательные модели основываются на методах прогнозирования и имеют важное значение в исследовательских проектах, направленных на повышение эффективности экономики.

Оптимизация с использованием математического программирования теория и методам решения задач о нахождении экстремумов функций на множествах, определяемых линейными и нелинейными ограничениями (равенствами и неравенствами). Наиболее известным методом этого семейства инструментов является линейное программирование. Однако практическое применение находят и другие методы математического программирования: нелинейное программирование, динамическое программирование, стохастическое программирование.

Большое число экономических задач, требующих учета разнообразных факторов и характеристик, изменяющихся во времени и влияющих друг на друга и на экономические процессы, сводится к линейным математическим моделям. Традиционно оптимизационные линейные математические модели называют моделями линейного программирования.

Таким образом, оптимальное решение – это решение, наиболее предпочтительное перед другими по определенному критерию эффективности.

Наиболее распространенный способ решения задачи – симплекс – метод. Симплекс-метод является методом направленного перебора решений системы. Каждое следующее решение улучшает значение целевой функции.

Симплекс- метод включает два этапа:

1. Определение начального решения, удовлетворяющего ограничениям;
2. Последовательное улучшение начального решения и получение оптимального решения задачи.

Статистические методы в научных исследованиях

Достаточно часто при выполнении научных исследований описываются ситуации, при которых рассматриваются детерминированные события функции, когда каждое событие является следствием другого, а физические законы описываются математическими зависимостями одной величины от другой. Человеческая деятельность достаточно часто опровергает это правило, при этом фиксируются не только отклонения измеряемых параметров при описании ситуаций, но и значительная неопределенность некоторых из них. Например, невозможно точно предсказать количество покупателей в магазине или объем продаж в определенный момент времени. Величины, точное значение которых не известно называют случайными.

Несмотря на случайный характер большинства параметров, анализируемых в экономике и менеджменте, научные исследования позволяют сформулировать и описать некоторые закономерности, которые широко используются в практической деятельности. В основе таких исследований лежит статистика - совокупность методов, позволяющих принимать решения в условиях неопределенности.

Статистическая наука состоит из нескольких разделов, каждый из которых имеет большое значение для исследований в экономике и менеджменте. Основу статистики представляет статистическая теория, которая отражает общие принципы статистической науки, ее категории и методы исследования. Ядром является математическая статистика - направление математики, базирующееся на численных методах обработки статистических данных, полученных в результате наблюдения и измерения случайных величин. Еще один значимый раздел статистической науки - прикладная статистика. Она исследует статистические проблемы, касающиеся непосредственно практической деятельности: измерение, сбор, обработка и анализ данных в различных областях исследования.

Основным условием правильного восприятия и практического использования статистической информации в экономике является знание статистической методологии.

Статистическая методология как система приемов, способов, методов, направленных на изучение количественных закономерностей, проявляющихся во взаимосвязи с социально-экономическими явлениями, включает следующие этапы исследования:

1. Организованная регистрация собранных фактов о массовых социально-экономических, общественных явлениях и процессах.

2. Сводка и группировка статистических данных – это систематизация первичных данных по признакам, объединяющим в качественно однородные группы.

3. Анализ совокупных данных, полученных в результате сводки и группировки, при этом используются обобщенные показатели: абсолютных, относительных и средних величин, показатели вариации, ряды динамики, анализ взаимосвязей и индексы.

Таким образом, основой статистических исследований являются метод статистического наблюдения, группировка и методы статистического анализа.

Метод статистического наблюдения является предварительной стадией статистического исследования, которая представляет собой планомерный, научно организованный учет (сбор) первичных статистических данных о массовых социально-экономических явлениях и процессах. Статистическое наблюдение должно отвечать ряду важнейших требований:

- проводиться непрерывно и систематически;
- учет массовых данных должен быть таким, чтобы не только обеспечивалась полнота данных, но и учитывалось их постоянное изменение;
- данные должны быть достоверными и точными;
- данные должны соответствовать принципу единообразия и сопоставимости.
- исследуемые явления должны иметь не только научную, но и практическую ценность.

При проведении научных исследований не удастся получить всю совокупность значений изучаемой величины (так называемую генеральную совокупность), поэтому в большинстве случаев используется только часть данных (выборка).

При проведении исследований необходимо:

- применять однородную, репрезентативную выборку, чтобы результаты исследования могли бы быть распространены на генеральную совокупность;
- выполнять анализ точности полученных результатов;
- обосновывать способы формирования выборки;
- применять для обработки данных способы, соответствующие их особенностям.

Сбор и регистрация статистических фактов предполагает возможность их измерения - присвоения чисел исследуемым предметам, событиям, явлениям, процессам. Для регистрации измерений применяются шкалы:

- шкала классификации (наименования);
- шкала порядка;
- шкала интервалов;
- шкала отношений.

Шкала классификации (наименований) используется для идентификации и классификации объектов и позволяет выполнить сравнение на уровне «равно» и «не равно». Шкала порядка (порядковая шкала) - это шкала рангов, позволяющая исследователям определить больше или меньше характеристика одного объекта по сравнению с другим. Однако, используя эту шкалу, мы не можем определить на сколько больше или меньше величина исследуемой характеристики. Например, исследование с использованием порядковой шкалы позволяет выявить отношение респондентов к характеристике объекта в виде «неудовлетворен», «скорее не удовлетворен», «скорее удовлетворен», «удовлетворен», но не позволяет выявить, в какой степени респондент не удовлетворен исследуемым аспектом. Шкала интервалов позволяет сравнивать величины и определять «насколько больше», «насколько меньше». Шкалы отношений ориентированы на выяснение вопроса «во сколько раз». В экономике к шкалам отношений относятся такие переменные как доля рынка, объем продаж, количество потребителей.

Случайная величина может быть дискретной или непрерывной. Если множество значений случайной величины конечно или счетно, т.е. их можно пронумеровать, то случайная величина называется дискретной. Случайная величина называется непрерывной, если она принимает все возможные значения из некоторого промежутка или на все числовой оси. Дискретная величина обычно задается рядом распределения, непрерывная величина - функцией или плотностью распределения.

Для обработки данных, собранных в результате наблюдений, используют метод группировки. Группировка представляет собой распределение множества единиц исследуемой совокупности по группам в соответствии с существенным для данной группы признаком. Несомненным достоинством данного метода возможность первичного обобщения данных, сравнительный анализ, обобщение показателей по группам и внутри группы. Появляется возможность сравнивать, анализировать причины различий между группами, изучать взаимосвязи между признаками.

Признак, по которому происходит выделение групп или типов явлений, называется группировочным или основанием группировки. Основание может быть количественным или атрибутивным. Атрибутивный - это признак, имеющий наименование (например, профессия: инженер, учитель и т.д.). Если в основе группировки находится количественный признак, то возникает вопрос об исчислении интервалов группировки, которые могут как равные так и неравные и, как правило, представляет собой промежуток между максимальными и минимальными значениями признака в группе.

Устойчивое разграничение объектов представляет собой классификацию. Она основывается на устойчивых признаках, (например, классификация отраслей народного хозяйства, классификация основных фондов и т.д.).

Статистический анализ можно рассматривать как количественное исследование в экономике и менеджменте, проведенное в соответствии с требованиями статистической теории и методологии.

Вопросы для собеседования

1. Специальные методы исследования
2. Математические методы.
3. Метод формализации.
4. Сетевые модели.
5. Социологические методы. Социально-психологические методы.
6. Метод экспертных оценок.
7. Статистическая сводка. Группировка.
8. Корреляционный анализ. Принцип ковариации.
9. Размах вариации (диапазон колебаний). Дисперсия. Среднеквадратическое (стандартное) отклонение. Коэффициент вариации.
10. Детерминированный факторный анализ. Аддитивные, мультипликативные, кратные, смешанные детерминированные модели.

Практическая работа №1

1 Расширенный план магистерской диссертации.

Здесь студент по выбранной теме своего исследования составляет расширенный план магистерской диссертации.

2 Формирование списка источников для написания магистерской диссертации (по правилам оформления в соответствии с ГОСТ) – не менее 25 источников включить на этом этапе – не старше 5 лет. Для статей – не старше 3 лет.

Начало выполнения магистерской диссертации связано с подбором литературы, который целесообразно начинать с изучения тех работ, которые были рекомендованы научным руководителем и которые близки к выбранной тематике магистерской диссертации.

Знакомиться с литературой целесообразно в следующей последовательности: руководящие документы (вначале законы, затем законодательные акты), научные издания (первоначально книги, затем периодические издания), статистические данные.

При этом вначале целесообразно изучить самые свежие публикации, затем – более ранние.

При подборе нормативно-правовых актов целесообразно использовать возможности тематического поиска документов в справочной правовой системе «Гарант», а также в других справочных системах («Консультант», «Кодекс» и др.). Эти справочно-информационные системы значительно облегчают тематический поиск необходимых нормативных документов.

Статистический и аналитический материал, связанный с протекающими в экономике процессами, можно получить через Интернет. При этом очень важным является умение работать в поисковых системах. Для подбора изданий по интересующей теме могут быть использованы списки литературы, содержащиеся в уже проведенных исследованиях (диссертации на соискание ученых степеней, отчеты по НИР и т.д.).

Одним из основных аспектов работы с литературой является определение главного в данном источнике, не читая его целиком. В издании изучается: заглавие; фамилия автора; наименование издательства (или учреждения, выпустившего книгу); время издания; аннотация; оглавление; введение или предисловие; справочно-библиографический аппарат (список литературы, указатели и т.д.), иллюстративный материал.

При изучении заглавия работы следует сопоставить его с темой магистерской диссертации. Если заглавие совпадает с темой или уже ее, то для Вас будет представлять интерес весь материал публикации. Если заглавие шире темы, то Вам будет интересна только часть издания.

Известность автора важна для исследователя. Если автор широко известен, то работа будет содержать устоявшиеся положения. Если автор малоизвестен, то к материалу следует подходить с критических позиций и обращать внимание на доказательство приводимых положений. Работы малоизвестных авторов не менее важны, чем публикации известных

ученых, поскольку малоизвестные авторы – это начинающие исследователи. Именно они выдвигают и доказывают новые положения.

Контрольные вопросы

1. Актуальность научного исследования. Необходима ли актуальность для фундаментального исследования?
2. Объект и предмет научного исследования. Каков практический и теоретический смысл различения объекта и предмета?
3. Проблема и тема научного исследования. Целесообразно ли изменять тему по мере исследования?
4. Формулировка цели научного исследования. Каково соотношение абстрактной и конкретной цели?
5. Задачи научного исследования. Как они соотносятся с логикой исследования?

Практическая работа №2

В данной практической работе студент должен составить отчет, содержащий:

1. Формулировку проблемы (гипотезы) исследования.
2. Постановку целей и задач (задачи можно сначала укрупненно) магистерской диссертации.
3. Определение объекта и предмета магистерской диссертации.
4. Определение общих методов, с помощью которых будет выполняться работа.
5. Определение специальных методов, с помощью которых будет выполняться работа.

Проблема (гипотеза) исследования – это область неизвестного, но востребованного в научном знании. Грамотно сформулированная проблема – это указание на противоречие, образовавшееся (выявленное) в изучаемой области, на знание, которого еще нет, но которое обязательно необходимо получить, чтобы разрешить обозначенное противоречие.

Цель исследования определяет, для чего проводится исследование, что планируется получить в результате. Цель работы ориентирует на анализ и решение проблемы в двух основных направлениях – теоретическом и прикладном.

Задачи исследования – это алгоритм достижения цели исследования. Это ступеньки, на каждой из которых производится та или иная исследовательская операция (изучение необходимой литературы, сбор эмпирических данных, их анализ, сопоставление: построение классификаций разработка методик и их реализация и т.д.).

Задачи исследования могут быть условно разделены на основные и дополнительные. Основные предполагают поиск ответа на его центральный вопрос: каковы пути и средства решения исследуемой проблемы? Дополнительные задачи помогают выяснить сопутствующие главной проблеме исследования обстоятельства, факторы, причины. Нельзя допускать перемешивания целей и задач, основных и неосновных задач.

Объект исследования – это то, на что направлен процесс познания.

Предмет исследования – это наиболее значимые с теоретической или практической точки зрения свойства, стороны, проявления, особенности объекта, которые подлежат непосредственному изучению в рамках намечающегося исследования. Это угол зрения на объект, аспект его рассмотрения, дающий представление о том, что конкретно будет изучаться в объекте, как он будет рассматриваться, какие новые отношения, свойства, функции будут выявляться.

Далее необходимо выбрать из всех возможных общих методов научного познания методы, которые можно использовать для написания работы по определенной теме.

Например, при раскрытии терминологии объекта исследования предполагается использование диалектического, исторического подхода; при рассмотрении феномена (или свойств, характеристик и др.) объекта будет использован системный и компаративистский методы.

Затем определить специальные методы, с помощью которых будет выполняться работа.

Например, при проведении анализа финансового состояния предприятия будут использованы современные методики сбора и обработки данных, методы финансового анализа, а также комплекс современных методов сбора и обработки исходной информации: экономико-статистический, сравнительный анализ; при обосновании результатов исследования: системный анализ, прогнозирования и прогнозных сценариев для получения и обоснования полученных результатов.

Контрольные вопросы

1. Понятие и признаки новизны научного исследования.
2. Критерии новизны эмпирических исследований.
3. Разработка новых методов и методик осуществления эмпирических исследований.
4. Критерии новизны теоретических исследований.
5. Разработка новых методов и методик осуществления теоретических исследований.
6. Критерии новизны прикладных исследований.
7. Выработка прогнозов развития определенных отраслей экономической деятельности.

Практическая работа №3 и 4

В соответствии с основными структурными компонентами научных исследований представить теоретический отчет по выбранной теме исследования. Подготовить статью для публикации.

Носителями информации могут быть различные документы:

- книги (учебники, учебные пособия, монографии);
- периодические издания (журналы, бюллетени, труды институтов, научные сборники);
- нормативные документы (стандарты, СНИПы, ТУ, инструкции, временные указания, нормативные таблицы и др.);
- каталоги и прейскуранты;
- патентная документация (патенты, изобретения);
- отчеты о научно-исследовательских и опытно-конструкторских работах;
- информационные издания (сборники НТИ, аналитические обзоры,
- информационные листки, экспресс-информация, выставочные проспекты и др.);
- переводы иностранной научно-технической литературы;
- материалы научно-технических и производственных совещаний;
- диссертации, авторефераты;
- производственно-техническая документация организаций (отчеты, акты приемки работ и др.);
- вторичные документы (реферативные обзоры, библиографические каталоги, реферативные журналы и др.).

Эти документы создают огромные информационные потоки, темпы которых ежегодно возрастают.

Работа с источниками НТИ

Каждое научное исследование после выбора темы начинают с тщательного изучения научно-технической информации.

Цель поиска, проработки, анализа информации — всестороннее освещение состояния вопроса по теме, уточнение ее (если это необходимо), обоснование цели и задач научного исследования. Следует уделить внимание изучению различных литературных источников как в оригинале, так и по переводным изданиям. Анализ иностранной информации позволит исключить дублирование по исследуемой теме. Это требует от научного работника знания одного или двух иностранных языков (предпочтительны английский, немецкий, французский).

Без личного ознакомления с оригиналом или квалифицированным переводом базироваться на литературном анализе иностранной информации других авторов не рекомендуется, поскольку каждый автор прорабатывает литературу применительно к своей

теме исследования. Кроме непосредственно относящейся к теме информации, необходимо проработать основную литературу по родственным специальностям.

Очень важно ознакомиться с циклом дисциплин, близких к теме, анализ которых может быть полезен при разработке отдельных вопросов темы.

Для всестороннего анализа информационного материала необходимо ознакомиться с тематикой научных исследований, которые проводятся в различных научно-исследовательских и организациях и университетах. Прорабатывая архивный материал этих организаций, нужно делать записи лишь необходимого по теме материала с указанием номера отчета, года, темы, исполнителей.

На стадии сбора и анализа информации полезны командировки в проектные учреждения, особенно на крупные передовые предприятия. Такие командировки позволяют выяснить, в какой степени исследуемая тема решается на производстве, на какие стороны темы следует обратить особое внимание, какие вопросы представляют первоочередной практический интерес. Желательно иметь мнение производственных коллективов по теме научного исследования.

После сбора литературных, архивных, производственных и других информационных данных и их обобщения полезно узнать мнение крупных ученых. Они могут оказать существенную помощь в разработке темы и определении объема собираемой информации.

Таким образом, научный работник, прорабатывая тему, накапливает большое количество различной информации. В зависимости от наименования и научной значимости темы объем информации может достигать 100—200 наименований и более.

Для эффективного анализа этой информации необходимо знать методы ее учета, проработки и анализа.

Прорабатывая архивный материал этих организаций, нужно делать записи лишь необходимого по теме материала с указанием номера отчета, года, темы, исполнителей.

Учет проработанной информации сводится к составлению библиографии. Библиография — это перечень различных информационных документов с указанием следующих определенных данных: фамилия и инициалы автора, название источника, место издания, издательство, год издания, объем источника в страницах.

Библиографический перечень составляют в алфавитном порядке по фамилиям авторов (для ускорения поиска нужной информации).

Научные тексты — главный источник исследовательской работы. Выбор литературы для чтения и изучения — важная составляющая труда исследователя.

Выбор и изучение литературы осуществляется поэтапно:

- Первичное ознакомление и беглое прочтение источника.
- Глубокое чтение и анализ.

Специальная литература для чтения и изучения отбирается: во-первых, по ключевым понятиям, составляющим тему исследования; во-вторых, по рекомендации научного руководителя; в-третьих, из имеющихся «под рукой» источников.

Чтение научной литературы должно сопровождаться работой со словарями, учебниками, записями лекций. Это помогает адекватно понимать научную терминологию, актуализировать знания и полнее их использовать.

При выборе книги или статьи для чтения целесообразно установить степень сложности источника. Это определяется по количеству непонятных, малознакомых и незнакомых терминов, по наличию неясных положений и утверждений, по сложной конструкции предложений. Определив степень сложности источника, можно более рационально спланировать изучение источников, начав с более лёгких для понимания, постепенно переходя к более трудным. Последними являются, как правило, теоретические тексты, менее сложными являются методические (эмпирические, описательные).

Целесообразно начать чтение научной литературы с источника, в котором интересующая вас проблема представлена более широко или даже целиком. Вы получите общее представление о теме и вопросах, её касающихся. Таким источником может быть даже учебник или учебное пособие.

Вторичное чтение литературы – чтение более медленное, продуманное, глубокое, с обязательным конспектированием, целенаправленное и ведущееся по плану, составленному в соответствии с задачами исследования и планом написания работы.

Выбор источников диктуется целью и планом работы, но при этом не следует забывать о целесообразности разнообразия источников.

Контрольные вопросы

1. Формулирование и обоснование результатов исследования.
2. Требования к использованию литературы в исследовании.
3. Особенности научного стиля речи.
4. Правила оформления исследовательской работы.
5. Формулирование положений, выносимых на защиту.
6. Оформление результатов научных и прикладных исследований.

Список используемой литературы

1. Липчиу Н.В. Методология научного исследования: учебное пособие / Н.В. Липчиу, К.И. Липчиу. – Краснодар: КубГАУ, 2013. – 290 с.
2. Методология и методы научных исследований в экономике и менеджменте: пособие для вузов / Завьялова Н.Б., Головина А.Н., Завьялов Д.В., Дьяконова Л.П., Мельников М.С. и др.; под ред. Н. Б. Завьяловой, А.Н. Головиной – Москва- Екатеринбург:, 2014. – 282 с.
3. Ромашенко Т. Д. Методология научного исследования экономики, управления и финансов : учеб. пособие / Т. Д. Ромашенко, Ю. И. Трещевский, Д. Ю. Трещевский ; под ред. Т. Д. Ромашенко ; Воронежский государственный университет. – 2-е изд., с изм. и доп. – Воронеж : Издательский дом ВГУ, 2018. – 232 с.

МЕТОДОЛОГИЯ НАУЧНОГО ПОЗНАНИЯ:
методические рекомендации
по выполнению практических работ по дисциплине
«Методология научного познания»

В.И. МЕНЬЩИКОВА

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

В. Г. Тронин, А. Р. Сафиуллин

МЕТОДОЛОГИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

Ульяновск
УлГТУ
2020

УДК 001.89(075.8)

ББК 72ся73

Т 73

Рецензенты:

Директор Ульяновского филиала Института радиотехники и электроники имени В.А. Котельникова Российской академии наук, доктор технических наук, доцент В.А. Сергеев

Ведущий инженер-программист ФНПЦ АО «НПО «Марс», кандидат технических наук Ю.А. Радионова

Утверждено редакционно-издательским советом университета
в качестве учебного пособия

Тронин, Вадим Георгиевич

Т 73 **Методология научных исследований : учебное пособие /**
В.Г. Тронин, А.Р. Сафиуллин. – Ульяновск : УлГТУ, 2020. – 86 с.

ISBN 978-5-9795-2046-9

Рассмотрены общие вопросы методологии диссертационного исследования. Методология научного исследования показана с точки зрения применения системного подхода и теории решения изобретательских задач. Пояснены свойства систем и кривая развития технической системы с выявлением общих принципов эволюции. Приведены законы развития технических систем, применяемые на различных этапах развития. Описана последовательность формулирования исследовательской задачи и уровни изобретений. Учебное пособие содержит примеры к материалу на различных технических системах, контрольные вопросы и задания для практического применения материала в исследованиях аспирантов.

Пособие предназначено для аспирантов первого года обучения по дисциплине «Методология научных исследований» направления 09.06.01 «Информатика и вычислительная техника» и других специальностей.

Работа подготовлена на кафедре «Информационные системы» Ульяновского государственного технического университета.

УДК 001.89(075.8)

ББК 72ся73

ISBN 978-5-9795-2046-9

© Тронин В. Г., Сафиуллин А.Р., 2020

© Оформление. УлГТУ, 2020

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	5
1. ОБЩИЕ ВОПРОСЫ ОРГАНИЗАЦИИ И МЕТОДОЛОГИИ НАУЧНОГО ИССЛЕДОВАНИЯ В АСПИРАНТУРЕ	7
1.1. Нормативно-правовые основы научно-исследовательской деятельности и подготовки диссертации в аспирантуре	7
1.2. Требования к диссертации на соискание ученой степени	11
1.3. Основные категории в методологии научных исследований	13
2. ЗАКОНЫ РАЗВИТИЯ ТЕХНИЧЕСКИХ СИСТЕМ	24
2.1. Системный подход	24
2.2. Кривая развития системы	25
2.3. Матрица бостонской консультационной группы	32
2.4. Законы статики: полноты частей системы, «энергетической проводимости», согласования ритмики частей системы	33
2.5. Законы кинематики: увеличения степени идеальности системы, неравномерности развития частей системы, перехода в надсистему	39
2.6. Законы динамики: перехода с макроуровня на микроуровень, перехода к более управляемым ресурсам	44
2.7. Частные случаи законов: самосборки, повышения свернутости системы, вытеснения человека	51
3. УРОВНИ ИЗОБРЕТАТЕЛЬСКИХ ЗАДАЧ	59
3.1. Главная полезная функция системы	59
3.2. Идеальный конечный результат	61
3.3. Типы противоречий: административное, техническое, физическое	62
3.4. 1-й уровень изобретательских задач	67
3.5. 2-й уровень изобретательских задач	68
3.6. 3-й уровень изобретательских задач	68
3.7. 4-й уровень изобретательских задач	69
3.8. 5-й уровень изобретательских задач	70

3.9. Изменение уровня изобретений и их экономической эффективности на разных этапах развития системы	71
ЗАКЛЮЧЕНИЕ	74
ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ	75
ЗАДАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ	81
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	83

ПРЕДИСЛОВИЕ

Человеческая деятельность может быть направлена на воспроизведение уже существующего опыта – репродуктивная деятельность, либо на создание чего-либо нового – продуктивная деятельность. При этом уровень новизны может быть как мирового уровня, так и субъективно новым для конкретного человека, сообщества.

Научное исследование претендует на получение уникального результата, расширение возможностей человечества в некотором узком направлении. Чтобы обнаружить то, чего никто ранее не обнаружил, вообразить и реализовать то, до чего никто ранее не додумался, требуется:

- особое мышление: нестандартное, латеральное, изобретательское, творческое;

- методология, позволяющая упорядочить, структурировать исследование и искать (создавать) новое целенаправленно, продуктивно, максимально используя уже созданные человечеством эффективные инструменты (математический аппарат и т. д.).

Фантаст Станислав Лем еще в 1963 год прогнозировал, что с развитием науки и огромным числом публикаций в будущем, потребуется оставлять только те исследования, которые действительно приносят нечто новое. Иначе в потоке работ, не приносящих ничего нового в мир, найти действительно ценное будет практически невозможно [1]. В соответствии с прогнозом в XXI веке наблюдаем экспоненциальный рост числа публикаций, что приводит к развитию наукометрии, как попытки структурировать разрастающиеся объемы публикаций. Вопросы продвижения результатов исследований, оценки научных публикаций для аспирантов рассмотрены в учебных пособиях [2, 3].

Для аспиранта важно понимание как последовательности выполнения исследования, ведущей к результату, так и понимание методологии проведения научного исследования. Поэтому мы рассмотрим нормативную составляющую подготовки аспиранта и формальные требования, основные определения, касающиеся методологии научного исследования. Изучим теорию развития технических систем и некоторые другие инструменты ТРИЗ, позволяющие аспиранту понять место своего исследования в эволюции систем, более четко сформулировать требования к нему и приблизиться к сильным решениям имеющихся проблем в улучшаемой, модернизируемой или создаваемой технической системе.

В середине XX века в СССР Генрихом Сауловичем Альтшуллером разработана теория решения изобретательских задач (ТРИЗ), которая позволяет улучшать любые системы (в первую очередь технические), основываясь на общих закономерностях эволюции технических систем, которые были выявлены на основе анализа тысяч патентов и изобретений. С тех пор ТРИЗ применяют в конструкторских бюро, отделах R&D большинства компаний мирового уровня: Intel, Samsung и т. д. Применение инструментов ТРИЗ в качестве методологии научного исследования рассмотрим в совокупности с системным подходом, функционально-стоимостным анализом и рядом других инструментов.

1. ОБЩИЕ ВОПРОСЫ ОРГАНИЗАЦИИ И МЕТОДОЛОГИИ НАУЧНОГО ИССЛЕДОВАНИЯ В АСПИРАНТУРЕ

1.1. Нормативно-правовые основы научно-исследовательской деятельности и подготовки диссертации в аспирантуре

Научно-исследовательская деятельность и подготовка научно-квалификационной работы, а именно диссертации на соискание ученой степени кандидата наук, являются одним из важнейших элементов подготовки аспиранта. В соответствии со статьей 2 Федерального закона от 23.08.1996 №127-ФЗ (ред. от 31.07.2020) «О науке и государственной научно-технической политике» [4] научная деятельность направлена на получение и применение новых знаний для решения технологических, инженерных, экономических, социальных, гуманитарных и иных проблем, обеспечения функционирования науки, техники и производства как единой системы. Научно-исследовательская деятельность включает:

– фундаментальные научные исследования, направленные на получение новых знаний об основных закономерностях строения, функционирования и развития человека, общества, окружающей среды;

– прикладные научные исследования, направленные преимущественно на применение новых знаний для достижения практических целей и решения конкретных задач;

– поисковые научные исследования, направленные на получение новых знаний в целях их последующего практического применения (ориентированные научные исследования) и (или) на применение новых знаний и проводимые путем выполнения научно-исследовательских работ.

Заслушивание и оценка научного доклада об основных результатах научного исследования и подготовленной диссертации, которая должна быть оформлена в соответствии с установленными требованиями, является частью Государственной итоговой аттестации аспиранта, которая завершается присвоением квалификации «Исследователь. Преподаватель-исследователь» и выдачей диплома об окончании аспирантуры (рис. 1.1).

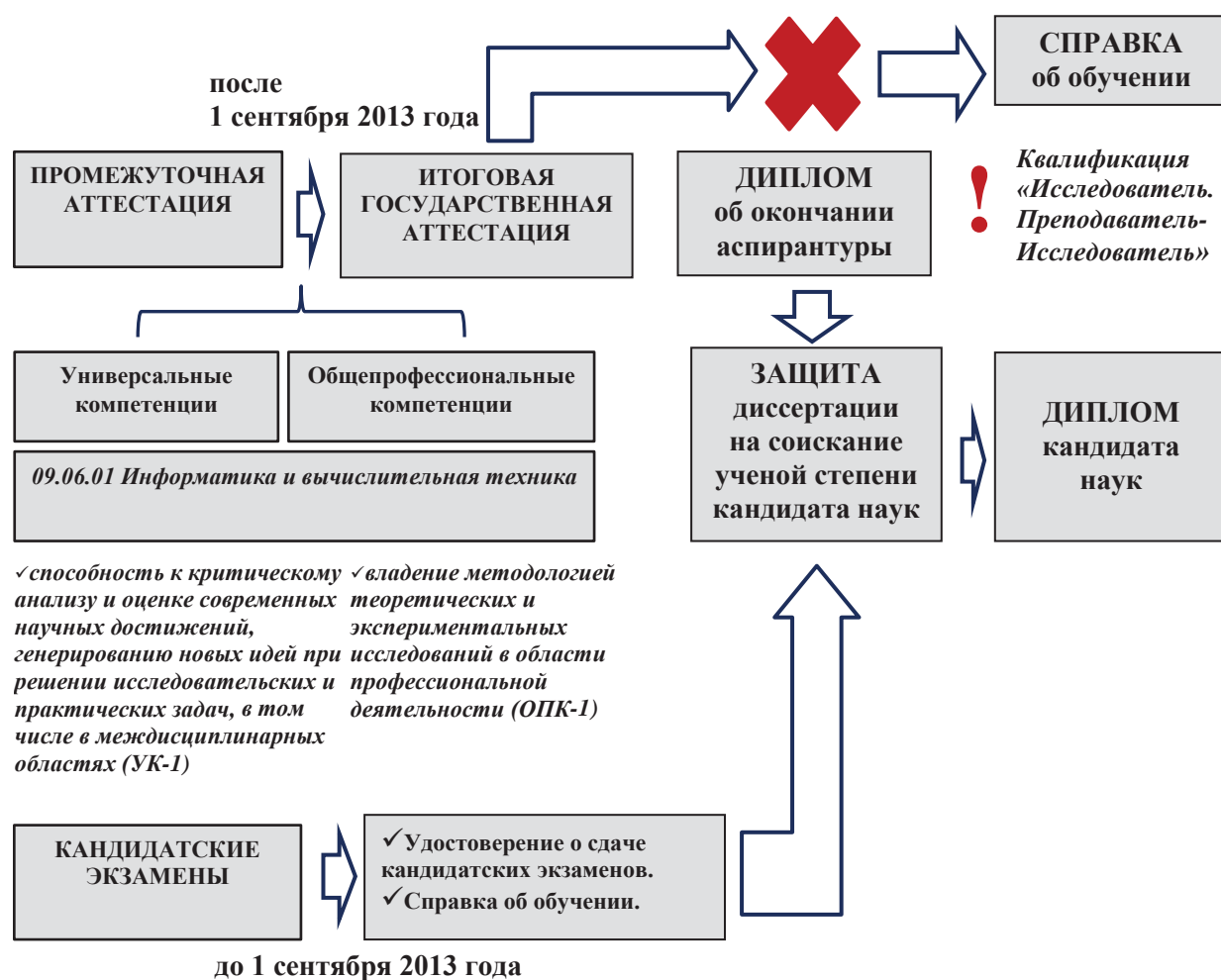


Рис. 1.1. Общая характеристика освоения программы аспирантуры.

Данные требования установлены федеральными государственными образовательными стандартами высшего образования по уровню подготовки кадров высшей квалификации (далее - ФГОС ВО). В соответствии с нормами ФГОС ВО выпускник аспирантуры должен быть подготовлен к двум видам

профессиональной деятельности: научно-исследовательская деятельность в соответствующей области и преподавательская деятельность по образовательным программам высшего образования. Основным результатом освоения программы аспирантуры в соответствии с ФГОС ВО является формирование у выпускника набора исследовательских компетенций в определенной профессиональной области, в том числе [5]:

- критический анализ и оценка научных достижений, результатов исследований и разработок,
- генерирование новых научных идей,
- способность проектировать и осуществлять научные исследования,
- владение методологией и культурой научного исследования,
- способность к разработке новых методов и их применению в самостоятельной научно-исследовательской деятельности.

Область научного исследования определяет объект и предмет исследования, цель и задачи научного исследования, выбираемые методы научного исследования. В некоторых исследованиях в применении терминов «объект» и «предмет исследования» нет различий, так как слово «объект» произошло от латинского *objectum* – предмет [6]. В других исследованиях считается, что объектом исследования может быть только явление и процесс, который «существует независимо от нашего сознания», а предметом – те аспекты, на которые направлена познавательная деятельность.

Содержание научного исследования при освоении программы аспирантуры, как правило, определяется паспортом соответствующей научной специальности [7], по которой аспирант подготавливает диссертацию на соискание ученой степени кандидата наук. В таблице 1.1 представлены области научных исследований, соответствующие профилям подготовки по направлению 09.06.01 Информатика и

вычислительная техника в Ульяновском государственном техническом университете.

Таблица 1.1

Шифр и наименование научной специальности [8]	Область исследования
05.13.05 Элементы и устройства вычислительной техники и систем управления	<ol style="list-style-type: none"> 1. Разработка научных основ создания и исследования общих свойств и принципов функционирования элементов, схем и устройств вычислительной техники (ВТ) и систем управления. 2. Теоретический анализ и экспериментальное исследование функционирования элементов и устройств ВТ и систем управления в нормальных и специальных условиях с целью улучшения технико-экономических и эксплуатационных характеристик. 3. Разработка принципиально новых методов анализа и синтеза элементов и устройств ВТ и систем управления с целью улучшения их технических характеристик. 4. Разработка научных подходов, методов, алгоритмов и программ, обеспечивающих надежность, контроль и диагностику функционирования элементов и устройств ВТ и систем управления
05.13.12 Системы автоматизации проектирования (промышленность)	<ol style="list-style-type: none"> 1. Методология автоматизированного проектирования в технике, включая постановку, формализацию и типизацию проектных процедур и процессов проектирования, вопросы выбора методов и средств для применения в САПР. 2. Разработка научных основ создания САПР и автоматизации технологической подготовки производства (АСТПП). 3. Разработка научных основ построения средств САПР, разработка и исследование моделей, алгоритмов и методов для синтеза и анализа проектных решений, включая конструкторские и технологические решения в САПР и АСТПП. 4. Разработка принципиально новых методов и средств взаимодействия проектировщик – система. 5. Разработка научных основ обучения автоматизированному проектированию. 6. Разработка научных основ реализации жизненного цикла проектирование – производство – эксплуатация, построения интегрированных средств управления проектными работами и унификации прикладных протоколов информационной поддержки. 7. Разработка научных основ построения средств автоматизации документирования, безбумажного документооборота, процессов работы электронных архивов технической документации, взаимодействия с изготовителем и потребителем изделий. 8. Разработка научных основ построения средств компьютерной графики, методов геометрического моделирования проектируемых объектов и синтеза виртуальной реальности

Шифр и наименование научной специальности [8]	Область исследования
05.13.18 Математическое моделирование, численные методы и комплексы программ	<ol style="list-style-type: none"> 1. Разработка новых математических методов моделирования объектов и явлений. 2. Развитие качественных и приближенных аналитических методов исследования математических моделей. 3. Разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий. 4. Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента. 5. Комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента. 6. Разработка новых математических методов и алгоритмов проверки адекватности математических моделей объектов на основе данных натурного эксперимента. 7. Разработка новых математических методов и алгоритмов интерпретации натурного эксперимента на основе его математической модели. 8. Разработка систем компьютерного и имитационного моделирования

1.2. Требования к диссертации на соискание ученой степени

Основные положения, регламентирующие требования к присуждению ученых степеней, изложены в Положении о присуждении ученых степеней (далее - Положение), утвержденном Постановлением Правительства РФ от 24.09.2013 №842 [9].

В соответствии с данным Положением диссертация на соискание ученой степени кандидата наук должна соответствовать ряду критериев.

1. Должна содержать решение научной задачи, которая имеет значение для развития соответствующей отрасли знаний. Либо в диссертации должны быть изложены новые научно обоснованные

технические, технологические или иные решения и разработки, имеющие существенное значение для развития страны.

2. Должна быть написана автором самостоятельно, обладать внутренним единством, содержать новые научные результаты и положения, выдвигаемые для публичной защиты, и свидетельствовать о личном вкладе автора диссертации в науку.

В диссертации, имеющей прикладной характер, должны приводиться сведения о практическом использовании полученных автором диссертации научных результатов, а в диссертации, имеющей теоретический характер, – рекомендации по использованию научных выводов. Предложенные автором диссертации решения должны быть аргументированы и оценены по сравнению с другими известными решениями.

3. Основные научные результаты диссертации должны быть опубликованы в рецензируемых научных изданиях.

Количество публикаций, в которых излагаются основные научные результаты диссертации, в рецензируемых изданиях должно быть: в области искусствоведения и культурологии, социально-экономических, общественных и гуманитарных наук – не менее 3; в остальных областях – не менее 2.

К таким публикациям приравниваются патенты на изобретения, патенты (свидетельства) на полезную модель, патенты на промышленный образец, патенты на селекционные достижения, свидетельства на программу для электронных вычислительных машин, базу данных, топологию интегральных микросхем, зарегистрированные в установленном порядке.

4. В диссертации соискатель ученой степени обязан ссылаться на автора и (или) источник заимствования материалов или отдельных результатов.

Диссертацию на соискание ученой степени кандидата наук представляют в виде специально подготовленной рукописи, оформленной в соответствии с требованиями ГОСТ [10].

1.3. Основные категории в методологии научных исследований

Понятие о методе и методологии научного исследования.

В переводе с греческого: «methodos» – «способ, путь следования, прием осуществления той или иной деятельности», «logos» – «слово, смысл, суждение, учение». Следовательно, в буквальном смысле методология – это учение об основных методах какой-либо науки. В самом обобщенном виде методология есть совокупность общих мировоззренческих положений и принципов, которые обуславливают позицию исследователя или группы исследователей и являются обоснованием для построения и организации познавательной деятельности в научных исследованиях. Методология представляет собой учение о познавательных процедурах и результатах данных процедур, учение о структуре и логической организации научного исследования.

Необходимо различать метод и методику. Метод – это подход к изучаемому материалу, его систематизация и теоретическое осмысление (теория), обусловленный особенностями объекта и предмета исследования. Методика же в большей степени ориентирована на совокупность приемов наблюдения, эксперимента и описания, представляет собой некоторый свод правил или алгоритм каких-либо действий при изучении какого-либо явления.

Методология практически любого научного исследований в своей основе имеет всеобщие философские принципы (табл. 1.2).

Таблица 1.2

Принцип	Содержание
Принцип познаваемости мира	Позволяет исследователю, приступая к изучению, быть уверенным, что данный исследуемый феномен познаваем, а непознаваемых процессов не существует
Принцип развития	Диалектический подход в исследовании явлений означает рассмотрения их как развивающихся в пространстве и времени
Принцип практической осуществимости	Возможность управления процессами и явлениями, способность целенаправленно воздействовать на них. Все, что теоретически возможно и не запрещено законами природы и общества, то практически осуществимо
Принцип детерминизма	Признание, что за каждым следствием скрывается своя причина и что все процессы в мире так или иначе связаны между собой
Принцип простоты (принцип «экономии мышления»)	«Бритва Оккама» (XIV в.): «Не умножай сущностей сверх необходимого», то есть стремление искать в отношении всякого факта наиболее простое объяснение из всех существующих вариантов

Метод научного исследования всегда отличает строгость, однозначность, эффективность, простота (экономность) и эвристичность, то есть способность приносить новые научные результаты. Ненаучные методы, напротив, имеют в своей основе отсылку к авторитетам или веру в авторитеты (религиозное или мистическое знание), некое «идеологическое» или же обыденное знание.

Методы научного исследования в современной науке.

Целесообразно разделить методы научного исследования на несколько групп [11]:

1. Всеобщие методы (применяются не только в научных исследованиях, но и в других отраслях человеческой деятельности) (табл. 1.3).

Таблица 1.3

Метод	Содержание	Пример
Анализ	Мысленное или реальное разложение объекта исследования на составные части	– статистический анализ научных публикаций; – анализ контента социальных сетей;
Синтез	Соединение отдельных частей объекта исследования в единое целое	– синтез структуры информационной системы; – синтез композиционных материалов;
Абстрагирование	Мысленное отвлечение от ряда свойств и отношений объекта исследования, которые являются несущественными в рамках данного исследования, с одновременным выделением существенных свойств, интересующих исследователя в данный момент	– класс как разновидность абстрактного типа данных в объектно-ориентированном программировании, классы объектов являются «шаблонами», определяющими наборы свойств, методов и событий, по которым создаются объекты;
Обобщение	Сумма нечетных простых чисел есть всегда четное число	– теория информации, теория автоматов, теория алгоритмов, теория графов, теория множеств, теория чисел и теория игр – это теоретические элементы базиса компьютерных наук; – радиоволны, тепловое излучение, видимый свет, ультрафиолетовое и рентгеновское излучение – это электромагнитные волны;
Индукция	Мысленное движение познания от фактов и отдельных случаев к общему положению, от частного к общему	– Д. И. Менделеев, используя частные факты о химических элементах, сформулировал периодический закон; – наблюдая некоторые числа и их комбинации, мы сталкиваемся с фактом, что сумма нечетных простых чисел есть всегда четное число;
Дедукция	Мысленное движение познания от общих утверждений к отдельным явлениям, выведение частного из общего; термин дедукция» в переводе с латинского означает вывод	– закономерность: все объекты в этом массиве точечные; ситуация: эти объекты – из этого массива; результат: эти объекты точечные; – античный силлогизм: Все люди смертны. Кай – человек. Следовательно, Кай смертен;

Окончание табл.1.3

Метод	Содержание	Пример
Аналогия	Получение знание об одних предметах или явлениях на основании их сходства с другими; рассуждение, в котором из сходства изучаемых объектов в некоторых признаках делается заключение об их сходстве и в других признаках	– электрический ток – поток воды; – аналог обыкновенного вируса в памяти компьютера является компьютерный вирус - вирус информационный, не имеющий физического тела; – лист преобразует световую энергию солнечного луча в энергию химических реакций, а также в электрическую энергию внутриклеточных процессов, аналогия - фотоэлемент преобразует световую энергию в электрическую
Моделирование	Получение знаний об объекте исследования через воспроизведение его характеристик на его заместителе – модели (мысленно представляемый или материально существующий аналог объекта)	– в зависимости от средств построения: натурные модели (макет Солнечной системы) и абстрактные (знаковые) модели (математические модели явления и компьютерные модели); – в зависимости от области применения: учебные (тренажеры, обучающие программы), опытные (модели корабля, машины), научно-технические (прибор, имитирующий разряд молнии), игровые (деловые, военные, спортивные игры), имитационные (испытание нового лекарства на мышах)
Классификация	Распределение тех или иных объектов по классам (отделам, разрядам) в зависимости от их общих признаков	– номенклатура научных специальностей, по которым присуждаются ученые степени; – классификация компьютерных наук (computer science): computer engineering и software engineering

2. Общенаучные методы (используются во всех разделах и областях науки), которые можно разделить на две группы.

Эмпирические методы:

– наблюдение – метод исследования, основанный на непосредственном и целенаправленном восприятии свойств объекта исследования при помощи органов чувств (ощущения, восприятия,

представления), наблюдение как научный метод отличается, как правило, применением специальных технических средств, важным приемом наблюдения является интерпретация его результатов (расшифровка показаний приборов, кривой на электрокардиограмме и тому подобное), обычно наблюдение включается в качестве составной части в процедуру эксперимента;

- описание – фиксация признаков объекта исследования, которые устанавливаются, например, путем наблюдения или измерения;

- измерение – определение численного значения некоторой величины путем сравнения ее с эталоном, единицей измерения;

- эксперимент – метод исследования объекта в контролируемых и управляемых условиях, искусственное воспроизведение исследуемого явления или процесса в заданных условиях; всякий научный эксперимент направляется какой-то гипотезой, идеей.

Теоретические методы:

- аксиоматический метод – метод исследования и способ построения научных теорий, основу которого составляет некое множество утверждений (аксиомы или постулаты), которое принимается без доказательств, и из которого выводится остальное знание (доказуемые положения или теоремы) по определенным логическим правилам (например, геометрия Эвклида);

- идеализация – особая разновидность абстрагирования, которая представляет собой переход к конструированию не существующих в реальном мире объектов, но при этом для которых есть прообразы в реальном мире (например, в физике – идеальный газ, а в экономике – *Homo economicus*, или экономический человек);

- формализация – перевод знания об объекте исследования в знаково-символическую форму с помощью языков, предназначенных

для точного и строгого выражения знания (язык математики, логики, химии и так далее);

– системный метод – метод исследования объекта как системы, то есть определенной совокупности материальных элементов или идеальных представлений, связей между ее компонентами, связей между внутренней средой системы и внешней средой;

– гипотетико-дедуктивный метод.

К методам теоретического уровня также можно отнести методы абстрагирования, обобщения, методы системного анализа, моделирование.

3. Конкретно-научные (частные или специальные) методы (применяются в отдельных разделах или отраслях науки).

Уровни научного исследования.

С позиции теории познания можно выделить два уровня научного исследования: теоретическое и эмпирическое [12].

Для теоретического уровня исследования характерно применение преимущественно логических методов познания. На данном уровне исследуемые факты и объекты анализируются и обобщаются для постижения их сущности, выявления внутренних связей и законов развития.

Теоретический уровень научного исследования связан с такими категориями, как проблема, гипотеза и теория.

Проблема – это теоретическая или практическая задача, решение которой неизвестно или известно не полностью, это «сложноорганизованная система исследовательских задач принципиального характера, обладающих существенной неопределенностью» [13].

Гипотеза – это предположение в области научного исследования, которое требует проверки и доказывания, это

допущение, истинное значение которого неопределенно. Научная гипотеза должна удовлетворять определенным требованиям [14]:

- должна формулироваться по отношению к определенной предметной области, ей следует быть объяснением некоторых наличествующих явлений и процессов, которым она должна соответствовать;

- «должна предлагать ответ на проблему, изначально ее породившую» [15];

- должна быть оптимально сформулирована на языке определенной науки;

- ей следует быть логически корректной, что обеспечивает ее доступность к восприятию и применению другими учеными;

- ей необходимо на некотором отрезке времени существования науки быть пригодной для объяснения определенных классов явлений и событий;

- она должна быть обоснованной, то есть положения, составляющие суть гипотезы, выдвинуты не случайно, а вытекают из анализа и синтеза данных науки, результатом обоснованности гипотезы выступает ее непротиворечивость и информативная направленность;

- должна быть доступной проверке (критерии принципиальной проверяемости, совместимости с фактическим материалом и с утвердившимися теоретическими положениями и принципиальной приложимости);

- гипотеза должна стимулировать дальнейшее изучение той предметной области, для объяснения которой она создавалась (многие гипотезы выдающихся физиков XX века, например, гипотеза Марии Кюри-Склодовской о радиоактивности урана, гипотеза Макса Планка о дискретном характере процессов излучения и поглощения направляли других исследователей не только на изучение свойств

атомов вещества, но и на выработку новых средств и методов их познания);

– гипотезе необходимо быть открытой для развития с целью преобразования в теорию или опровержения и элиминации из науки.

Теория – это концептуальная система знаний, это «наиболее организованная форма научного знания, дающая целостное представление о закономерностях и существенных связях определенной области действительности» [16]. Понятие «теории» ассоциируется с именем создателя, примерами теории являются классическая механика Ньютона, электромагнитная теория Дж. К. Максвелла [17]. Структуру теории образуют определения, понятия, законы, научные положения и учения, и другие элементы.

В структуре теории можно выделить следующие элементы [18]:

– исходные основания – фундаментальные понятия, принципы (руководящие идеи, которые составляют исходные положения), законы (объективные, существенные, внутренние, необходимые и устойчивые связи между явлениями, процессами), уравнения, аксиомы (исходные положения, которые являются недоказываемыми и из которых по установленным правилам выводятся другие положения);

– идеализированные объекты – абстрактные модели существенных свойств и связей изучаемых предметов;

– логика теории – совокупность определенных правил и способов доказательства, нацеленных на прояснение структуры и изменения знания;

– философские установки и ценностные факторы;

– совокупность законов и утверждений, выведенных в качестве следствий из данной теории в соответствии с конкретными принципами.

Эмпирический уровень исследования характеризуется преобладанием чувственного познания, то есть изучения внешнего мира посредством органов чувств. Данный уровень научного исследования, как правило, связан с практической или опытно-экспериментальной деятельностью исследователя.

Структуру эмпирического уровня исследования составляют факты, эмпирические обобщения и законы (зависимости). Факт – это событие, результат, нечто реальное в противоположность вымышленному; конкретное, единичное в отличие от абстрактного и общего; как форма эмпирического знания факт противопоставляется теории или гипотезе [16]. Одни и те же факты могут быть успешно объяснены несколькими теориями, которые зависят от фактов. Понятие «научный факт» значительно шире и многограннее, чем понятие «факт» [19], применяемое в обыденной жизни. Научные факты характеризуются особыми свойствами – новизной, которая говорит о принципиально новом, не известном до сих пор предмете, явлении или процессе, а также объективностью, точностью и достоверностью. Новизна научного факта не обязательно означает научное открытие, но это новое знание о том, чего мы до сих пор не знали. Система определенных научных фактов позволяет исследователю перейти к эмпирическому обобщению. Регулярность в наблюдаемых явлениях и устойчивость в отношениях между ними есть эмпирические законы, которые на данном уровне исследования еще не являются теоретическим знанием.

Терминология научного исследования.

Каждый термин имеет определенный смысл в заданном контексте, без точного определения базы терминов исследования неизбежны разночтения, недопонимание, ложные выводы.

При выполнении научного исследования, результаты которого будут отражены в тексте диссертации, научной статьи или научного

доклада, рекомендуется применять единую терминологию. Такой подход при выполнении научного исследования:

- формирует научный характер исследования и научный стиль изложения результатов,
- позволяет выстроить логику исследования и систематизировать используемый категориальный аппарат,
- свидетельствует о понимании автором содержательных различий в терминах,
- помогает автору доказательно отстаивать собственную позицию.

Термин есть слово или словосочетание, которое призвано обозначить объект, предмет, явление или понятие о них в пределах некой области исследования (познания). Совокупность терминов, которые используются в определенной науке, составляют ее понятийный аппарат. Основу терминологии научного исследования составляют определения и понятия (рис. 1.2).

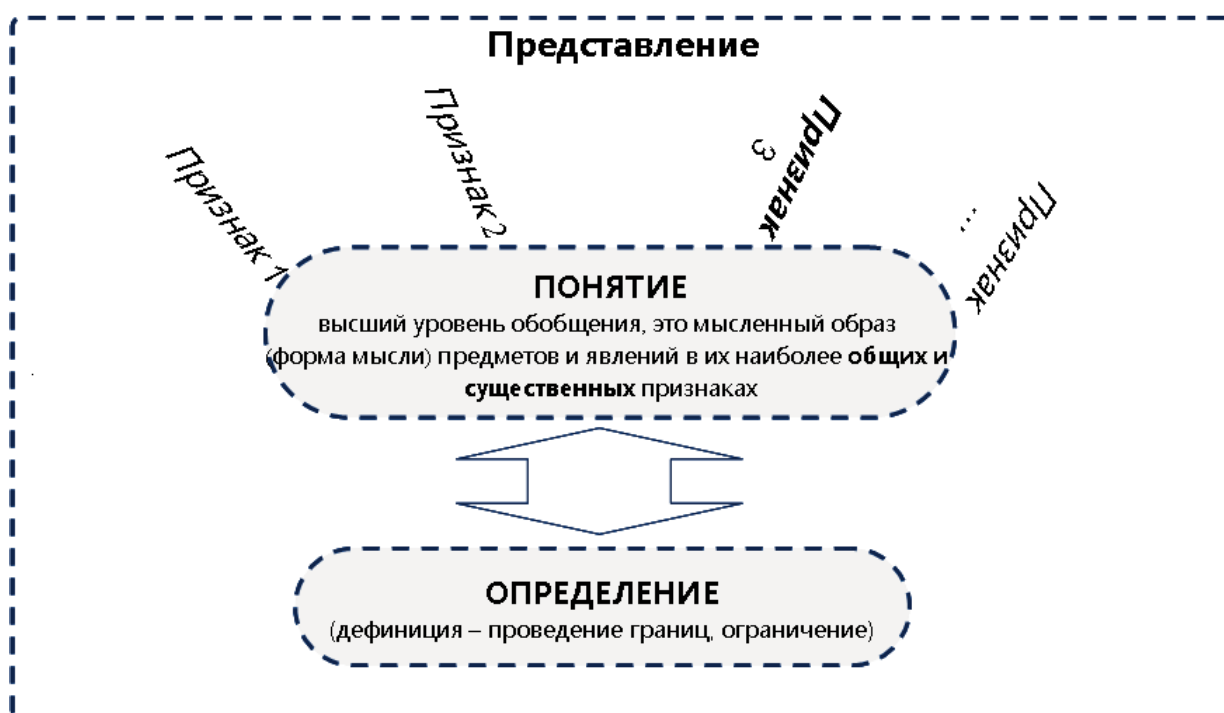


Рис. 1.2. Терминология научного исследования

Раскрыть содержание научного понятия можно через такие приемы, как:

- описание – перечисление внешних черт предмета или явления, включающих как существенные, так и несущественные признаки, позволяющие отличить его от других;

- характеристику – перечисление лишь наиболее важных в том или ином отношении признаков предмета или явления;

- сравнение – указание на сходство предметов, что позволяет лучше понять или по-новому осветить их черты и свойства.

Важным элементом познания в научном исследовании является классификация – группировка объектов исследования в соответствии с их общими признаками. Классификация, как и оценка явления или процесса в научном исследовании, строится на основе определенного критерия (выделяемый исследователем признак как средство для построения соответствующего суждения).

2. ЗАКОНЫ РАЗВИТИЯ ТЕХНИЧЕСКИХ СИСТЕМ

2.1. Системный подход

Для выявления возможности развития технической системы необходимо использовать системный подход, который активно применяется в ТРИЗ. Под системой понимается совокупность взаимосвязанных элементов, обладающая свойствами, не сводящимися к свойствам отдельных элементов, и предназначенная для выполнения определенных функций.

Главное свойство новых систем – появление новой функции. Функциональность – это проявление определенных свойств (функций) при взаимодействии с внешней средой. Назначение системы (цель создания) – достижение определенного желаемого конечного результата.

Вторым важнейшим свойством системы является эмерджентность, достигаемая за счет упорядоченных взаимосвязей и взаимодействий элементов системы и проявляющиеся в возникновении таких свойств, которыми элементы системы не обладают. В ЭВМ основной функцией является проведение вычислений, но отдельные элементы ЭВМ (плата памяти, процессор или блок питания) выполнить такую задачу не могут.

Вкратце перечислим другие свойства систем.

Целостность – каждый элемент системы вносит вклад в реализацию целевой функции системы.

Организованность – заключается в наличии структуры и функционирования (поведения).

Структурность – упорядоченность системы, определенный набор и расположение элементов со связями между ними. Структура соответствует функциям, при изменении функций системы проходит реструктуризация.

Поведение – целенаправленное изменение во времени состояния системы, поведение реализуется самой системой, исходя из собственных целей. Поведение системы имеет задачи сохранения устойчивости и развития системы.

От того, насколько система устойчива и способна противостоять внешним возмущающим воздействиям, зависит продолжительность жизни системы.

Простые системы имеют пассивные формы устойчивости: прочность, сбалансированность, регулируемость, гомеостаз. А для сложных систем определяющими являются активные формы: надежность, живучесть и адаптируемость.

Надежность – система способна сохранять свою структуру при уничтожении отдельных ее элементов благодаря механизму их замены или дублирования. Живучесть системы проявляется в активном подавлении вредных факторов. Адаптируемость заключается в том, что в условиях изменения внешней среды система меняет поведение (структуру) с целью сохранения, улучшения или приобретения новых качеств. Адаптация возможна благодаря наличию обратных связей в системе.

Многие свойства систем мы будем использовать при описании законов развития систем.

Любая реальная система существует во взаимодействии с внешней средой, при этом границы системы следует определять в зависимости от постановки задачи. Например, оператор ЭВМ может рассматриваться как внешняя среда или как часть автоматизированной системы [20].

2.2. Кривая развития системы

Жизненный цикл большинства технических систем может быть изображен в виде S-образной кривой (рис. 2.1), показывающей изменение во времени основных характеристик системы (мощность,

производительность, скорость, тиражи и т. д.). S-образная кривая не отражает существа происходящих в системах изменений – она лишь демонстрирует их результат.

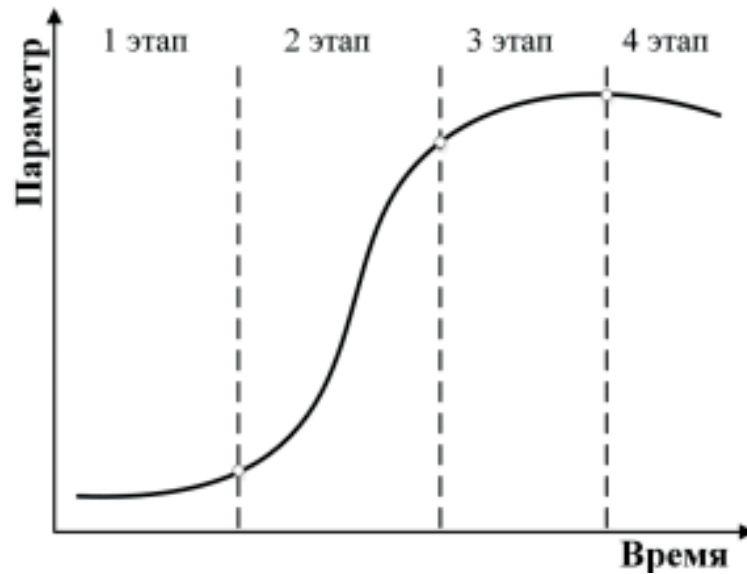


Рис. 2.1. График S-образной кривой развития системы

Конечно, у разных технических систем кривая имеет свои индивидуальные особенности, но обычно отображает характерные этапы (рис.2.2).

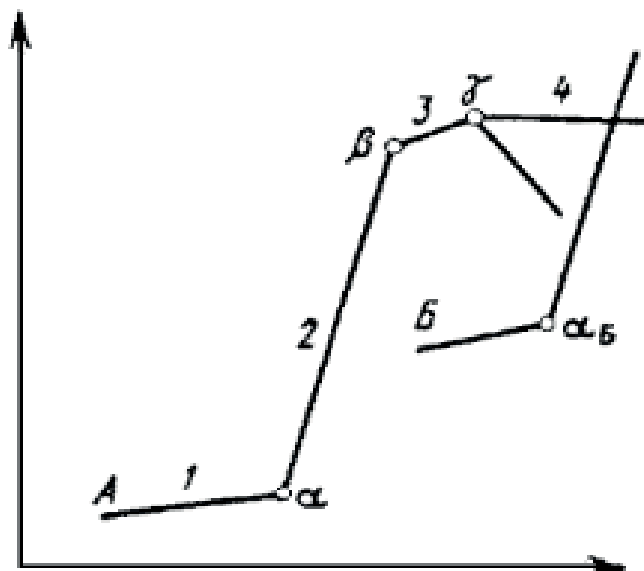


Рис. 2.2. График S-образной кривой развития системы и следующего поколения системы

В «детстве» (этап 1) техническая система развивается медленно. Затем наступает пора «зрелости» (этап 2), на котором техническая система быстро совершенствуется, обеспечивая массовое ее применение. С некоторой точки темпы развития начинают спадать (этап 3), далее наступает пик развития, скат и «старость» системы (этап 4).

Сценарий жизненного цикла не предопределен жестко, техническая система А в пиковой точке может:

- деградировать, уступив место принципиально другой системе Б;
- на долгое время законсервировать достигнутые показатели (в конструкцию велосипеда за XX век внесены непринципиальные улучшения и он продолжает существовать наравне с мопедом, скутером, мотоциклом).

Рассмотрим, какие же факторы определяют длительность каждого этапа и интенсивность роста и спада на «жизненной кривой» технической системы?

Изучение Альтшуллером Г.С. [21] истории развития параметров различных технических систем заставило прийти к выводу, что реальные кривые развития технических систем во многом не похожи на ожидаемые теоретические кривые (рис. 2.3).

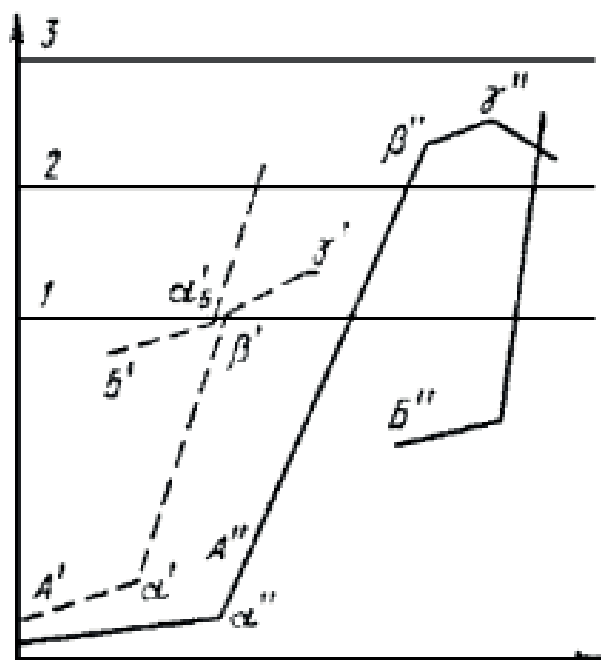


Рис. 2.3. График S-образной кривой развития системы

Логично предположить, что после появления технической системы, основанной на новом техническом принципе (эффекте, конструкции), она должна всесторонне изучаться, модернизироваться с переходом к массовому применению. Для реальных новинок переход к массовому применению начинается с большим временным интервалом и на более низком техническом уровне даже в сравнении с существующими системами.

С чем связано такое расхождение? При быстром развитии признанной учеными (сообществом инженеров, бизнесменов, потенциальных пользователей) технической системы достигается уровень, где исчерпываются возможности модернизации использованного в системе принципа, достигается потолок экономической эффективности (этап 2). Когда кривая А доходит до завершения этапа 2, в дальнейшем развитии системы оказываются заинтересованными многие люди. К этому времени накапливается инерция интересов участников:

- финансистов, ждущих прибыли от инвестиций;

- инженеров, зарабатывающих на обслуживании системы;
- ученых, занимающихся моделированием и улучшением системы;
- создано большое число рабочих мест в производстве и эксплуатации системы;
- привыкли конечные потребители.

Сложившаяся схема устраивает участников, и они не заинтересованы в появлении чего-то нового, даже если это новое потенциально намного лучше. Под интересы участников эксплуатации старой системы подстраивается экономическая модель, и вплоть до завершения этапа 3 система продолжает оставаться рентабельной за счет выноса отрицательной стоимости за пределы системы в виде разрушения, загрязнения внешней среды (природной, человеческой).

Приведем пример строительства больших танкеров. Катастрофа, произошедшая в 1967 году с танкером «Torrey Canyon», при которой 120 тыс. тонн нефти попали в море, привела к тяжелейшим экологическим последствиям (рис.2.4).

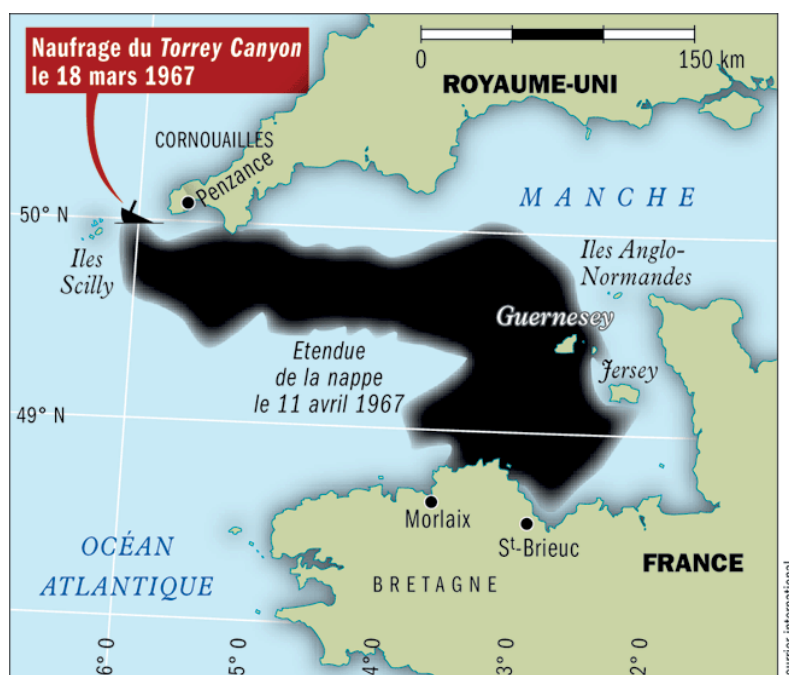


Рис. 2.4. Масштаб разлива нефти при катастрофе танкера «Torrey Canyon»

Если в 1960-х годах дедвайт (полная грузоподъемность, или водоизмещение) танкеров составлял не более 200 тыс. тонн, то в начале XXI века он уже составляет до 500 тыс. тонн нефти. Количество нефти, оказавшейся в океане, растет прямо пропорционально количеству и дедвайту танкеров, серьезные аварии случаются регулярно [22]. Иногда количество танкеров с нефтью в океане увеличивается до максимально возможного, и риски суперкатастроф многократно возрастают. Например, в апреле 2020 года, когда из-за остановки экономик промышленных стран возник избыток предложения и сотни груженых танкеров курсировали в океане и у береговых линий. Прибыль для судовладельцев и продавцов нефти обеспечена за счет постепенного уничтожения экосистем Мирового океана.

Приоритет экономической выгоды за счет причинения вреда внешней среде значительно продляет срок жизни системы. Однако дальше достигаются непреодолимые физические ограничения. Например, невозможно втиснуть в вагон метро больше пассажиров, чем там может поместиться, когда пассажиры стоят вплотную друг к другу, занимая всю возможную площадь (рис.2.5).



Рис. 2.5. Пример физического ограничения по размещению пассажиров

Следующее поколение системы, работающее на других принципах, получает возможность развития только тогда, когда предыдущая система прошла этап 3. А быстрый подъем новой системы происходит лишь после того, как кривая развития предыдущей минует пик эффективности и пойдет на спад.

Сильные решения нетривиальных задач возникают в случае их соответствия объективным законам развития систем. Каждому этапу развития системы соответствует группа законов, сформулированных Альтшуллером и дополненных, уточненных его учениками:

- Статика – применяются вначале жизни системы;
- Кинематика – позволяют достичь пика развития технической системы;
- Динамика – актуальны для применения после достижения пика развития при переходе к новой системе [23].

Эволюция технических систем заключается в том, что созданная система развивается, доходит до предела своих возможностей, далее замещается другой системой, основанной на неких других принципах (рис.2.6).

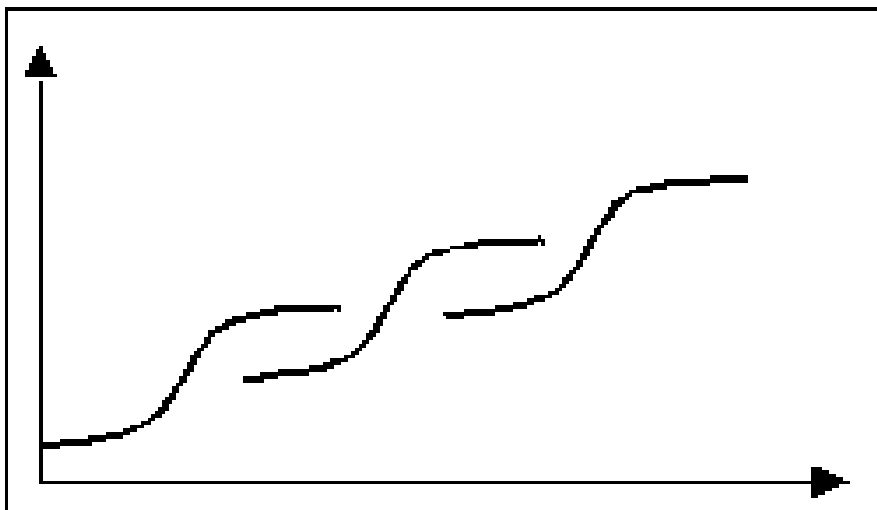


Рис. 2.6. Замена системы на другую систему, основанную на других принципах с улучшением основного показателя

2.3. Матрица бостонской консультационной группы

Для понимания S-кривой развития систем в экономике хорошо подходит матрица бостонской консультационной группы (рис.2.7).

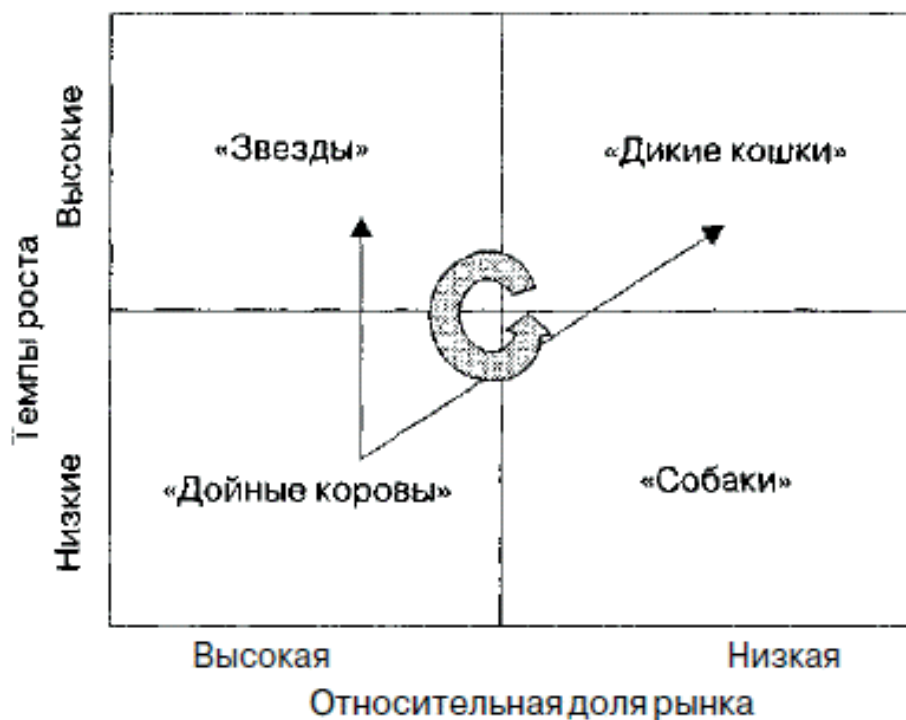


Рис. 2.7. Матрица бостонской консультационной группы

На определенной нише фирма выходит со своей разработкой («Дикая кошка»). Эта разработка имеет много технических недостатков, но в случае нахождения потребителей тема имеет дальнейшие перспективы развития. Если удалось эту разработку довести до рабочего состояния и найти платежеспособный спрос, то разработка переходит в квадрант «Звезда». На этом шаге обеспечен высокий спрос, растут тиражи изделия и решаются проблемы экспансии (новые производства, новые модификации, новые рынки сбыта). Со временем рынок насыщается, и изделие переходит в квадрант «Дойная корова» – когда имеется платежеспособный спрос, изделие выпускается большими тиражами и приносит максимальную прибыль. В данной ситуации усилия компании направлены на

оптимизацию логистики, снижение себестоимости продукции, ведь при больших тиражах это приводит к высокой отдаче. Именно на этой стадии лучше всего подходит применение функционально-стоимостного анализа. Фаза «Дойной коровы» может длиться достаточно долго, до тех пор, пока система с альтернативной технологией не выйдет в фазу «Звезда». Тогда предыдущая система переходит в квадрант «Собака» – емкость рынка уменьшается, инвестиции в улучшение системы не оправдывают себя, от такого актива лучше избавляться. Задачи компании на разных фазах S-кривой жизненного цикла представлены на рис. 2.8 [24].



Рис. 2.8. Основные задачи компании по мере развития системы

2.4. Законы статики: полноты частей системы, «энергетической проводимости», согласования ритмики частей системы

Рассмотрим законы «Статики», которые определяют начало жизни технических систем:

- Закон 1. Закон полноты частей системы;
- Закон 2. Закон «энергетической проводимости»;
- Закон 3. Закон согласования ритмики частей системы.

Закон 1. Закон полноты частей системы

Необходимым условием принципиальной жизнеспособности технической системы является наличие и минимальная работоспособность основных частей системы: двигатель, трансмиссия, рабочий орган и орган управления.

Недостаточно, чтобы отдельная часть была работоспособной, она должна быть работоспособной в составе конкретной технической системы для выполнения определенной функции.

Двигатель – элемент системы, являющийся источником или накопителем энергии для выполнения требуемой функции.

Рабочий орган – элемент системы, передающий энергию элементам окружающей среды и совершающий выполнение требуемой функции.

Трансмиссия – элемент системы, транспортирующий энергию от двигателя к рабочему органу.

Средство управления – элемент системы, регулирующий поток энергии по ее частям и согласующий их работу во времени и пространстве.

Обобщенный состав технической системы представлен на рис.2.9.

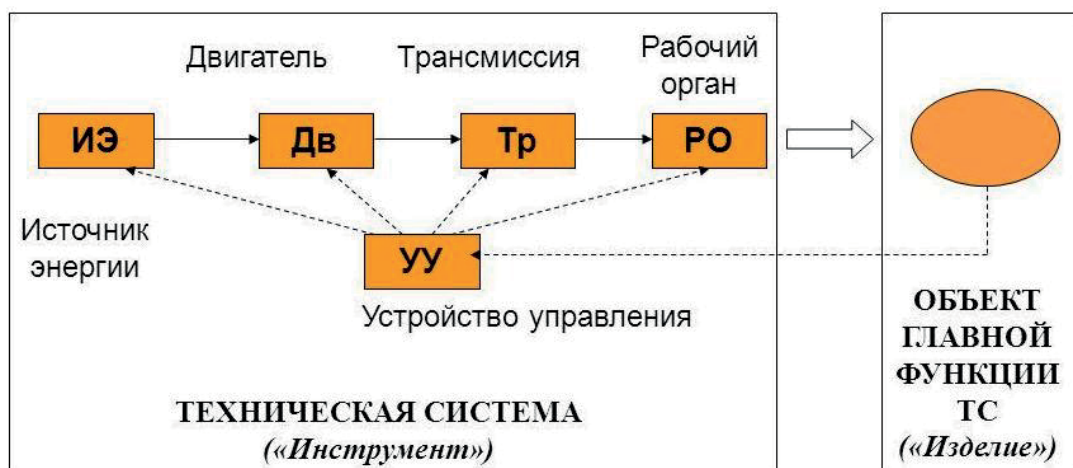


Рис. 2.9. Обобщенный состав технической системы

Для осуществления главной функции вычислений в составе персональной ЭВМ (рис.2.10):

- источником энергии является блок питания,
- трансмиссия – провода, шлейфы, шины,
- устройство управления – клавиатура, манипулятор, компьютерные программы,
- рабочий орган – процессор.

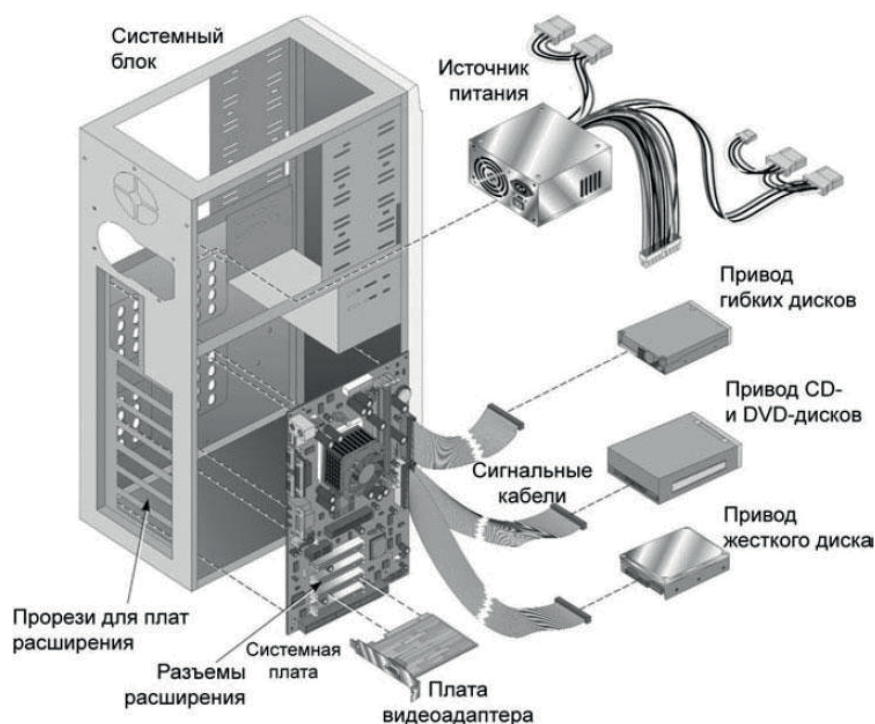


Рис. 2.10. Состав технической системы ПЭВМ

Существует много систем, которые состоят из неполного набора элементов (например, многие компьютерные приложения, реализующие конкретную функцию). В этих случаях остальные элементы системы и соответствующие связи мы обнаружим, если будем рассматривать систему более высокого уровня, где анализируемая система является подсистемой.

Эффективность технической системы зависит от возможностей управления, для этого в составе системы должны быть управляемые элементы. Некоторые специальности ВАК РФ направлены именно на

контур управления: например группа специальностей 05.13.00 «Информатика, вычислительная техника и управление». Программное обеспечение по большей части выполняет задачи управления в системе и является частью технической системы.

Развитие управления в системе осуществляется в целях того, кто управляет. Например, аэростат для вертикального подъема является управляемой технической системой с помощью регулирования температуры подаваемого воздуха. Для обеспечения управляемости аэростатом в горизонтальном направлении требуется ввести дополнительный управляемый элемент (двигатель с винтом), и тогда он становится дирижаблем (рис.2.11). Слово «дирижабль» происходит от слова дирижировать – управлять.



Рис. 2.11. Элементы управления в дирижабле

Закон 2. Закон «энергетической проводимости» системы

Необходимым условием принципиальной жизнеспособности технической системы является сквозной проход энергии по всем частям системы.

Проблемы с прохождением энергии сквозь всю систему приводят к тому, что какая-то часть системы не будет получать энергию и не сможет функционировать. Энергия помимо обеспечения работы всех частей самой технической системы, также выполняет

компенсацию потерь и утечек в системе, расходуется на измерение параметров работы частей системы и обрабатываемого изделия.

Передача энергии от одной части системы к другой может быть

- вещественной (например, вал, шестерни, рычаги и т. д.),
- полевой (например, магнитное поле),
- вещественно-полевой (например, передача энергии потоком заряженных частиц).

Для информационных систем проводимость энергии следует воспринимать как проводимость информации.

Ухудшение энергопроводимости в каком-либо месте системы приводит к ухудшению какого-либо показателя функционирования системы и является источником изобретательской задачи. По мере развития техническая система становится не только хорошим проводником энергии, но и снижает потери энергии: при преобразовании, на стыках элементов системы, за счет сокращения отходов.

На изменении энергопроводимости основаны правила создания новых технических решений:

- Если элементы при взаимодействии образуют энергопроводящую систему с полезной функцией, то для повышения ее работоспособности в зонах контакта следует применять вещества с близкими или одинаковыми уровнями развития.

- Если элементы системы при взаимодействии образуют энергопроводящую систему с вредной функцией, то для ее разрушения в зонах контакта элементов желательно использовать вещества с различными или противоположными уровнями развития.

- Если элементы при взаимодействии друг с другом образуют энергопроводящую систему с вредной и полезной функцией, то в зонах контакта элементов нужно помещать вещества, уровень

развития которых и физико-химические свойства изменяются под воздействием какого-либо управляемого вещества или поля.

Закон 3. Закон согласования ритмики частей системы

Эффективно работают и жизнеспособны только системы, в которых вид колебаний подобран так, чтобы части системы не мешали друг другу и наилучшим образом выполняли полезную функцию.

Частота собственных колебаний присутствует в каждой части системы и зависит от параметров самого объекта (например, от массогабаритных и механических, емкостных и индукционных характеристик).

Вынужденные колебания связаны с воздействием внешней среды системы, при совпадении частот внешнего воздействия и элемента системы возникает резонанс. Для того чтобы улучшить работу системы, следует или согласовать колебания частей или, наоборот, рассогласовать их. Использование резонанса (или предупреждение его появления) значительно улучшает эффективность функционирования системы и достигается изменением характеристик элементов системы (размеров, массы, частоты), при этом в систему не требуется добавлять ничего нового.

В авиации известно такое явление, как земной резонанс – самовозбуждающиеся колебания вертолета на земле с нарастающей амплитудой, которое может привести к разрушению вертолета.

Применяется явление резонанса и в электронике. Колебательный контур, состоящий из емкости и индуктивности, используется в элементах настройки и электрических фильтрах. Резонанс может быть и вредным, если вызывает искажение сигнала или паразитные шумы.

Другими словами, эффективное выполнение главной функции возможно, если периодичность работы подсистем системы

согласована по частоте колебаний, взаимосвязей, параметров (материал, размер, форма, прочность и др.), процессов, происходящих в надсистемах и т. д.

Для повышения эффективности работы системы элементы системы должны быть не лучшими, а согласованными по своим характеристикам, режимам функционирования.

2.5. Законы кинематики: увеличения степени идеальности системы, неравномерности развития частей системы, перехода в надсистему

Законы кинематики определяют развитие технических систем, независимо от конкретных технических и физических факторов:

- Закон 4. Закон увеличения степени идеальности системы;
- Закон 5. Закон неравномерности развития частей системы;
- Закон 6. Закон перехода в надсистему.

Закон 4. Закон увеличения степени идеальности системы

Развитие всех систем идет в направлении увеличения степени идеальности. Сумма полезных функций по отношению к затратам на существование системы или коэффициент полезного действия (КПД) системы постепенно увеличивается. Под идеальной подразумевается система, которая фактически отсутствует, а функция ее сохраняется и выполняется.

При приближении к идеалу многие реальные технические системы становятся все более крупноразмерными и тяжелыми (грузовые самолеты, автомобили (рис. 2.12)).

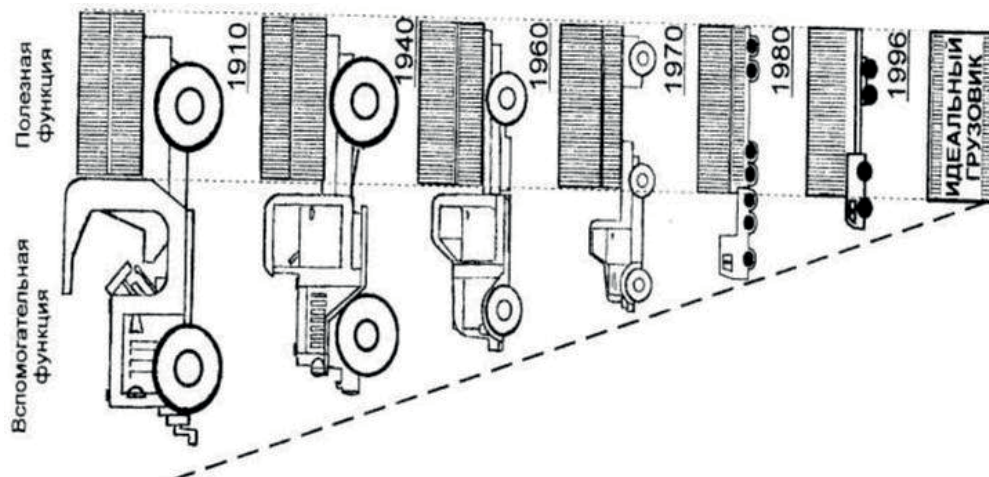


Рис. 2.12. Повышение уровня идеальности грузовых автомобилей

Основные пути приближения к идеалу реализуются в других законах развития:

- повышение количества выполняемых функций;
- «свертывание» в рабочий орган;
- переход в надсистему.

Степень идеальности может меняться и многократно переформулироваться. Например, мобильный телефон (рис. 2.13) помимо звонков стал обеспечивать обмен сообщениями (повышение количества выполняемых функций), клавиатура соединилась с сенсором («свертывание» в рабочий орган), телефон стал частью планшета или автомобиля (переход в надсистему).



Рис. 2.13. Повышение уровня идеальности мобильных телефонов, сенсоров мобильных телефонов

При приближении к идеалу техническая система вначале борется с силами природы, затем приспосабливается к ним и, наконец, использует их для своих целей. Закон увеличения идеальности наиболее эффективно применяется к тому элементу, который непосредственно расположен в зоне возникновения конфликта или порождает нежелательные явления. При этом повышение степени идеальности, как правило, осуществляется применением незадействованных ранее ресурсов (веществ, полей), имеющихся в зоне возникновения задачи. Например, антенна в телефонах стала уменьшаться, частично функцию антенны выполняет человек, а мощность сигнала стала переменной в зависимости от условий приема.

По тому, насколько дешевые и доступные ресурсы использованы, насколько снимается конфликт предложенным решением, можно судить о том насколько удалось приблизиться к идеалу.

Закон 5. Закон неравномерности развития частей системы

Развитие частей системы происходит неравномерно, и чем система более сложна, тем сильнее будет проявляться неравномерность развития ее элементов.

Причиной неравномерности становится требование увеличения главной полезной функции системы. Для этого приходится усилить определенное свойство элемента системы – это становится началом специализации элемента, дифференциации свойств в системе. При усилении одних свойств элемента нарушается согласованность взаимодействия с другими элементами, возникает противоречие. Противоречие разрешается появлением новой модификации системы в которой достигается уровень согласованности между элементами системы на новом уровне. Затем цепочка повторяется.

Из элементов системы приоритетной модернизации требует рабочий орган, именно он должен опережать в своем развитии остальные части системы, то есть обладать большей степенью динамизации по веществу, энергии или организации. Качество рабочего органа, непосредственно соприкасающегося с изделием, задает потенциал возможностей всей системы в целом. Пример, если мы стремимся улучшить проходимость транспортного средства, то следует заняться улучшением и модификацией колес. Именно потенциал колес задаст потолок проходимости, а двигатель, трансмиссия лишь помогут этот потенциал реализовать. Платформа-эллипсоид: Арктический разведчик/спасатель «Богомол» с принципиально другой формой колес [25] демонстрирует данный закон (рис. 2.14).

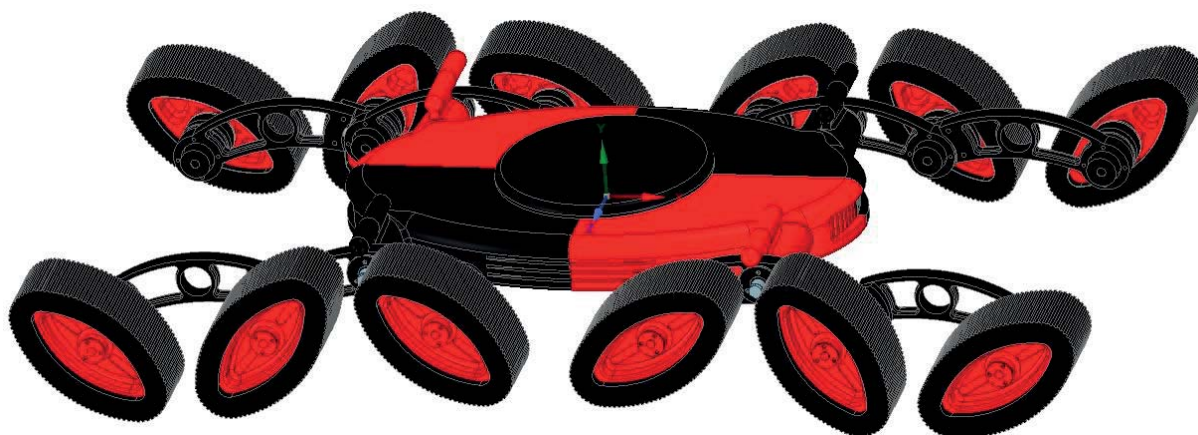


Рис. 2.14. Платформа-эллипсоид: Арктический разведчик/спасатель «Богомол»

В вычислительной технике рабочим органом является вычислительное устройство (процессор), все возможности вычислительной техники ограничиваются потенциалом процессора, и качество других элементов (устройства ввода вывода, оперативная и постоянная память) позволяет раскрыть возможности вычислителя. Смена поколений ЭВМ основана на обновлении элементной базы, на которой строится вычислительное устройство.

Закон 6. Закон перехода в надсистему

После того как система исчерпала возможности развития, она включается в состав надсистемы, и развитие продолжается на уровне надсистемы.

В качестве иллюстрации можно привести историю создания транзисторов, затем интегральных микросхем от малых до сверхбольших (рис.2.15). Эволюция продолжается на уровне компоновки, программирования микросхем.

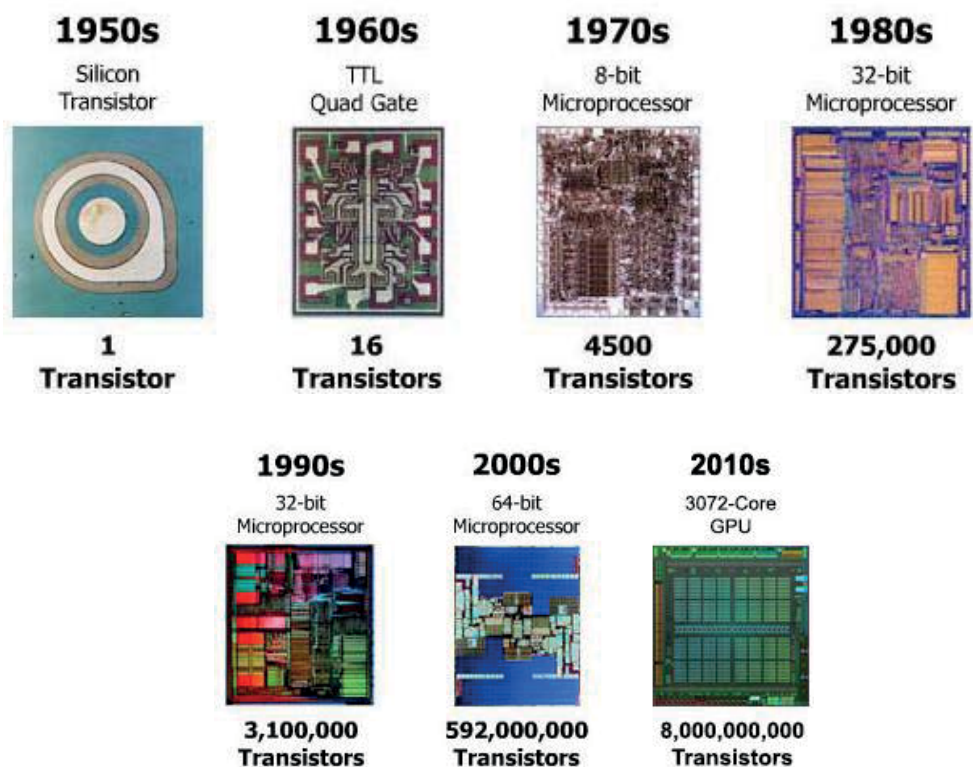


Рис. 2.15. Закон перехода в надсистему на примере транзисторов

2.6. Законы динамики: перехода с макроуровня на микроуровень, перехода к более управляемым ресурсам

Законы «Статики» и «Кинематики» универсальны – они применимы не только к техническим, но и к биологическим, социальным системам. Законы «Динамики» отражают развитие современных технических систем под действием конкретных технических и физических факторов:

- Закон 7. Закон перехода с макроуровня на микроуровень.
- Закон 8. Закон перехода к более управляемым ресурсам.

При создании новой технической системы в ней присутствуют жесткие внутренние связи и мало специализированных подсистем. При эксплуатации при изменении внешних условий выясняется, что система часто выходит из строя и недолговечна. Возникает потребность в адаптации системы. Жесткие системы для повышения их эффективности должны становиться динамичными, то есть

переходить к более гибкой, быстро меняющейся структуре и к режиму работы, подстраивающемуся под изменения внешней среды.

Для механических систем совершенствование начинается с перехода от неподвижных частей к движущимся, там, где имеется излом, вводится шарнир, жесткие элементы заменяются на гибкие, на гидро- и пневмоконструкции, используется вибрация, периодическое изменение формы и др. (рис. 2.16).



Рис. 2.16. Динамизация механической системы «шлагбаум»

Пример динамизации механической системы – эволюция автомобильных шин, главной задачей которых является обеспечение сцепления с дорогой. На первых автомобилях применялся стальной обод на колесе, не обеспечивающий амортизации, изобретение резины привело к изготовлению более мягких покрышек (1844 г.), улучшить сцепление и амортизацию при движении позволили надувные покрышки (1889 г.). В XXI веке в концептах созданы безвоздушные покрышки, не требующие накачивания, устойчивые к проколам и динамично меняющие свои характеристики под рельеф дороги (рис. 2.17).



Рис. 2.17. Динамизация автомобильных шин: резиновые, надувные, безвоздушные

Дальнейшая динамизация основана на применении физических и химических эффектов и явлений, введении обратной связи, применении приемов самоорганизации, «интеллектуализация» техники. Динамизация вещества системы проходит в определенной последовательности (рис.2.18). Затем следует динамизация поля – переход от постоянного действия к импульсному, к переменным и нелинейным полям. Программное обеспечение в системе, моделирование системы (ее элементов, параметров, режимов и т. д.) может обеспечить максимальную динамизацию при использовании преимущественно виртуальных ресурсов вместо материальных.

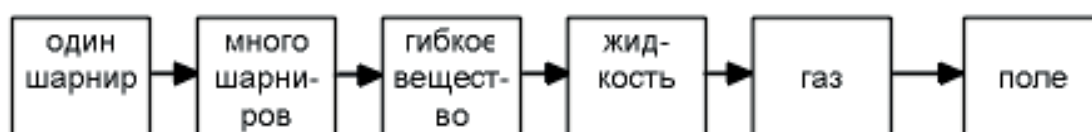


Рис. 2.18. Последовательность динамизации вещества системы

Закон 7. Закон перехода с макроуровня на микроуровень

Развитие рабочих органов системы идет сначала на макро-, а затем на микроуровне. Переход с макро- на микроуровень – одна из основных тенденций развития современных технических систем.

В большинстве современных технических систем рабочими органами являются конструкции: винты самолета, колеса автомобиля, резцы токарного станка, ковш экскаватора, пластины жесткого диска и т. д. Для повышения эффективности системы проводится развитие рабочих органов в пределах макроуровня до тех пор, пока не будет достигнуто физическое ограничение. Следующее существенно улучшение системы при сохранении функции требует перестраивать рабочий орган на микроуровне. Вместо элементов конструкции работу выполняют частицы вещества и т. п.

Например, вместо печатающей головки с набором иголок в матричном принтере в лазерном принтере формирование изображения осуществляется модулем лазерного сканирования, магнитным и термическим элементом с запеканием мельчайшего порошка.

Для достижения высоких результатов задействуются возможности структуры вещества. Вначале используется кристаллическая решетка, затем ассоциации молекул, единичная молекула, часть молекулы, атом. Одна из специальностей ВАК как раз посвящена наноуровню – 05.16.08 «Нанотехнологии и наноматериалы (по отраслям)».

Эволюция двигателей самолетов также демонстрирует вынужденный переход с макроуровня на микроуровень. В погоне за грузоподъемностью самолёты снабжались большим числом моторов (на рис.2.19 советский семимоторный самолет-гигант К-7 1933 года). Важнейшим изобретением стала замена макрорабочего органа – винта

на микроуровень – газовую струю (на рис.2.20 первый реактивный итальянский самолет КК-1 1940 года).

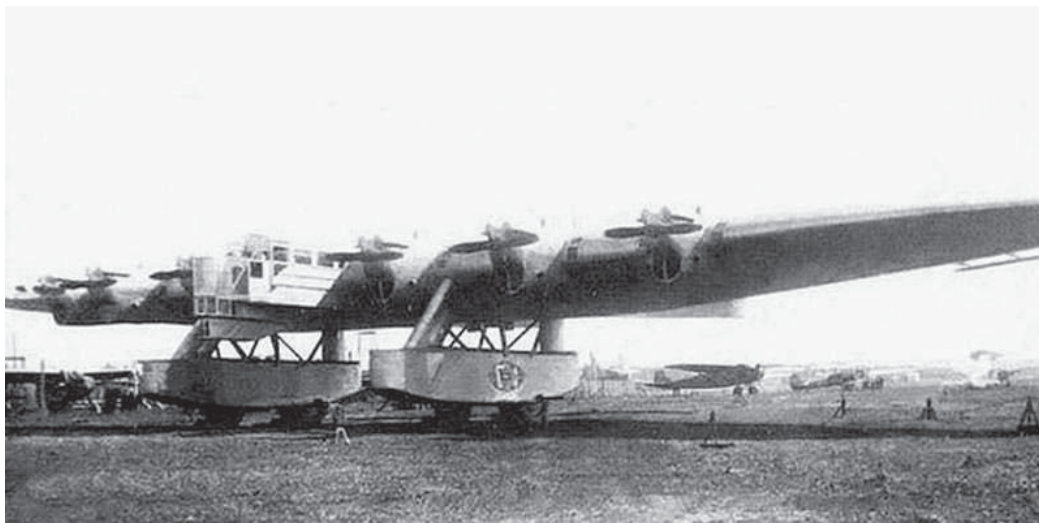


Рис. 2.19. Семимоторный самолет-гигант К-7 1933 года



Рис. 2.20. Первый реактивный самолет КК-1 1940 года

Закон 8. Закон перехода к более управляемым ресурсам

Развитие технических систем идет в направлении увеличения степени управляемости.

Повышение управляемости в технических системах идет в направлении:

- перехода от механического воздействия к применению различных полей (электромагнитных);
- увеличения степени дисперсности веществ;
- числа связей между элементами;
- отзывчивости системы.

Само воздействие может быть в одну или обе стороны, как полезным, так и вредным, может представлять переменное воздействие.

В качестве примера закона перехода к более управляемым ресурсам можно рассмотреть эволюцию управления вычислительной техникой. Первые компьютеры управлялись на аппаратном уровне (hardware) – электронные реле, транзисторы и т. п. Следующим этапом стало появление низкоуровневых языков программирования (Assembler). Позже были разработаны языки более высоких уровней – Fortran, С с управлением не на уровне отдельных команд, а на уровне классов, модулей и библиотек. Более высокий уровень управляемости – объектно-ориентированное программирование и создание баз данных. Объединение компьютеров в сети от локальных до интернета позволило расширить возможности управления информационными ресурсами. Последовала оцифровка и рост мультимедиа контента. Пользователями вычислительных сетей стала большая часть населения Земли, объектами управления стало огромное количество устройств (интернет вещей). Объектами оцифровки стали биологические и медицинские данные, с бурным развитием биоинформатики. К огромным массивам распределенных данных стали применяться методы машинного обучения. Управление стало более гибким, масштабным и тонко настроенным с namного порядков большим количеством задействованных объектов. Задачами эффективного применения информационных технологий в различных

аспектах занимается группа научных специальностей 05.13.00 «Информатика, вычислительная техника и управление».

К управляемым ресурсам, позволяющим повысить эффективность системы, относятся различные поля и вещества.

Для запоминания основных полей мастером ТРИЗ Злотиным Борисом Львовичем [21] предложена аббревиатура «МАТХЭМ»:

- М – механическое – механические усилия, перемещения; гравитационные и центробежные силы; вибрации, удары; аэро- и гидродинамические эффекты

- А – акустическое – звуковые колебания, ультра- и инфразвук, стоячие волны, резонансные колебания

- Т – тепловое – нагрев и охлаждение

- Х – химическое – использование химических реакций

- Э – электрическое – действие постоянного или переменного токов, электростатическое поле

- М – магнитное действие постоянных и электромагнитов

Перечень полей аббревиатуры МАТХЭМ не полон, предлагается аспиранту самому дописать поля, используемые в области его исследования.

Вещества с особыми характеристиками: легкоиспаряемые (газотворные); легкорастворимые; легковытравливаемые; легкосгораемые; легкоплавящиеся; с эффектом памяти формы (металл, пластик); с эффектом Кюри; увеличивающие свой объем при застывании; полимеризующиеся; легкоразрушаемые. Научной задачей создания материалов с заданными свойствами занимаются аспиранты группы специальностей ВАК 05.17.00 «Химическая технология».

Управляемость системой может быть обеспечена применением бесплатных (дешевых) веществ, имеющихся внутри и в ближайшем окружении системы: пустота, пена, воздух, вода, сыпучие тела,

отходы. Известные примеры с управлением высотой полета на воздушных шарах, глубиной погружения подводных лодок, используя балласт.

2.7. Частные случаи законов: самосборки, повышения свернутости системы, вытеснения человека

Рассмотрим частные случаи законов, которые раскрывают варианты, а как же именно может происходить приближение к идеальной системе.

Закон 9. Закон самосборки. Переход от систем, которые требуется детально создавать, продумывать и контролировать, к «самособирающимся» системам.

Как итог развития нелинейной неравновесной термодинамики появилась наука о самоорганизации и устойчивости структур различных сложных неравновесных систем: физических, химических, биологических и социальных – синергетика. Самоорганизация реализуется в физике неравновесных процессов, а также в химических реакциях, где она часто описывается как самосборка.

Термин самосборка относится к методам получения наноструктур (наноматериалов) «снизу–вверх». Самосборка представляет собой процесс, в котором неупорядоченная система уже существующих компонентов образует организованную структуру, как следствие конкретных локальных взаимодействий между самими компонентами, без внешнего направления. Основная задача для реализации самосборки – необходимость повлиять на параметры системы и задать свойства отдельных частиц таким образом, чтобы они организовывались с образованием желаемой структуры [26].

Самоорганизация в природных системах имеет следующие особенности:

1. Распределение суммарного знания по всей системе и самоуправляемость без предписаний сверху вниз – избыточность командных пунктов.

2. Определенную независимость и автономию каждой части, из которых составлена система, при нежизнеспособности отдельных элементов вне системы (термит, термитник, лес, экосистема).

3. Постоянную адаптацию, похожую на поиск нового равновесия. Биосистемы стремятся к равновесию, но никогда его не достигают сохраняя неравновесность, нестабильность, возможность перестройки к новым условиям.

4. Самоизменяющиеся правила, вокруг которых и возникает сложность поведения сообразная сложности среды.

5. Отсутствие четкой границы между системой и внешней средой. Биосистемы постоянно стремятся к захвату любого ближнего участка, перекраивают границы, которые постепенно приобретают фрактальные очертания. В биосистемах ярко выражен эффект эмерджентности [27].

Самоорганизация в кибернетике – процесс, в ходе которого создается, воспроизводится или совершенствуется организация сложной динамической системы. Процессы самоорганизации могут иметь место только в системах, обладающих высоким уровнем сложности и большим количеством элементов, связи между которыми имеют не жесткий, а вероятностный характер. Важно наличие распределение элементов управления по системе и большого количества обратных связей (образование положительной обратной связи для регуляции в системе). Как можно отметить, для успешной реализации эффективной самособирающейся искусственной системы требуется скопировать многое у биосистем.

При успешной реализации самособирающейся системы небольшие и понятные правила на микроуровне трансформируются в сложное организованное поведение на макроуровне [28].

Из ДНК происходит самосборка клеток, и все животные являются результатом самосборки (рис. 2.21). Правила дорожного движения, которые соблюдает каждый участник исключительно в части себя касающейся, выливаются в организованный поток на трассе (рис. 2.22). Секрет успеха концепции Интернет связан с использованием правил самосборки.

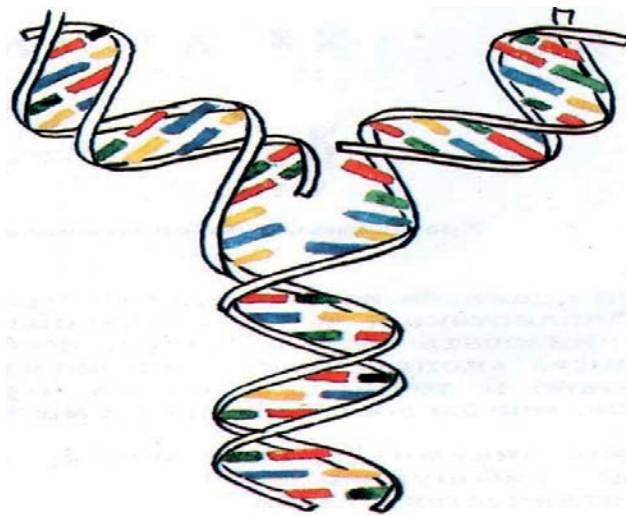


Рис. 2.21. Самосборка ДНК



Рис. 2.22. Самосборка участников дорожного движения

Закон 10. Закон повышение свернутости системы

Функции, которыми перестают пользоваться, исчезают вместе с элементами системы, которые их выполняют. Элементов становится меньше за счет того, что один элемент становится способен выполнять несколько функций.

В каких случаях элемент может быть свернут:

- если исчезла выполняемая им функция; например, шумоизоляция в мобильном телефоне не нужна, поскольку он не издает шума по сравнению с более мощным персональным компьютером;

- если объект функции сам выполняет эту функцию. Например, в магазинах самообслуживания продавец не подбирает товары – этим занимаются сами покупатели. Если видеокарта по своей конструкции имеет эффективный теплоотвод, то вентилятор не нужен;

- если функцию выполняют оставшиеся элементы системы или надсистемы. Например, в автомобилях функцию обеспечения жесткости конструкции может выполнять рама, а может выполнять сам кузов, и рама в легковых автомобилях в большинстве случаев отсутствует.

По достижению цели существования системы или при исчезновении цели система распадается.

Последовательность свертки системы манипулятор типа «мышь» с повышением уровня идеальности (рис. 2.23):

- исчез провод, а функция выполняется;
- исчезли кнопки, а функция выполняется;
- исчезло колесико, а функция выполняется;
- исчез манипулятор типа «мышь» (появились сенсорные экраны), а функция выполняется.



**Рис. 2.23. Свертка системы манипулятор типа «мышь»:
исчезли кнопки, колесико, шнур**

Источник освещения поэтапно может быть свернут до самого излучающего световой поток элемента (рис. 2.24).



Рис. 2.24. Последовательность свертки системы источник освещения

Свертывание системы можно проследить на широко используемом в менеджменте методе функционально-стоимостного анализа (ФСА), или в английской аббревиатуре Activity Based Costing (ABC).

В 30-е годы XX века эмигрировавший из Италии советский авиаконструктор Роберт Людвигович Бартини разработал метод, основанный на функциональной модели объекта, для которой формулировал идеальный конечный результат и обострял противоречие в системе. Функциональный подход Бартини стал основой функционально-стоимостного анализа.

В конце сороковых годов XX века Юрий Михайлович Соболев, инженер-конструктор Пермского телефонного завода СССР, применил системный анализ и поэлементную отработку изделий. Он рассматривал каждый конструктивный элемент как самостоятельную часть конструкции, формулировал его функциональное назначение как основное или вспомогательное.

Во время Второй мировой войны американская компания «Дженерал электрик» вынужденно удешевляла изделия из-за дефицита материалов. После войны инженер компании Лоуренс Д. Майлс, сотрудник отдела снабжения, знавший о работах Соболева, занялся анализом отзывов пользователей о работе изделий. Он выявил, что пользователи в большинстве случаев не заметили ухудшения характеристик изделий. Л. Майлс стал анализировать поэлементно каждое изделие на возможность его удешевления, сэкономил фирме огромные суммы и в 1952 году зарегистрировал метод стоимостного анализа.

Методы ФСА активно развивались в СССР, США, Японии и других развитых странах [29].

Резюмируем, метод ФСА заключается в том, что пользователи оплачивают нужные им функции, и пользователей не интересуют издержки компании. Часть функций изделий являются вспомогательными, некоторые редко используются – именно связанные с ними элементы системы должны быть проанализированы на возможность снижения издержек или устранения из конструкции.

По данным американской статистики, каждый доллар, вложенный в ФСА, может принести от 7 до 20 долларов экономии за счет снижения себестоимости продукции.

Закон 11 Закон вытеснения человека

В процессе развития технической системы техника начинает выполнять функции, которые ранее выполнял человек, постепенно приближаясь к полностью «бесчеловечной» системе.

Термин «Робот» придумал чешский писатель Карел Чапек в начале XX века и роботами обозначал механизмы, созданные для выполнения тяжелой работы с заменой людей [30]. Робототехника, механизация и автоматизация – одни из основных направлений развития современных технологий.

Вытеснение человека происходит не сразу и не всегда полностью. Анализ по ФСА позволяет выяснить, какие функции экономически эффективнее выполняет человек, а какие технические устройства.

Возможно вытеснение человека как объекта, замена его деятельности устройствами, выполняющими те же операции. Но также возможен реинжиниринг технологий с принципиальным изменением выполняемых операций. Например, беспилотные летательные аппараты (рис. 2.25) и грузовые автомобили весьма существенно меняют свои характеристики: исчезают место для размещения пилотов и системы жизнеобеспечения, транспорт становится способен выдерживать перегрузки, которые раньше были несовместимы с сохранением здоровья пилотом и поэтому были невозможны.



Рис. 2.25. Беспилотный летательный аппарат – пример вытеснения человека из системы

При эволюции системы за человеком может оставаться творческая, аналитическая нестандартная работа, но также может быть и наоборот, что человеку останется менее квалифицированная работа. Например, внедрение систем навигации снизило требования к таксистам и их оплату.

Примеры вытеснения человека – магазин самообслуживания, банковский автомат, сборочный конвейер (рис. 2.26).

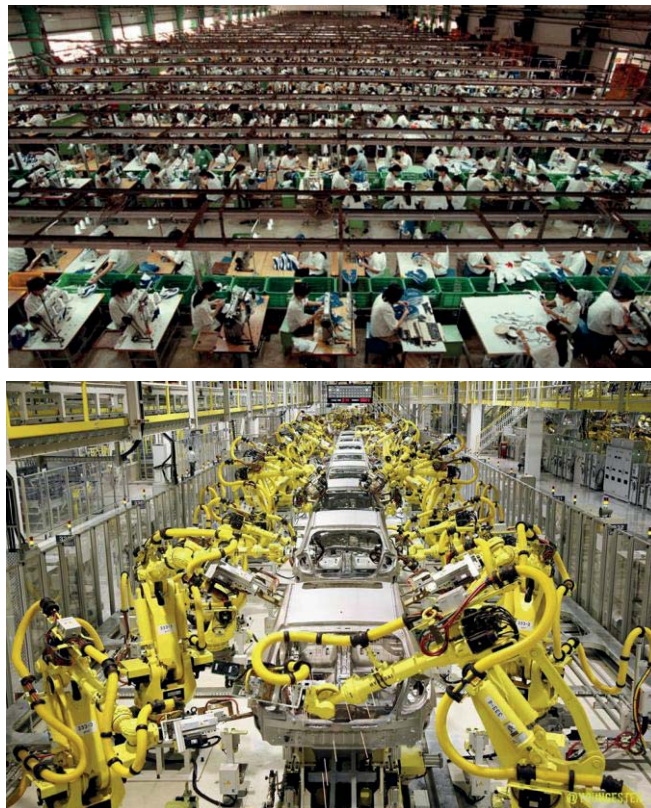


Рис. 2.26. Вытеснение человека в производстве

3. УРОВНИ ИЗОБРЕТАТЕЛЬСКИХ ЗАДАЧ

3.1. Главная полезная функция системы

Любая техническая система создается для того, чтобы выполнять свою главную полезную функцию.

Иными словами, пользователю нужна не система сама по себе, а ее функция. А наличие системы – лишь необходимость, в которой всегда присутствуют факторы расплаты.

Так, автомобиль помогает человеку перемещаться из одного места в другое (выполняет полезную функцию). Издержки присутствуют в виде необходимости ремонтировать автомобиль, оплачивать налоги и топливо, простаивать в пробках, искать место для парковки, очищать машину и дорогу в непогоду, риска аварий и угона (рис. 3.1). Набирают популярность альтернативные варианты, уменьшающие издержки: велотранспорт, общественный транспорт, такси, каршеринг и т.д.



Рис. 3.1. Факторы расплаты на наличие автомобиля

Как определить, какая из функций технической системы является главной? Следует попробовать упростить систему, отказываясь от функций, которые она выполняет. И как только мы попробуем устранить функцию, которая является главной, система становится ненужной. Нет главной функции – нет и системы.

Например, главная функция интернет браузера – отображение информации.

Помимо главной функции, система может выполнять и другие, второстепенные. Например, для автомобиля вторичных функций множество: создание комфорта для пассажиров, источник электричества и тепла, место хранения и т. д.

Если техническую систему использовать не по назначению, то в ней проявляются латентные функции. Например, автомобиль можно использовать вместо грузила или подставки. Иногда решение изобретательской задачи сводится к нахождению необычного применения технической системы.

Фактор расплаты от технической системы присутствует как вредные функции. Вредные функции – загрязнение выхлопами, ухудшение здоровья водителя и пассажиров.

3.2. Идеальный конечный результат

При разработке новых решений крайне важно представлять себе идеальный конечный результат. Напомним, что это такое решение проблемы (задачи), при котором системы нет, а функция выполняется. Такой подход может потребовать многократной переформулировки исходной задачи.

Например, задача безопасности дорожного движения может решаться на уровне охоты патрульной службы за пьяными водителями, судов и вырезвителей, устранения последствий пьяных аварий. Возможно решение на уровне «умных» автомобилей, не дающих возможности пьяным ездить. Может быть решение на уровне

ограничений продажи алкоголя, а может быть в сообществе недопустимо употребление алкоголя по каким-либо причинам. В последнем случае решение близко к идеальному – нет пьяных водителей и не нужна целая система борьбы с водителями и последствиями [31].

3.3. Типы противоречий: административное, техническое, физическое

Экономическая эффективность технической системы зависит от того, насколько хорошо она выполняет главную полезную функцию, ради которой она и создавалась. Чтобы повысить КПД системы, требуется усилить некоторое свойство одного из элементов. При изменении свойств элемента нарушается согласованность между элементами технической системы. Например, более мощный двигатель в автомобиле становится слишком тяжелым, более объемным, усиливаются требования к подвеске, подаче топлива, системе охлаждения. Для выпуска нового автомобиля возникла необходимость решать все эти проблемы, конфликты в системе, противоречия.

Противоречие – проявление несоответствия между разными требованиями, предъявляемыми человеком к системе, и ограничениями, налагаемыми на нее законами природы, социальными, юридическими и экономическими законами, уровнем развития науки и техники, конкретными условиями применения и т. п.

Пример. Для увеличения скорости полета самолета логично уменьшить площадь крыла. Это обеспечит снижение сопротивления движению и меньше расход горючего. Но малая площадь крыла не обеспечивает требуемой подъемной силы при малых скоростях. Посадка и взлет будут возможны лишь на более высокой скорости, а значит возникает потребность увеличения длины взлетно-посадочной

полосы. Использовать существующие аэродромы будет затруднительно, что недопустимо.

Решением стали самолеты с изменяемой стреловидностью крыла, первые образцы которых пробовали создавать во Вторую мировую войну (рис. 3.2).



Рис. 3.2. Первый советский самолет с изменяемой стреловидностью крыла СУ-17

На начальных этапах развития в системе имеются огромные неиспользованные скрытые резервы для улучшения, и противоречия разрешаются путем компромисса. Используя резервы системы, отыскиваются варианты конструкции, обеспечивающие приемлемые значения обеих конкурирующих характеристик. Такой путь инженерного улучшения возможен до некоего предела, после которого невозможно сохранять баланс характеристик. Такое противоречие может быть разрешено лишь качественным изменением – созданием принципиально новых технических решений. Развитие

техники, как мы видим, происходит по диалектическому чередованию количественных и качественных изменений.

В ТРИЗ для постановки и пошаговой детализации задач используются три типа противоречий:

- административные,
- технические,
- физические.

Большинство проблем в технических системах, как совершенно новых, опытных образцах, так и серийных, выявляются при их эксплуатации. Претензии к работе системы обычно формулируются потребителем в виде нечетких пожеланий: «хочется, чтобы было лучше...», «нужно что-то сделать» и т. д. Административное противоречие: нужно что-то предпринять, а что именно и как, совершенно непонятно. Административное противоречие формулируется крайне обобщенно и допускает огромное количество вариантов уточнения задачи и, следовательно, подхода к ее решению.

Формулировка технического противоречия направлена на уточнение постановки задачи и сводится к тому, что при улучшении известными путями одного свойства (параметра) системы недопустимо ухудшается другой параметр.

Любая техническая система характеризуется большим числом пользовательских характеристик: экономичность, надежность, эргономичность, эстетичность, транспортабельность, безопасность, экологичность, технологичность и т. д.

Все эти характеристики относятся к степени выполнения техническим объектом главной и вспомогательных полезных функций, либо вредным функциям (факторы расплаты, отрицательная стоимость). Стремление улучшить одни характеристики технической системы, как правило, приводит к ухудшению других, которые могут быть не менее важны.

Рассмотрим примеры формулирования противоречий в сфере информационных технологий:

- Снижение нагрузки на сервер приводит к увеличению требований к клиентским приложениям, проблемам с актуальностью данных;

- Попытки уменьшить объем передаваемой по вычислительной сети информации приводит к задержкам работы пользовательских приложений из-за ожидания, проблемам с актуальностью отображаемых данных;

- Для удобства пользователей веб-сайта функции должны быть доступны в один клик, но это приведет к загромождению главной страницы;

- Для защиты от вредоносного программного обеспечения требуется проверка файлов на жестком диске, оперативной памяти и интернет-страниц, но работа антивирусного ПО забирает много вычислительных ресурсов;

- Датчиков должно быть много, чтобы получать более точную информацию об объекте, и должно быть мало чтобы снизить затраты ресурсов на их обслуживание [32];

- На канал связи требуется низкая нагрузка, чтобы не было помех, и требуется высокая нагрузка для передачи большого объема данных;

- Система должна отображать данные в реальном режиме времени для обеспечения управления процессом, но это невозможно, поскольку данные до отображения требуют обработки.

Итак, сформулировано техническое противоречие, которое касается внешней характеристики системы, следующий шаг уточнения постановки задачи – выявление противоречия на уровне внутреннего функционирования системы – физического противоречия.

Административное противоречие: Робот должен доехать быстро к некоторой точке на складе, чтобы успеть к разгрузке. Но ехать быстро нельзя, так как маршрут захламлен неопознанными предметами, и это опасно.

Техническое противоречие: Ехать надо быстро и в то же время медленно. Два противоречащих свойства процесса, которые обусловлены разными требованиями: необходимостью успеть вовремя и безопасностью. Поэтому физическое противоречие можно сформулировать следующим образом.

Физическое противоречие: Скорость должна быть большая, чтобы успеть, и скорость должна быть маленькая, чтобы доехать.

Таким образом, физическое противоречие — это ситуация, когда к объекту или его части предъявляются несовместимые противоположные требования. Оно строится по схеме: объект должен обладать свойством, вместе с тем, иметь противоположное антисвойство.

Для ускорения решения задач на уровне выявленных противоречий в ТРИЗ на основе анализа патентного фонда СССР систематизированы 40 типовых приемов разрешения противоречий: во времени, в пространстве, за счет изменения структуры внутри системы, за счет использования возможностей надсистемы.

Рассмотрим применение пары приемов для решения данной задачи:

- прием сделай наоборот – робот едет в противоположном направлении или объект склада сам оказывается рядом с роботом,
- прием сделать заранее – робот заранее подъезжает к нужной точке.

Все 40 приемов разрешения противоречий с примерами приведены в источнике [33].

3.4. 1-й уровень изобретательских задач

Технические науки направлены на изучение естественных законов с целью практического применения для создания и модернизации техники. Прикладные исследования и фундаментальная наука неразрывно связаны: поскольку результаты фундаментальных исследований являются теоретической основой для проведения прикладных исследований, а результаты научно-технической деятельности предоставляют набор фактов, которые могут подтверждать или опровергать научные теории.

В ТРИЗ задачи классифицируются по тому, насколько они являются принципиально новыми, от 1-го уровня («неизобретательские изобретения») до 5-го уровня (никем не исследованная тема, фантастика). Конечно, наука должна быть направлена на решение задач на более высоком уровне, и аспирант должен иметь ориентиры в выборе темы.

К первому уровню относятся задачи, решение которых не связано с устранением технических противоречий и приводит к частным улучшениям. Решение задачи первого уровня доступно каждому специалисту, поскольку:

- средства решения задачи относятся к той же области, что и задача,
- объект задачи указан точно и правильно,
- вариантов изменений немного,
- вносимые изменения локальны и не меняют баланса подсистем [21].

Решение задач 1-го уровня находит каждый специалист, перед которым такая задача поставлена. Задачу первого уровня, на которую имелся зарегистрированный патент, Альтшуллер Г.С. задал читателям газеты «Пионерская правда» (номер от 4 мая 1985 г.) и в ответ

получил 5272 писем от учащихся второго – седьмого классов из которых правильных ответов было 4570 [34].

3.5. 2-й уровень изобретательских задач

Ко второму уровню относятся задачи с техническими противоречиями, преодолеваемыми с помощью способов, известных применительно к родственным системам. Например, задача в области приборостроения решается способами, уже известными в приборостроении, но применительно к другим техническим системам. Изменения затрагивают один элемент системы и не меняют принцип действия системы. Решения задачи второго уровня – мелкие изобретения. Для получения ответа обычно требуется анализ нескольких десятков вариантов решения. Например, телефон, дополненный определителем номера, автоответчиком, но остался все же устройством, предназначенным для проведения переговоров на расстоянии.

Изобретения первого и второго уровней творчества составляют три четверти мирового патентного фонда. Они совершенствуют технику и улучшают ее показатели, в основном это конструкторские решения, которые признаются изобретениями. Изобретения первого и второго уровней может сделать любой грамотный инженер [21].

Изобретения первого и второго уровней характерны для фазы зрелости системы, и приносят крупный экономический эффект за счет больших тиражей изделий, где они могут быть применены и улучшают различные характеристики: технологичность изготовления, устранение конструктивных недочетов и т. д.

3.6. 3-й уровень изобретательских задач

Противоречие и способ его преодоления находятся в пределах разных наук, но смежных между собой (физика и химия, химия и биология). Решения задач третьего уровня связаны с устранением

противоречий, возникших на уровне системных элементов. Полностью меняется один из элементов системы, частично – связанные с ним. Количество вариантов, рассматриваемых в процессе решения, измеряется сотнями. В итоге – качественное изобретение, возможный результат целенаправленной работы исследователя за долгий интервал времени. Эта ниша для прикладной науки. К третьему уровню относится примерно пятая часть патентного фонда. Проводились опыты по решению инженерами задач третьего уровня, имеющих контрольный ответ. Статистика такова, что лишь единицы приближались к контрольному решению. Задачи третьего уровня не решаются методами мозгового штурма и подобными ему, а требуют направленной исследовательской работы [21].

В области электроники к третьему уровню можно отнести создание флеш-памяти, полупроводниковая перезаписываемая память существовала и ранее, но изобретение сотрудника фирмы Тошиба Фудзиро Масуоки 1984 года породило создание огромного количества устройств высокой компактности, энергоэффективности и надежности.

3.7. 4-й уровень изобретательских задач

Для решения задачи создается новая техническая система, позволившая устранить противоречия, присутствовавшие на уровне надсистемных элементов в прототипе – старой технической системе. В задачах четвертого уровня противоречия устраняются средствами, далеко отстоящими от науки, к которой относится задача. Если задача возникла в механике, то средства для ее решения могут быть взяты, например, из химии, биологии, оптики. В современной науке активно развиваются междисциплинарные направления: бионические, гибридные, нейронные системы, биоинформатика и т. д. Число вариантов, среди которых скрыто решение проблемы, измеряется десятками тысяч. В результате – крупное изобретение. Зачастую

найденный принцип может применяться к решению других задач второго-четвертого уровней.

Изобретения 4-го уровня составляют несколько процентов от патентного фонда. Как правило, такого рода задачи не могут быть решены в одиночку и требуют усилий больших исследовательских коллективов (научной школы) [21].

3.8. 5-й уровень изобретательских задач

Изобретательская ситуация представляет собой комплекс сложных проблем (например, строительство подводной или наземной цивилизации, колонизация других планет), количество вариантов решения которых бесконечно велико. Изобретение 5-го уровня приводит к созданию принципиально новой системы, которая постепенно дополняется другими изобретениями. Часть вспомогательных задач для решения проблемы 5-го уровня сами по себе являются новыми, не имеющими в мире аналогов, и их решение приводит к созданию новых отраслей науки и техники. Изобретения 5-го уровня всегда помимо решения технических задач затрагивают комплекс социальных проблем: появление новых профессий, изменение законодательства, смена образа жизни людей и системы ценностей в обществе и т. д. Достаточно упомянуть, как поменялась жизнь людей после изобретения автомобиля, массового распространения интернета.

В патентном фонде доли процента изобретений относятся к пятому уровню, но именно такие изобретения создали нынешнюю цивилизацию. Экономическая эффективность таких систем оказывается сильно отложенной во времени [21].

Примерами могут служить изобретение самолета, радио, киноаппарата, ЭВМ, лазера.

Особую задачу представляет собой создание фонда задач, важных для развития человечества. Уровень, к которому относится

задача, актуальна на определенный момент времени. Когда решение задачи 4-го уровня найдено и реализовано, то в дальнейшем подобное решение в другой области человеческой деятельности можно отнести лишь ко второму уровню.

Великие изобретения (четвертого и пятого уровней) обычно свершаются как результат эстафеты с большим числом участников-исследователей, в которой изобретение приписывается последнему участнику, хотя возможно он не был самым выдающимся. Как утверждал французский философ Бернар Шартрский еще в XII веке: «Мы подобны карликам, усевшимся на плечах великанов; мы видим больше и дальше, чем они, не потому, что обладаем лучшим зрением, и не потому, что выше их, но потому, что они нас подняли и увеличили наш рост собственным величием».

Увидеть этот принцип можно на примере теории относительности, создание которой приписывается Альберту Эйнштейну в 1905 году. Формулу зависимости энергии от скорости света описывал русский ученый, уроженец Симбирска Николай Алексеевич Умов (1873 год), гениальный французский математик Анри Пуанкаре сформулировал принцип относительности (1902 год), голландский Нобелевский лауреат по физике 1902 года Хендрик Лоренц сформулировал преобразование координат – основу специальной теории относительности. Приведенный список ученых, сделавших возможным появление теории относительности, далеко не полон.

3.9. Изменение уровня изобретений и их экономической эффективности на разных этапах развития системы

Создателем ТРИЗ Альтшуллером было исследовано, как меняются число, уровень и экономическая эффективность изобретений по разным этапам развития системы (рис. 3.3, а).

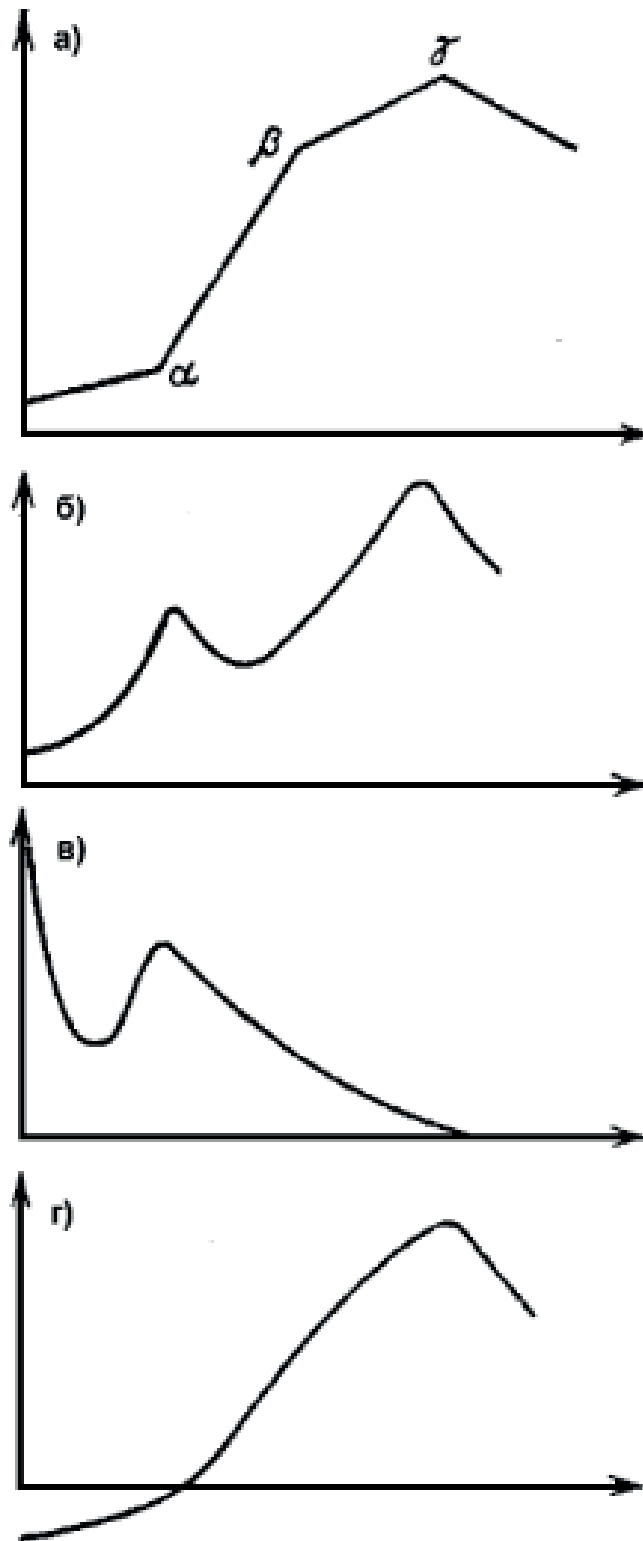


Рис. 3.3. График изменения экономической эффективности изобретений на разных этапах развития системы

Количество изобретений при создании системы невелико, значительно увеличивается в период перехода к массовому

применению системы, второй пик количества изобретений служит цели продления срока существования системы (рис. 3.3, б).

Первые изобретения, благодаря которым возникла техническая система, всегда высокого уровня, новые, значительно опережающие время (рис. 3.3, в). Постепенно уровень изобретений снижается. Высокий уровень у тех модернизаций системы, которые обеспечивают системе возможность массового применения. Далее уровень изобретений неуклонно снижается по мере исчерпания потенциала повышения эффективности системы. Когда приближается крах в существующей системе, появляются новые изобретения высокого уровня, относящиеся к системе следующего поколения, и ждут своего звездного часа.

Средняя экономическая эффективность конкретного изобретения в разные периоды развития технической системы также сильно меняется (рис. 3.3, г). Первые изобретения, несмотря на максимально высокий уровень, не дают прибыли: техническая система существует в прототипе и опытных образцах, в ней предостаточно недостатков и недоработок. Производители стараются выжать прибыль из вложенных в предыдущую систему средств, потенциальный потребитель еще не готов к использованию новой технологии. Прибыль начинает появляться после перехода к массовому применению, когда даже небольшое усовершенствование приносит большую экономию и соответственно большое вознаграждение авторам [23].

Здесь же можно отметить, что различные инновационные конкурсы требуют экономической окупаемости от проектов за 1-2 года, и поддерживаются чаще инновации 1-2 уровней, которые могут такое требование выполнить. Для изобретений высокого уровня перспектива прибыли всегда отсрочена во времени, хотя потенциальная емкость создаваемого рынка может быть огромной.

ЗАКЛЮЧЕНИЕ

Первая глава содержит схему подготовки диссертации, описание нормативно-правовых основ научно-исследовательской деятельности и подготовки диссертации в аспирантуре, основные категории в методологии научных исследований

Во втором разделе рассмотрен системный подход на основе теории решения изобретательских задач к развитию технических систем, кривая развития системы и матрица бостонской консультационной группы, приведены законы развития технических систем, действующие на разных этапах развития систем и позволяющие находить сильные решения по улучшению функционирования системы.

В третьем разделе даны определения главной полезной функции, вторичных функций и факторов расплаты системы, идеального конечного результата развития системы. Описан механизм появления противоречий при развитии системы и как грамотное формулирование противоречий служит для постановки задачи по улучшению систем, упомянуты типовые приемы разрешения противоречий, применяемые в ТРИЗ. Пояснены уровни изобретательских задач от тех, которые может решить каждый (1-й уровень), до задач создания таких уникальных систем, которые кажутся фантастическими и которые могут быть реализованы только сообществом исследователей (5-й уровень). Приведено изменение уровня изобретений, их количества и экономической эффективности по этапам развития системы.

В результате аспирант получает набор знаний, инструментов и практических навыков, понимание методологии научных исследований.

ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

Вопросы к главе 1

1) Какие виды научно-исследовательской деятельности предусмотрены нормами Федерального закона от 23.08.1996 №127-ФЗ (ред. от 31.07.2020) «О науке и государственной научно-технической политике»? Какой вид характерен для области научных исследований, в рамках которой вы подготавливаете научно-квалификационную работу (диссертаций)?

2) Охарактеризуйте область вашего научного исследования с учетом содержания паспорта соответствующей научной специальности, по которой вы подготавливаете научно-квалификационную работу (диссертацию).

3) Каково содержание и особенности объекта и предмета исследования в вашей научно-квалификационной работе (диссертации)? В чем различие объекта и предмета исследования?

4) Какие требования к соискателю ученой степени кандидата наук установлены Положением о присуждении ученых степеней?

5) Какие критериям должна соответствовать диссертация на соискание ученой степени кандидата наук в соответствии с требованиями Положения о присуждении ученых степеней?

6) Каковы структурные элементы диссертации на соискание ученой степени кандидата наук в соответствии с нормами ГОСТа Р 7.0.11-2011 Диссертация и автореферат диссертации. Структура и правила оформления? Какие элементы обязательно должны быть во введении диссертации на соискание ученой степени кандидата наук?

7) В чем заключается содержание терминов «методология», «метод», «методика», их общность и различие?

8) Какие всеобщие философские принципы составляют основу методологии большинства научных исследований? Какова взаимосвязь методологии научного исследования и философии?

9) Какие группы методов научного исследования выделяют? Приведите имеющиеся примеры применения соответствующих методов из разных групп, характерные для вашей отрасли наук и области ваших научных исследований? Какие методы в вашей области исследования относятся к общенаучным, а какие – к частным?

10) Каково содержание и различия теоретических и эмпирических методов научного исследования?

11) Какие методы в большей степени характерны для области научных исследований, соответствующей научной специальности, по которой вы подготавливаете научно-квалификационную работу (диссертацию)? Какие методы вы применяете при выполнении научного исследования с учетом темы вашей научно-квалификационной работы (диссертации)? Какие методы преобладают – теоретические или эмпирические?

12) Охарактеризуйте метод моделирования. Используете ли вы моделирование в научном исследовании по теме научно-квалификационной работы? Какие модели характерны для вашей области научного исследования?

13) Охарактеризуйте эксперимент как метод научного исследования. Используете ли вы эксперимент в научном исследовании по теме научно-квалификационной работы? Каковы особенности проведения эксперимента в области вашего научного исследования?

14) Каково содержание формализации как метода научного исследования? В чем заключается формализация в научном исследовании по теме вашей научно-квалификационной работы (диссертации)?

15) Раскройте содержание системного метода с учетом области и предмета вашего научного исследования.

16) Охарактеризуйте теоретический и эмпирический уровни вашего научного исследования.

17) Что такое научная проблема? Какой научной проблемой вы занимаетесь в рамках своего научного исследования?

18) Что такое научная гипотеза? Каким требованиям она должна соответствовать? Какова гипотеза вашего научного исследования?

19) Какие факты составляют эмпирическую основу вашего научного исследования?

20) С учетом каких рекомендаций выстраивается терминология научного исследования?

Вопросы к главе 2

1) Поясните, в чем заключается системный подход?

2) Какие основные свойства характеризуют систему? Приведите пример свойств для конкретной технической системы.

3) Опишите основные этапы развития технической системы.

4) Поясните отличия теоретической и практической кривой развития системы. В чем причины такого отличия?

5) Поясните этапы прохождения нового изделия в соответствии с Матрицей Бостонской консультационной группы.

6) В чем заключается закон полноты частей системы? Разберите состав технической системы «компьютер», «принтер», «осциллограф».

7) Сформулируйте закон энергетической проводимости системы.

8) Опишите правила создания новых технических решений с разными эффектами энергопроводимости.

9) Приведите примеры использования согласования, рассогласования ритмики частей системы, а также резонанса для решения задач.

10) Дайте определение идеальной системы. Приведите примеры увеличения степени идеальности по мере развития технической системы.

11) Назовите пути приближения к идеалу и приведите примеры.

12) Поясните, чем обусловлена неравномерность развития системы.

13) Какая часть системы должна получить опережающее развитие и почему?

14) Опишите закон перехода в надсистему и приведите пример.

15) Опишите этапы динамизации системы.

16) Приведите примеры перехода с макро- на микроуровень при развитии систем.

17) Расшифруйте аббревиатуру МАТХЭМ, приведите пример каждого из полей списка.

18) Приведите примеры полей, которые применяются в вашей области и не учтены в перечне МАТХЭМ.

19) Перечислите вещества и поля с особыми свойствами, которые активно применяются в технике.

20) В чем заключается закон самосборки? Объясните суть понятия синергетики, самоорганизации.

21) Поясните, какими путями может происходить свертка системы.

22) Объясните применение функционально-стоимостного анализа в свертке системы.

23) Поясните причины вытеснения человека из системы, перечислите варианты вытеснения.

Вопросы к главе 3

1) Приведите пример главной полезной функции, дополнительных функций и факторов расплаты для технической системы мобильный телефон.

2) Попробуйте по-разному сформулировать задачу и идеальный конечный результат для системы управления трафиком движения крупного мегаполиса.

3) Сформулируйте причины появления противоречий при развитии системы.

4) Приведите пример уточнения постановки задачи путем формулирования последовательно административного, технического и физического противоречий.

5) Приведите примеры применения приемов разрешения противоречий.

6) Сформулируйте характеристики задачи 1-го уровня. Приведите пример задачи первого уровня.

7) Сформулируйте характеристики задачи 2-го уровня. Приведите пример задачи второго уровня.

8) Сформулируйте характеристики задачи 3-го уровня. Приведите пример задачи третьего уровня.

9) Сформулируйте характеристики задачи 4-го уровня. Приведите пример задачи четвертого уровня.

10) Сформулируйте характеристики задачи 5-го уровня. Приведите пример задачи пятого уровня.

11) Объясните, как меняется уровень изобретений при создании новой технической системы, в период зрелости и старения системы.

12) Каково примерное соотношение количества изобретений по уровням? Поясните причины такой ситуации.

13) Поясните, как соотносится экономическая эффективность и уровень полученного изобретения.

14) Как вы считаете, на решение изобретательских задач какого уровня должны быть направлены диссертационные исследования?

15) Как меняется уровень сложности решения конкретной задачи со временем и что служит причиной этих изменений?

16) Решение каких задач под силу одному исследователю, а для решения каких задач требуется группа исследователей – научная школа?

17) Какие задачи сейчас вы бы отнесли к пятому уровню?

ЗАДАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ

Задание 1. Выбрать техническую систему (изделие, технологию), имеющую развитие во времени (предпочтительно связанную со своим научным исследованием):

- 1) Выполнить описание цикла развития системы с момента возникновения и по настоящее время (когда создана, на каком этапе находится сейчас, количественные параметры на каждом этапе (например, количество экземпляров)).
- 2) Разобрать состав системы.
- 3) Сформулировать главную полезную функцию, вторичные функции, вредные функции системы на каждом из этапов.
- 4) Под каждый закон развития систем (учитывая этап развития системы) подобрать пример со ссылкой на патенты (отечественные, зарубежные), публикации (открытые Интернет-ресурсы с описанием для старых открытий, базы Web of Science, Scopus, Google Scholar, РИНЦ для более современных изобретений).

Примеры систем:

- канал связи;
- телефон;
- мобильный телефон;
- сервер;
- ПЭВМ;
- тонкий клиент;
- клиент-серверные технологии;
- вычислительные устройства (процессоры);
- датчики (скорости, давления, высоты, температуры и т. д.);
- измерительные устройства (осциллограф, спидометр и т. д.);
- устройство управления (клавиатура, манипуляторы, сенсоры и т. д.);

- устройство отображения (монитор, принтер)

Задание 2. Описать место своего исследования в цикле развития технической системы:

– к какой системе относится;

- какие законы развития технических систем оно реализует;

- какому этапу развития технической системы соответствует;

- к какому уровню изобретательских задач относится

формулировка исследования. Пояснить, по каким критериям принято такое решение.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Лем, С. Сумма технологий / С. Лем. – Москва: Изд. АСТ. 2018. – 640 с.
2. Тронин, В.Г. Планирование и управление научными проектами с применением современных информационно-коммуникационных технологий : учебное пособие / В. Г. Тронин. – Ульяновск : УлГТУ, 2017. – 211 с.
3. Тронин, В.Г. Оценка результатов научно-исследовательской работы и наукометрия : учебное пособие / В. Г. Тронин, А. Р. Сафиуллин. – Ульяновск : УлГТУ, 2019. – 136 с.
4. Федеральный закон от 23.08.1996 №127-ФЗ (ред. от 31.07.2020) «О науке и государственной научно-технической политике». http://www.consultant.ru/document/cons_doc_LAW_11507/ (дата обращения: 10.08.2020).
5. Приказ Министерства образования и науки РФ от 30 июля 2014 г. №875 «Об утверждении федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.06.01 Информатика и вычислительная техника (уровень подготовки кадров высшей квалификации)». <http://base.garant.ru/70730842/> (дата обращения: 10.08.2020).
6. Мокий М.С. Методология научных исследований / М. С. Мокий, А. Л. Никифоров, В. С. Мокий ; под ред. М. С. Мокия. — М. : Изд-во Юрайт, 2015. — 255 с.
7. Паспорта научных специальностей ВАК. <http://arhvak.minobrnauki.gov.ru/316> (дата обращения: 10.08.2020).
8. Приказ Министерства образования и науки РФ от 23 октября 2017 г. №1027 «Об утверждении номенклатуры научных специальностей, по которым присуждаются ученые степени». <https://vak.minobrnauki.gov.ru/uploader/loader?type=34&name=3349241001&f=2958> (дата обращения: 10.08.2020).

9. Постановление Правительства РФ от 24.09.2013 №842 (ред. от 01.10.2018, с изм. от 26.05.2020) «О порядке присуждения ученых степеней». http://www.consultant.ru/document/cons_doc_LAW_152458/ (дата обращения: 10.08.2020).

10. ГОСТ Р 7.0.11-2011 Диссертация и автореферат диссертации. Структура и правила оформления. <http://protect.gost.ru/document.aspx?control=7&id=179727> (дата обращения: 10.08.2020)

11. Эйсмонт, Н. Г. Теоретические основы и практика научных исследований : учебное пособие / Н. Г. Эйсмонт, В. В. Даньшина, С. В. Бирюков ; Минобрнауки России, ОмГТУ. – Омск : Изд-во ОмГТУ, 2018. – 98 с.

12. Колмогоров, Ю. Н. Методы и средства научных исследований: учебное пособие / Ю. Н. Колмогоров [и др.]. – Екатеринбург : Изд-во Урал. ун-та, 2017. – 152 с.

13. Прытков, В.П. Структура научной проблемы / В.П. Прытков // Теория и практика общественного развития. – 2013. – №1. – С. 44-47.

14. Сидоренко, Н.И. Гипотеза как форма научного познания / Н.И. Сидоренко // Исторические, философские, политические и юридические науки, культурология и искусствоведение. Вопросы теории и практики. – Тамбов : Грамота. – 2014. – № 5 (43): в 3 ч. Ч. III. – С. 168-173.

15. Коэн, М. Введение в логику и научный метод / М. Коэн, Э. Нагель, пер. с англ. П. С. Куслия. – Челябинск: Социум, 2010. – 655 с.

16. Горский, Д.П. Краткий словарь по логике. / Д.П. Горский, А. А. Ивин, А. Л. Никифоров. – М. : Просвещение, 1991. – 208 с.

17. Котенко, В.П. Методологические проблемы анализа научных теорий / В.П. Котенко // Библиосфера. – 2013. – №13. – С. 29-38.

18. Обидина, Ю. С. Философия и методология науки: учебно-методическое пособие / Мар. гос. ун-т; авт.-сост. Ю. С. Обидина. – Йошкар-Ола, 2017. – 238 с.

19. Пономарев А.Б. Методология научных исследований: учебное пособие / А.Б. Пономарев, Э.А. Пикулева. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2014. – 186 с.

20. Волкова В.Н. Теория систем и системный анализ : учебник для бакалавров / В.Н. Волкова, А.А. Денисов. — М. : Изд-во Юрайт ; ИД Юрайт, 2012. – 679 с.

21. Альтшуллер, Г.С. Поиск новых идей: от озарения к технологии (Теория и практика решения изобретательских задач) / Г.С. Альтшуллер, Б.Л. Злотин, А.В. Зусман, В.И. Филатов. – Кишинев: Картя Молдовеняскэ, 1989. – 111 с.

22. Занин, С. Время «нефтяных пятен» <https://neftegaz.ru/analysis/ecology/330352-vremya-neftyanykh-pyaten/> (дата обращения: 28.07.2020).

23. Альтшуллер, Г.С. Творчество как точная наука / Г.С. Альтшуллер. – М.: Сов. радио, 1979. – 184 с.

24. Давыдова, А.А. Составление портфеля НИОКР на основе анализа матрицы БКГ / А.А. Давыдова, Д.В. Секерин // Экономические аспекты развития российской индустрии в условиях глобализации. – 2014. – С. 225-229.

25. Патент №2621530 «Колесный движитель» (Заявка № 2015148482 от 11.11.2015 г.)

26. Муртазина, Э.М. Нерешенные проблемы нанотехнологии: химическая обработка с помощью самосборки (краткий обзор зарубежных публикаций) / Э.М. Муртазина // Вестник Казанского технологического университета. 2011. №15. URL: <https://cyberleninka.ru/article/n/nereshennyye-problemy-nanotekhnologii->

himicheskaya-obrabotka-s-pomoschyu-samosborki-kratkiy-obzor-zarubezhnyh-publikatsiy (дата обращения: 28.07.2020).

27. Хищенко, В.Е. Самоорганизация: элементы теории и социальные приложения / В.Е. Хищенко. – М. : Урсс, 2005. – 224 с.

28. Хакен, Г. Синергетика. Иерархия неустойчивостей в самоорганизующихся системах и устройствах / Г. Хакен. – М.: Мир, 1985. – 420 с.

29. Рыжова, В.В. ФСА в решении управленческих задач по сокращению издержек / В.В. Рыжова. – М.: Риор, 2013. – 224 с.

30. Чапек, К. Собрание сочинений: В 3 т. Т. 2. / К. Чапек. – М. : Терра, 2004. – 416 с.

31. Тронин, В.Г. ТРИЗ в инженерном образовании / В.Г. Тронин // SMART-образование Ульяновской области. – 2018. – № 1(3). – С. 134.

32. Рубин, М. Применение ТРИЗ в проектировании и разработке ПО / М. Рубин, С. Сысоев <http://2017.secrus.org/program/submitted-presentations/applications-of-triz-methods-in-sw-development> (дата обращения: 28.07.2020).

33. Гин, А.А. Теория решения изобретательских задач : учебное пособие I уровня : учебно-методическое пособие / А.А. Гин, А.В. Кудрявцев, В.Ю. Бубенцов, А. Серединский. – 3-е изд. – Томск : Изд-во Том. политехн. ун-та, 2017. – 64 с.

34. Альтшуллер, Г.С. Найти идею. Введение в теорию решения изобретательских задач / Г.С. Альтшуллер. – Новосибирск: Наука, 1986. – 209 с.

Учебное издание

ТРОНИН Вадим Георгиевич,
САФИУЛЛИН Антон Рифкатович

МЕТОДОЛОГИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Учебное пособие

Редактор Н. А. Евдокимова

ЛР № 020640 от 22.10.97

Подписано в печать 30.09.2020. Формат 60×84/16 .

Усл. печ. л. 5,12. Тираж 100 экз. Заказ № 510. ЭИ № 1488.

Ульяновский государственный технический университет,
432027, г. Ульяновск, ул. Сев. Венец, д. 32.
ИПК «Венец» УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Н.А. ГУНИНА

**МЕЖДУНАРОДНАЯ ПРОФЕССИОНАЛЬНАЯ
КОММУНИКАЦИЯ**
методические рекомендации по дисциплине
«Международная профессиональная коммуникация»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

ГУНИНА Н.А.

Международная профессиональная коммуникация : методические рекомендации / Н.А. Гунина. – Ульяновск : УлГТУ, 2021. – 39 с.

Методические рекомендации адресованы студентам для усвоения дисциплины «Международная профессиональная коммуникация». В данных методических рекомендациях предоставлены необходимые для освоения дисциплины теоретические положения и проверочные задания. Методические рекомендации разработаны в соответствии с рабочей программой дисциплины «Методология научного познания». Методические рекомендации предназначены для студентов, обучающихся по направлению 09.04.03 «Прикладная информатика» профиль «Искусственный интеллект и бизнес-аналитика».

Работа подготовлена на кафедре «Иностранные языки».

СОДЕРЖАНИЕ

1. Профессиональная коммуникация	4
2. Научная коммуникация	25
3. Деловая коммуникация	34

1. Профессиональная коммуникация

Module 1. Professional Communication

ПР01. Устройство на работу. Основные виды работы, их краткая характеристика на английском языке; описание обязанностей, связанных с выполнением того или иного вида работы.

ПР01. Application for a job. Jobs and their brief description in English; a description of job responsibilities.

Exercise 1. Choose the correct answer.

1. A bank teller _____ in a bank.
A manages advertising B manages the credit department C receives and pays out money D tells banking stories
2. A _____ marks errors in the first printed copy of the text.
A caretaker B controller C printer D proofreader
3. He is looking for a _____ in electrical engineering.
A job B labor C occupation D work
4. What do you do in your spare time? What is your favorite _____? – I like to read books about traveling.
A job B labor C occupation D work
5. He is a member of staff and works from nine to five in the main office. He is a _____ worker.
A freelance B full-time C part-time D temporary
6. A _____ usually has quite a few vacancies for skilled and unskilled workers.
A art gallery B large construction company C local school D small travel agency

Exercise 2. Read the job ads and match the descriptions to the position.

A Digital Marketing Manager **B Sales Manager** C Electronics & Systems Engineer

Advertisement 1

Looking for your next big break? Join a progressive, forward thinking company with an exciting range of products and a fantastic reputation.

Role

You will be responsible for driving sales for Plant and Tool hire services to the construction industry. You will help to grow and develop business and maintain many existing accounts with small to large companies.

Company

A market leading hire and sales company with a fantastic reputation!

To be successful, you may have:

- Previous experience in a similar industry
- A proven track record of winning new business within the plant and tool hire markets or related.
- A high level of motivation, determination and passion.
- The ability to deal with any questions or issues to meet customer's expectations.
- A high level of communicative skills both verbally and written

Benefits

A basic salary of £43,000 - £47,000

Bonus

Company Car, Fuel Card

Advertisement 2

The

Opportunity

- An exciting opportunity to join our Navigation Sensors Group.
- The Team are responsible for the specification and assessment of inertial and satellite navigation equipment.
- The successful candidate will work within the inertial sensors team and apply proven understanding of electronics for the specification of inertial systems

The

Role

You would get detailed understanding of rate sensors and accelerometers to allow requirements to be created.

Monitoring supplier and technology roadmaps and identifying cutting edge and future technologies for investment.

Supporting sub assembly as part of an international team.

- Undertake research and design activities as part of an international team.
- Monitoring supplier and technology roadmaps and identifying cutting edge and future technologies for investment.
- Collaborating with suppliers and researchers of Navigation Sensors and related technologies.
- Designing and conducting laboratory tests.

What's

in

it

for

you?

This role offers an opportunity to be part of a successful team working to develop new and innovative solutions to address complex customer requirements.

- You will develop expertise in navigation sensors; both conventional and cutting-edge technologies.
- Involvement in the strategic growth of a rapidly evolving team.
- You will take a key role in several multi-national research programmes.
- The work is technically challenging, innovative and rewarding.

Advertisement 3

Mobile Fun- Who are we?

We are Europe's leading eCommerce retailer offering the latest mobile devices accessories from our websites, we were established in 2000. We are based in Birmingham and currently employ around 60 staff with a great reputation and having the highest possible standards. This is a fantastic place to work with a great team.

And what would the main responsibilities be?

Accountability and responsibility for the management and ongoing development of Mobile Fun's marketing team, ensuring effective coordination of marketing campaigns and activities across all Mobile Fun channels and the strategic development of all marketing activities and reach across all Mobile Fun's website portfolios.

Self Development

You will have the opportunity to access personalised learning and development opportunities that will expand your existing knowledge and challenge you and equip you with new skills.

Now, here's what we need from you:

- *Confidence and articulation in relation to the analysis of numerical data and statistics*
- *Hands-on with marketing activities*
- *Proven experience in Online marketing*
- *Proven team management experience*
- *Degree in a relevant subject*

As a part of a growing Mobile Fun team: You will receive access to a variety of our excellent benefits which could include;

- up to 25 days holiday,
- pension scheme matched up to 6%,
- Staff discounts (up to 90%)

Please apply by submitting a cover letter and a full CV!

Job Type: Full-time

Salary: £32,500.00 to £40,000.00 /year

Exercise 3. Read the job ads in 2 once again and answer the questions.

- Which job (jobs) requires previous experience in a similar industry?
- Which job (jobs) requires excellent communication skills?
- Which job (jobs) requires conducting laboratory tests?
- Which job (jobs) offers learning and development opportunities?
- Which job (jobs) requires working in a team?
- Which job (jobs) involves dealing with advanced technologies?

ПР02. Устройство на работу. Современные требования к кандидату при поступлении на работу. Основные документы при принятии на работу.

ПР02. Application for a job. Job requirements. Employment documents.

Exercise 4. Complete the sentences using the following words: *advertisement, applicant, to advertise, to apply for, requirement, position, to require, experience, to provide, curriculum vitae (CV) / resume, application, to assist, referee*

1. This company is looking for a person for theof a Sales Manager who will a Managing Director.
2. They placed an in the local newspaper two days ago.
3. Their main are experience and communicative skills. They also that an applicant should be self-disciplined.
4. should have a 3 years'
5. You must have two from your previous work and give names of your
6. Candidates should their and send to the address given in the newspaper where the company its products.

Exercise 5. Read Josh Reed's covering letter and complete it with the words and phrases from the box: *abilities experience am keen position knowledge skills*

Josh Reed
75 Berry St
Summerville QLD 4536
T: (07) 8222 1111
E: j.reed@email.com

Rachel Forrester
HR Manager
Brighton Mining
Green Plains NSW 2008
February 3, 2016

Dear Ms Forrester

Re: Mechanical Engineer Position

I am writing to apply for the ___ of Mechanical Engineer as recently advertised on SEEK.com.au.

I am a highly motivated Mechanical Engineer with a Bachelor of Engineering (Mechanical Major) and three years of practical on-site mining _____. I am very interested in joining the engineering team at Brighton Mining given your reputation for world-leading innovation in open cut and underground mining.

In my current position as a mechanical engineer at Newcrest Mining I have developed key project management ___ and the ability to improve communication with the broader project team.

I have a thorough ___ of the processes of open cut and underground mining. I possess excellent interpersonal and communication skills and my multitasking ___ are advanced.

Given my on-site experience I am accustomed to operating in a FIFO environment and working hard to keep projects running on time and within budget. I ___ to employ my skills and enthusiasm as an integral part of your team and I look forward to being able to discuss this position with you further.

Yours sincerely,
Josh Reed

Exercise 6. Read Carmen Frazier's CV and answer the questions.

1. What are the main parts of this CV?

2. What have you learned about Frazier's education?

3. What are her professional abilities?

4. Does Frazier speak any foreign languages?

5. Do you think she is a good candidate for the company Case Consultants? Why/ why not?

PERSONAL INFORMATION

Carmen Frazier

Date of Birth: May 6, 1979

Citizenship: American

PROFILE

- Extensive knowledge in Building Engineering
- Strong technical skills in AutoCad applications
- MS Windows, Mac OS, Archicad, CorelDraw, Sage
- Experienced working in large scale industries

EDUCATION

M.S. in Structural Engineering, 2006, Drexel University, Philadelphia, PA

B.S. in Civil Engineering, 2002, Drexel University, Philadelphia, PA

EMPLOYMENT HISTORY

Chief, Civil Engineer, 2007 - Present

RESPONSIBILITIES:

- Provided cost-effective solutions to recurring construction problems.
- Monitored the status of government projects and ensured compliance with civil engineering standards.

SKILLS

Adept with engineering tools and techniques

Extensive experience in residential, commercial and industrial projects

Certificate in AutoCAD

Exercise 7 Fill the gaps below with the correct present tense (Present Simple or Present Continuous)

Work

- a. What _____ (you work on) at the moment?
- b. What company/ division/ department/ section/ team _____ (you work) for?
- c. What _____ (you do)?
- d. What _____ (your company/ division/ department/ section/ team do)?
- e. What kind of company _____ (you work) for?
- f. Who _____ (you work) for?

English at work

- a. _____ (you in the middle of write) anything in English now?
- b. _____ (you ever make) English phone calls?
- c. _____ (you find) it difficult to write in English?
- d. _____ (you have problems) with teleconferences in English?
- e. _____ (you normally use) automatic translation?
- f. _____ (you often write) in English?
- g. How many English emails _____ (probably wait) for you right now?
- h. What _____ (you find difficult) difficult about English language meetings?
- i. As we speak, _____ (you think) in English?

English outside work

- a. _____ (you usually watch) movies with English subtitles, with Japanese subtitles or with no subtitles?
- b. How often _____ (you socialise) with English speakers?

Language learning

- a. _____ (you take) any other English classes now?
- b. _____ (you study) on the train on the way home in the evening?
- c. _____ (you listen) to anything in English?
- d. _____ (you listen) to NHK English radio in the morning?
- e. How _____ (you learn) vocabulary?
- f. How much English study _____ (you typically do)?
- g. _____ (your English improve)?
- h. What new English language apps _____ (you use)?

Exercise 8 Fill the gaps in the questions about your company and job (Present Simple and Continuous)

Your job

“What’s your job?”

“I _____ ¹(work) in the marketing department of a large insurance company.”

“How’s work?”

“Pretty busy. I _____ ²(get) ready to move to our New York office.”

“What _____ ³(you do)?”

“I’m an accountant.”

“What _____⁴(you do) here?”
 “I’m giving a presentation on our new product range.”
 “What exactly _____⁵(you do)?”
 “I negotiate deals with new suppliers.”
 “What _____⁶(you work on)?”
 “I’m hiring staff for a new branch in Kobe.”
 “How’s your project going?”
 “Not very well. We _____⁷(still work) on the initial plan.”
 “Do you often go abroad on business?”
 “No, very rarely, unfortunately. I mainly _____⁸(travel) around Japan.”

IP03. Компании. Структура компании, названия отделов.
IP03. Companies. Company structure, department names.

Exercise 1. Study the information about the people and fill in the gaps below.

<i>‘I’m Robert. I am responsible for the day-to-day running of the business. I represent the company in the business world’</i>
<i>‘Hi! I’m Marina. My job is to make sure that the company is producing what people want to buy.’</i>
<i>‘My name is Peter. I am responsible for the entire company when Robert is away on business. ‘</i>
<i>‘Hi! I’m Cecily. I deal with personnel matters and recruitment. I also do with issues of staff welfare.’</i>
<i>‘I’m John. My area of responsibility is financial issues and money planning’</i>
<i>‘Hi! My name is Liza. I do the bookkeeping and the payroll.’</i>
<i>‘Hi! I’m Jake. I lead the team which makes our products.’</i>
<i>‘My name is Sonya. I deal with developing and testing our new products’</i>
<i>‘I’m Ben. I’m in charge of people who sell our products.’</i>

- (1)..... Bradford: Managing Director
- (2) Thomson: Assistant Managing Director
- (3) Gates: Sales Director
- (4) Johnson: Finance Director
- (5) White: Marketing Director
- (6) Brown: R&D Manager
- (7) Tales: HR Director
- (8) Smith: Production Manager
- (9) Bay: Accountant

Exercise 2. Look at the list of departments in a company (a-h) and read people's situations (1-6) below. Decide which department each person should ask to speak to when phoning the company. There are more departments than you need.

Names of departments:

Human Resources

Sales

Production

Accounts

Quality Control

Research and Development (R&D)

Technical Support

1 Mr. Mitchell is a marketing executive who has received several complaints from customers about faulty goods.

2 Mr. Davies is a consultant who thinks he has not been paid for an invoice.

3 Mr. Finer has just received the results of the laboratory tests on a possible new product

4 Ms. Smith is a sales executive who is interested in working for the company.

5 Ms. Evans works in the company as a secretary and she has a problem with her computer.

6 Mr. Martins is a retailer who is interested in stocking the company's products.

Exercise 3. Match the words (1-10) to their definitions (A-J).

1 A.G.M.^{UK}

2 executive officer^{US}

3 board of directors

4 chairman^{UK}

5 reception

6 organisation chart

7 shareholder

8 vice president^{US}

9 headquarters

10 manager

11 managing director^{UK}

A any of several executive officers, each responsible for a separate division

B the place where visitors and clients report on arrival at a company

C person who heads a Board of Directors; head of a company; chairperson

D Annual General Meeting of a company's shareholders

E person who holds or owns shares in or a part of a company or corporation

F group of people chosen to establish policy for and control a company

G person managing the affairs of a corporation - chief executive officer

H a company's principal or main office or centre of control

I person responsible for day-to-day running of a dept.; executive officer^{US}

J senior director after the chairman responsible for day-to-day direction

K a table or plan showing a company's structure graphically

Exercise 4. Read the dialogues and fill in the missing phrases.

1) *annual turnover, employ, technical people, workforce*

A: How many people does your company ¹_____?

B: We have sixty employees. We have about forty factory workers and ² _____ and the rest are admin and sales staff. We started off with only ten people so our ³ _____ has grown a lot. What's your ⁴ _____?

A: It was over 2 million euro last year.

2) *do manufacture help*

A: What does your company ⁵ _____?

B: We ⁶ _____ and sell fire prevention and fire control equipment

A: What do your products do?

B: They ⁷ _____ prevent fires and help suppress fires once they have started.

3) *provide services guards*

A: What does your company do?

B: We ⁸ _____ security services to large businesses and hotels.

A: What kind of ⁹ _____ do you provide?

B: We provide security ¹⁰ _____, CCTV and 24-hour monitoring.

IP04. Компании. Характеристика обязанностей работников отделов, описание работы компании.

IP04. Companies. Description of the staff responsibilities and company activities.

Exercise 5. What do these companies do? Make sentences about the companies' activities, using the words below.

Example: Microsoft designs and sells IT software.

SONY

AUCHAN

CITYBANK

ADIDAS

APPLE

MICROSOFT

TOYOTA

COCA-COLA

Verbs: create, design, develop, manufacture, sell, market, offer, provide

Word combinations: banking services, cars, clothing, electronic goods, food and drinks, Internet services, IT software

Exercise 6. These sentences describe two companies, Autotech and Green Fingers. Choose pairs of sentences, which describe similar things and match them with the correct company. Underline the verbs which mean the same things.

GREEN FINGERS

A small garden-products company

1 George and Tames Hawkins began Green Fingers in the 1920s.

2

3

4

5

6

AUTOTECH
A large car-parts company

1 John Smith started Autotech in 1960.

2

3

4

5

6

John Smith started Autotech in 1960.

It has a workforce of 2,500.

Autotech exports to over 12 countries.

It manufactures car parts.

It introduces one or two new components each year.

It employs about 35 people.

Green Fingers sells some of its products abroad.

It makes garden products.

George and James Hawkins began Green Fingers in the 1920s.

Green Fingers supplies the gardening industry.

It launches 12 new products a year.

Autotech provides components for the car industry.

Exercise 7. Complete the text with the correct form of verbs: *launch* *have* *begin*
manufacture *provide* *export*

Sonara _____ in 1972 near Turin. Today, it _____ mainly aircraft engines, but in the 1970s it also _____ the car industry with components. It _____ a workforce of 2,000. Sonara _____ 75 % of its engines to other European countries. Last month, it _____ a new type of engine which burns 15 % less fuel than other models.

Exercise 8. Match the sentence halves.

1. Panetti employs over 3,500 people,
2. It introduced four new products last year
3. It makes bread and,
4. Panetti only supplies its own shops;
5. It doesn't sell any of its products abroad,

- a) but it plans to expand into France.
- b) including 1,400 in its own retail outlets.
- c) including sandwiches and pies.
- d) many other bakery products.
- e) it does not make products for anyone else.

Exercise 9. Match these words and phrases from exercise 7 with a word or phrase from exercise 8 that has a similar meaning.

- 1 manufactures makes
- 2 provided
- 3 has a workforce of
- 4 export

5 launched

Exercise 10. Read the text and name the most important inventions of the company.

WHAT IS SO GOOD ABOUT SONY CORPORATION?

Since 1946 Sony has been committed to bringing the world the best in technology. It is a leader in consumer electronics. This innovation comes with Sony not just simply recreating products, but actually inventing new technology. In May 1960 - Sony was the first company to create the world's first direct-view portable television. The first model was developed based on Sony's extensive experience in radio technology. This device opened the door to personal television use.

Sony made further advancements and in 1962 produced the world's lightest and smallest all transistor television (the TV5-303). The advancements however did not stop there. In 1963 the world's first compact transistor VTR, the PV-100, launched. And in 1965 the world's first home-use open-reel VTR, the CV-2000, launched, paving the way to allow people to record and playback over 1 hour of video. There are so many new offerings that Sony has brought us, which we did not have before. The Walkman is probably one of the greatest inventions from Sony. Sony also brought us the world's first CD player also!

Sony is an electronics manufacturer that has brought us many great inventions, which have benefited humanity. Sony continues to innovate and be a world class leader in reliable electronics, with that elegant and ergonomic Sony style.

Make sentences about the most important events in the history of the company using the Sony timeline:

- 1946 - _____
- 1960 - _____
- 1962 - _____
- 1963 - _____
- 1965 - _____

Exercise 11. Put the verbs into Past Simple

1. I (meet) Managing Director at the airport at 7.00 in the morning.
2. We (take) our visitors to the plant.
3. On Wednesday I (fly) to Moscow for a conference.
4. My presentation (not go) very well.
5. Yesterday evening I (write) a proposal for an American company.

Exercise 12. Make questions to the answers.

1. What _____?
They arrived at 10 o'clock.
2. Why _____ the meeting?
I left the meeting because I had an urgent phone call.
3. When _____ the company?
She joined the company in 2015.
4. Who _____ at the conference?
We saw our colleagues from Moscow.
5. How long _____ with the visitors?
I spent 2 days with them.

ПР05. Инновации в производственной сфере. Описание товаров, их особенностей.

ПР05. Innovation in industry. Description of goods and their characteristics.

Exercise 1. Read the text and decide whether the statements are true or false.

1. Product is anything which is produced by people or machines and can be sold.
2. Products are always physical. You can touch or smell them.
3. Products can be delivered in the form of services or ideas.
4. Products cannot be virtual.

1. What is a product?

In general, a product is defined as a "thing produced by labor or effort" or the "result of an act or a process".

In marketing, a product is anything that can be offered to a market that might satisfy a want or need. In retail, products are called merchandise. In manufacturing, products are purchased as raw materials and sold as finished goods. Commodities are usually raw materials such as metals and agricultural products, but the term can also refer to anything widely available in the open market.

2. Goods, services, or ideas

Goods are a physical product capable of being delivered to a purchaser and involve the transfer of ownership from seller to customer.

A service is a non-material action resulting in a measurable change of state for the purchaser caused by the provider.

Ideas (intellectual property) are any creation of the intellect that has commercial value, but is sold or traded only as an idea, and not as a resulting service or good. This includes copyrighted property such as literary or artistic works, and ideational property, such as patents, appellations of origin, business methods, and industrial processes.

3. Product classification: tangible or intangible

A product can be classified as tangible or intangible.

A tangible product is a physical object that can be perceived by touch such as a building, vehicle, or gadget. Most goods are tangible products. For example, a soccer ball is a tangible product.

An intangible product is a product that can only be perceived indirectly such as an insurance policy. Intangible data products can further be classified into virtual digital goods ("VDG"), which are virtually located on a computer OS and accessible to users as conventional file types, such as JPG and MP3 files, and real digital goods ("RDG"), such as 3-D objects or presentational items.

Exercise 2. Fill in the missing words in the sentences: *practical, economical, functional, user-friendly, well-designed*

1. The new air conditioning system in our office is much cheaper than the old one. It is more

2. It took us quite a long time designing the new office furniture. Now it is very
3. The operating system on my office computer is easy to use. It is very
4. My new car is much easier to park. It is very ... for driving in the city center.
5. The new office equipment is exactly what we needed. It is very

Exercise 3. Read the tips on writing a good product description and match the heading to the paragraphs.

Make it Easy to Scan

Focus on the Product Benefits

Know Who Your Target Audience is

Use Power Words That Sell

Tell the Full Story

1. _____

The first step to writing product descriptions is to define your target audience.

You want to be able to define which features would be of most interest to your potential buyers.

As you are writing your product description, keep these questions in mind:

- How did this person arrive to your page?
- What are his or her interests, generally?
- Why would this person be interested in your Shopify store, specifically?
- How would this person describe the product to a friend?
- What features or benefits would interest this person the most?

By keeping these questions in mind as you write your product copy, you will be better able to write a product description that sells.

2. _____

As a business owner, you are understandably excited to share all of the qualities of your products. You want to show that your product has the best features and most unique specs.

The buyer, however, is not necessarily interested in the mundane features of the product. Instead, they want to know how it can benefit them.

A product feature is a factual statement about the product that provides technical information. A product benefit, on the other hand, tells **how the product can improve the buyer's life**.

3. _____

A good product description should give all relevant details, convince the buyer of its benefits, and pack an emotional punch.

Emotions influence buyer behavior, so your product description is the perfect place to elicit emotions.

How do you do this?

By filling in any gaps that potential buyers may have about the product.

4 _____

There are certain words and phrases that naturally elicit an emotional response in humans. Luckily for Shopify store owners, this also increases sales.

By being mindful of these words and phrases, you can more easily convince your customers to take the leap and make the purchase.

5 _____

People have short attention spans and read only about 16% of what's on the page. So you have to make your descriptions super scannable.

As in, the buyer is able to find exactly the information he or she wants without wasting time looking through other pieces of information.

Make your product descriptions easy to scan by including bullet points, short paragraphs made up of just a few sentences each, lots of white space, and different size fonts.

Exercise 4. Describe the company product (service) and activities using the information below.

1 Company: Translations R Us

Product: Electronic Translation Dictionary

What it does: Translates 45 languages into English, Chinese, Japanese and Spanish

2 Company: Colorado Hot Air, Inc.

Product: Hot Air Balloons

What it does: Takes people on flights up to 10,000 feet (about 3100 meters).

3 Company: Yoga Sensation

Service: Yoga Instruction

What is provided: Yoga classes at all levels worldwide for corporations and resorts.

4 Company: The Pool Doctor

Service: Swimming Pool Maintenance

What is provided: Cleaning, servicing and repair of swimming pools. Everything from small private pools to huge resort and country club pools.

Exercise 5. Match the adjectives with the opposites above.

1. unreliable
2. large
3. heavy
4. unpopular
5. short

a cheap

b unattractive

c fast

d boring

e bad

Fill in the gaps in the sentences.

1. The consumers like *Margin*, it is very _____.
2. But they think they have to pay a lot, because the product is _____.
3. Most people say the bottle is nice, it looks _____.
4. Some people think the packages is not _____, it is heavy.
5. The product meets customers' expectations. It is _____.

Exercise 6. Divide the words into three groups: *metal* *big* *square* *tiny* *glass*
rectangular *wood* *huge* *triangular*

Shape:

Material:

Size:

IP06. Инновации в производственной сфере. Анализ рыночной продукции и конкурентоспособности товаров.

IP06. Innovation in industry. Product analysis and the competitiveness of goods.

Exercise 7. Read the text about product development.

The development stages of a new product

Before a product can embark on its journey through the four *product life cycle stages*, it has *to be developed*. New product development is typically a huge part of any manufacturing process. Most organizations realize that all products have a limited *lifespan*, and so new products need to be developed to replace them and keep the company in business. Just as the *product life cycle* has various stages, new product development is also broken down into a number of specific phases.

Developing a new product involves a number of stages which typically center on the following key areas:

The original idea: Every product has to start with an original idea. In some cases, this might be fairly simple, basing the new product on something similar that already exists. In other cases, it may be something revolutionary and unique, which may mean *the idea generation* part of the process is much more involved.

Market Research: An organization may have plenty of ideas for a new product, but once it has selected the best of them, the next step is to start re-searching the market. This enables them to see if there's likely to be *a demand* for this type of product, and also what specific features need to be developed *in order to best meet the needs* of prospective customers.

Design and Development of the Product: The next stage is *the design and development* of the product. Prototypes may be modified through various design and manufacturing stages in order to come up with a finished product that consumers will want to buy.

Product Trials (Testing): Before most products *are launched* and the manufacturer spends a large amount of money on *production and promotion*, most companies will test their new product with a small group of actual consumers. This helps to make sure that they have *a viable product* that will be *profitable*, and that there are no changes that need to be made before it's launched.

Analysis: Looking at the feedback from consumer testing enables the manufacturer to make any necessary changes to the product, and also decide how they are going to launch it to the market. With information from real consumers, they will be able to make a number of *strategic decisions* that will be crucial to the product's success, including what price to sell at and how the product will be marketed.

Launch of the Product (Introduction): Finally, when a product has made it all the way through the new product development stage, the only thing left to do is *launch* it to the market. Once this is done, good product life cycle management will ensure the manufacturer *makes the most of all their effort* and investment.

Thousands of new products go on sale every year, and manufacturers invest a lot of time, effort and money in trying to make sure that any new products they launch will be a success. Creating a profitable product isn't just about getting each of the stages of new product development right, it's also about managing the product once it's been launched and then throughout its lifetime.

Exercise 8. Fill in the gaps using the words in italics from the text.

1. Without a proper ... it is impossible to ... a product successfully.
2. It is essential for any product to be
3. Companies spend a huge amount of money on of a new product.
4. Prototypes may be modified through variousstages.
5. Any new product has to start with ...

Exercise 9. Answer the questions:

1. What is the most important stage in the development of a new product?
2. How many stages are there in the product development process?
3. On which stage the companies can realize that their new product may be of demand among prospective customers?
4. How can companies know that their new product will be profitable?
5. How can you explain the statement "to create a profitable product"?

Exercise 10. Choose any product and prepare to give a brief product review.

- Your product review should include:
- a description of the product
 - the way the product is used
 - the cost of the product

- a comparison with other, similar products
- a recommendation to buy or not buy the product
- a reason for your recommendation
- a rating on a scale of 1-5

PRODUCT REVIEW

PRODUCT:

PURPOSE:

AUDIENCE:

USE:

COST:

COMPETITORS:

RECOMMENDATION:

REASON:

RANKING

PP07. Дизайн и спецификация товара. Описание дизайна и спецификации товара.

PP07. Design and product specification. Description of design and product specifications.

Exercise 1. Read the text and correct wrong statements:

1. Design is something related to fashion and style.
2. A good design begins with a good idea.
3. Design is not connected with people's quality of life.
4. Designers try to realize only sensible ideas.
5. Scientists and manufacturers don't have anything in common with designers.

WHAT IS DESIGN?

Design is everywhere. The word "design" means different things to different people. One definition given by designer Richard Seymour is 'making things better for people'. The design activity is focused first and foremost on human behaviour and quality of life. It transfers any idea into a blueprint for something useful, whether it's a car, a building, a graphic, a service or a process.

Scientists can invent technologies, manufacturers can make products, engineers can make them function and marketers can sell them, but only designers can combine insight into all these things and turn a concept into something that's desirable, viable, commercially successful and adds value to people's lives.

A good design begins with the needs of the user. No design, no matter how beautiful and ingenious, is any good if it doesn't fulfill a user need. Finding out what the customer wants is the first stage of what designers do. The designer then builds on the results of that inquiry with a mixture of creativity and commercial insight. Different designers use different methods-combining market research, user testing, prototyping and trend analysis. These methods lead to innovative products and services. Designers learn that ideas that may seem strange are worth exploring and that the 'common-sense' solution is not always the right one.

Exercise 2. Choose one of the products and write its description using the words and the outline

Words: *elegant functional futuristic handmade innovative retro stylish simple mass-produced traditional up-to-date streamlined ergonomic durable easy to use*

Useful phrases:

Visual appeal : It looks like (a design from the 2010s)

Material: It has a metal/wooden/golden (top/side/base)

Features: It has several (qualities / special features)....

It has a unique feature...

One of its weak/strong points is...(that it is very difficult/easy to use because...)

Use: It is designed for (opening/keeping)...

It is used for.....

ПР08. Дизайн и спецификация товара. Характеристика и сравнение дизайна различных товаров, представленных на современном рынке. Тест 1.

ПР08. Design and product specification. Description and comparison of design of various products in the modern market. Test 1.

Exercise 3. Complete this presentation introduction with the words: *talk about look at points of view questions brief finally hear act as go along*

Good afternoon and thank you for making the effort to be here with us today. My name's John Smith and I'm responsible for marketing. What I'd like to do today is 1 _____ our recent product promotion campaign. This 2 _____ talk will hopefully 3

_____ a springboard for discussion. I'm going to 4 _____ the marketing campaign from three 5 _____ : firstly, the customers; secondly, the financial institutions; and 6 _____, the shareholders. If you have any 7 _____, just interrupt me as I 8 _____. Your point of view may well be different, and we'd like to 9 _____ from you.

Exercise 4. Complete the following presentation common expressions and phrases:

- i. Today we will _____ at ...
- ii. I'd like to _____ by ...
- iii. This _____ us to the next _____ ...
- iv. Let me _____ you an example ...
- v. A case in _____ is ...
- vi. On _____ whole ...
- vii. Let me end by _____ ...

Exercise 5. Choose one of the products given below and complete the product design specification. Invent any additional information you need.

PRODUCT DESIGN SPECIFICATION

Product name:

Product description:

Product performance:

Ergonomics features:

Dimensions:

Design:

Safety:

Product 1:

OUTDOOR HEATER

Function: to heat the air outside a building

- gives a lot of heat for 26 hours
- powered by propane gas
- easy to regulate the heat
- light and easy to move
- easy to clean
- can be used in all weathers

Product 2:

BABY MONITOR

Function: to check the health of a sleeping baby

- works up to a 100-metre range
- powered by mains or battery

- low battery indicator
- adjustable volume
- has a belt clip and also a stand
 - easy to use and very light

Exercise 6. Read a report on two mobile phones and answer the questions.

1. What mobile phones are compared?
2. Do these phones have any similar features? What are they?
3. What are their peculiarities which differ these models from each other?
4. How many parts does the report consist of? What are they?

Two models of mobile phones the Nokia 6230i and the Samsung SGH-D500 are compared in this report. We have studied their features and found out that they have some similarities and differences.

First, the Samsung SGH-D500 is a little cheaper than the Nokia 6230i. It is very stylish and one of the most popular models on the market. Its features include a 1.3 Mp camera and music player. However, it does not have a memory card so you can store only a limited amount of music tracks and photos on this phone.

Second, the Nokia 6230i is small and light. It weighs only 99g and is very functional. It has a 1.3 Mp camera like the Samsung phone, but it also has a memory card. The panel of the camera can be replaced easily. In addition, covers are available in five different colours.

In conclusion, test results showed that the Samsung SGH-D500 was not very good in capturing movement. Despite this fact both the Nokia and Samsung models are very good buys.

Exercise 7. Write a report on two hand dryers.

AYT

- * uses principle of evaporation
- * may take up to 44 seconds to dry hands
- * warm air flows out at moderate speed
- * most of water removed by slow evaporation
- * bacteria are not filtered out of washroom air
- * unfiltered air blows around the room
- * machine starts when user presses start button
- * user usually has to press start button several times
 - * user often walks out with damp hands, wiping on clothes to dry

Dyson

- * uses process of scraping water off hands like windscreen wiper
 - * dries hands in 10-12 seconds
 - * cold air forced out at high speed
 - * long ultra-thin apertures run along two blades
 - * filter removes over 99.9% of bacteria from air* filtered air stays within the machine, not around room
 - * uses 80% less energy than conventional dryer

- * hands completely dry after use
- * turns on automatically when user inserts hands

2. Научная коммуникация

Module 2. Scientific Communication

ПР09. Предоставление исследовательского проекта. Форма заполнения заявки с описанием исследовательского проекта.

ПР09. Submitting a research proposal. Application form with a description of the research project.

Exercise 1. Read the following extract from a website. Your colleague has decided to apply to SARF and has downloaded an application form. Look at the list of sections on the form (1-10) below and match each one to your colleague's notes on the information she needs to provide.

The Sheridan Australian Research Fellowship (SARF) aims to develop science in Australia by attracting outstanding scientists in their field to continue their research in an Australian university or research institution. SARF fellowships are awarded to individual scientists with future potential for leadership in their field. Successful applicants receive a 5-year grant covering salary, travel and relocation costs.

1. Applicant	a) an explanation of how I'll do the research and why it is important
2. Current appointment and address	b) a short description of what I'll research
3. Location of proposed study	c) a statement from a senior researcher explaining why I'm a suitable applicant
4. Sponsor's recommendation	d) how much I plan to spend on my research
5. Departmental support	e) the job I do now
6. Project title	f) the name of someone to support my application
7. Project summary	g) what I'll call my research
8. Details of proposed research	h) permission from my head of faculty to use his/her resources
9. Budget	i) where I plan to study
10. Nominated referee with personal knowledge of applicant	j) my personal info

ПР10. Предоставление исследовательского проекта. Варианты исследовательских проектов и их особенности в современном сообществе.

ПР10. Submitting a research proposal. Types of research projects and their features in the modern community.

Exercise 2. Match each highlighted section in the projects summary (A – F) to the correct function (1 – 6) from the list below.

PROJECT SUMMARY

Provide a brief summary of aims, significance and expected outcomes of the research plan

A 3-D odour-compass for odour-detecting robots

Odour-sensing robots offer many benefits over the current use of animals in similar roles, including safety, efficiency and durability. [A] However, the robots which have been developed to date are limited by the fact that they can only accurately detect and navigate towards odour plumes if they are within direct 'sight' of the chemical source. Clearly, in real world situations, obstacles may well impede the robot's detection ability, and at present, odour-sensing robots are therefore only of limited use. [B] The proposed research will concentrate on developing a robot which is able to gather readings in three dimensions and therefore overcome the limitations of current models in odour-detection. [C] This technology will make robots a more effective substitute for animals.

[D] This research aims to develop existing robotic technology to create a three-dimensional (3-D) odour compass to be used as a navigation tool in searching for an odour source. [E] This will then be tested experimentally in simulated environments where wind direction is not stable or where obstacles interfere with odour distribution. A second stage in the research will be to develop the robot's environmental sensors, thus allowing it to safely negotiate the terrain to reach the source of the odour. [F] This should produce a robot which is able to both detect and move to the source of an odour, even on difficult terrain.

1. State the aims of your research.
2. Define what the problem is.
3. Explain why your topic is worth researching.
4. Say what the expected outcomes of the research are.
5. Outline the procedures you will follow.
6. Outline how you will limit your investigation.

Exercise 3. Complete another project summary using the following correct word or phrase:

aims to however the initial phase
the proposed research the study will indicate

Consumer interest in wines produced in organic vineyards has increased significantly in the last few years (1) _____, to date it is unclear whether these production methods actually improve soil or grape quality. (2) _____ will be the first phase of a long-term study on a New Zealand vineyard. These results (3) _____ whether methods of viticulture improve grape quality.

The research (4) _____ investigate the effects of organic agriculture on soil and grape quality. (5) _____ will consist of two treatments, organic and conventional (the control), each replicated four times in a randomised, complete block design. All organic practices will follow the standards set out by the Food Standards Australia New Zealand (FSANZ).

_____ (6) will assess soil quality using physical, chemical and biological indicators over six years. The next phase will then assess the physiology of the vines.

Exercise 4. Match a sentence (a – e) below to the correct paragraph (1 – 5) of the full project summary.

NSF GRANT PROPOSAL FOR STABILIZATION AND OPTIMAL CONTROL OF DYNAMIC SHELL MODELS

[1] _____ We intend to investigate problems related to stabilisation and optimal control of dynamic shell models where control actions and sensing are put into place via smart materials technology.

[2] _____ The principle model considered in this proposal is a three-dimensional structural acoustic interaction with curved walls, which is modeled by shell equations. This model occurs in

the context of decreasing noise or pressure entering an acoustic chamber (e.g. airplane's cabin) and generated by an exterior source.

[3] _____ Thus mathematical investigation related to control problems of shell equations is challenging from a mathematical point of view and calls for the introduction of new tools and new techniques for the analysis and computations connected to the problem.

[4] _____ Two approaches will be considered. First, piezoelectric shell's modeling tracked by past researchers and a second centre on piezoelectric patches attached to the curved wall. These approaches will result in two different control models. Rigorous mathematical analysis of the problem, including comparative analysis, followed by numerical computations and experimental verification of the models will represent the essential part of the project.

[5] _____ Thus we wish to combine the teams' expertise to generate results leading to progress in the field.

a) Both teams for this project already have knowledge in various features of the problems described.

b) However there are very few outcomes and the methods that were developed have broken down.

c) The need for mathematical analysis of shell models is in response to a variety of technological demands, which call for more complex models.

d) This project will focus on stabilization and optimal control, particularly with boundary controls, of systems concerning dynamical shells.

e) We propose to carry out collaborative research between the French National Institute of Computer Science Research and the University of Ottawa in the general area of control theory for models illustrated by partial differential equations (PDEs).

ПР11. Участие в научной конференции. Описание форм участия в научных конференциях.

ПР11. Participating in a research conference. Forms of participation in scientific conferences.

Exercise 1. Read the following five extracts and then write which form (or forms) of communication (*an academic journal, a popular science magazine, a conference, a popular science book, an online forum or science blog, a newspaper*) each one comes from.

1) ... more people were pain-free when using the handheld device than those who had used an identical dummy device. Although the study by Lipton *et al.* (2010) has reliable results, there are some points to consider when putting these findings into context. Importantly, the results will need to be verified in larger trials that directly compare ...

2) Tea and coffee drinkers have a lower risk of developing type 2 diabetes, a large body of evidence shows. And the protection may not be down to caffeine since decaf coffee has the greatest effect, say researchers in *Archives of Internal Medicine*. They looked at...

3) ... can be rapidly generated by lentivirus-mediated transgenesis. RNAi also holds great promise as a novel therapeutic approach. This report provides an insight into the current gene silencing techniques in mammalian systems.

4) Hi! Has anyone had any experiences with nanoparticles sticking to glassware :-(? If so, does anyone know if there's a suitable silylation protocol to pre-treat the glassware to do something about this annoying non-specific adsorption? Thanks!

5) Animal and *in vitro* studies suggest that aspirin may inhibit breast cancer metastasis. We studied whether aspirin use among women with breast cancer decreased their risk of death from breast cancer. This was a prospective observational study based on ...

ПР12. Участие в научной конференции. Проведение научной конференции.

ПР12. Participating in a research conference. Conducting a scientific conference.

Exercise 2. Look at the online poster advertising a conference and complete the phrases below in accordance with the poster using the given words.

Keynote speakers

- Zoltan Szabo

European Institute of Malaria Research (EIMR)

- Mirembe Kabasomi

Makarere University, Kampala, Uganda

Preliminary Programme

A list of other invited speakers and preliminary session topics is currently being developed by the Conference Chair and will be announced in due course. Please check back for updates.

For further information about us see www.eimr.org

Online registration only

www.eimr.org/con7/registration

Registration is on a strictly first- come, first-served basis.

Application deadlines

1 April for abstract or poster presentation submissions

7 May for attendees

Registration fees

Academia – €450, Students – €350, Commercial/Industry – €650

basis course deadline keynote preliminary presentation registration (x2) strictly

1. application _____

2. on a _____ first-come, first-served _____

3. _____ speakers

4. online _____ only

5. poster _____

6. _____ programme

7. _____ fees

8. to _____ an abstract

9. in due _____

10. check back for _____

ПР13. Принципы составления и написания научной статьи. Анализ отрывков из научных статей по различным темам. Введение и отработка новой лексики, клише.

PP13. Writing a research paper. Analysis of extracts from scientific articles on various topics. Introduction of a new vocabulary.

Exercise 1. Tony is doing research into the panspermia hypothesis as part of a Master's degree in astrobiology. He has been investigating whether it is possible for bacteria and microorganisms to survive in an environment as harsh as the surface of Mars. He has been advised to organize the text of his introduction around five key questions. Match the beginnings to the endings of the questions.

- | | |
|---------------------|---|
| 1. What was I | a. approach the problem? |
| 2. Why was it | b. expect to know after doing the research? |
| 3. What was already | c. important? |
| 4. What did I | d. investigating? |
| 5. How did I | e. known about the subject of my research? |

Exercise 2. Read five extracts from the introduction to Tony's paper. Which question from Exercise 7 is each extract answering? Write the questions above the extracts.

1. _____

Such an extreme environment was thought to be uninhabitable, but microbial ecology studies reported the presence of microorganisms (Amaral-Zettler et al., 2002). Could the surface composition of Mars protect life against radiation?

2. _____

A number of studies have investigated different extreme Martian surface conditions on terrestrial microorganisms. Nicholson and Schuerger (2005) reported that the spores of *Bacillus subtilis* were able to survive for 19 days under Mars atmospheric pressure and composition. Saffary et al. (2002), however, found that survival decreased due to ...

3. _____

Potential habitability in the subsurface would increase if the overlaying material did play a protective role.

4. _____

For many years now, scientists have speculated about the possibility of life on Mars (Klein et al., 1976; McKay, 1997). The discovery of liquid water on Mars would increase its habitability ...

5. _____

We report here on our studies of protection by Rio Tinto Basin iron oxides and hydroxides on two microorganisms, *Acidithiobacillus ferrooxidans* and *Deinococcus radiodurans*, under simulated Mars surface conditions.

PP14. Принципы составления и написания научной статьи. Анализ различных частей научной статьи и их особенностей.

PP14. Writing a research paper. Analysis of various sections of a scientific article.

Exercise 3. Read an extract from the introduction of a paper about the ability of lichens and microbes to survive in deep space. Put the verbs into the correct form.

Recent advances in space technology (1)_____ (provide) the possibility of studying the

survival of different microorganisms in the harsh environment of space (*Demets et al., 2005; Baglioni et al., 2007*). So far, lichens (2) _____(be) the only organisms able to survive exposure to such extreme conditions (*Sancho et al., 2007; de los Rios et al., 2010*).

It is believed that, if sufficiently protected by meteorite-like material, microorganisms may also survive the journey through space. However, Brandstatter *et al.* (2008) (3) _____(report) that microorganisms embedded in 2 cm thick rocks on the outer surface of a re-entry capsule, simulating the entry of a meteorite, (4)_____ (not survive).

The aim of this work (5)_____ (be) to obtain further information on the resistance of rock-colonising microbial communities and lichens to outer space conditions, during the Biopan-6 flight of ESA on board a Russian Foton satellite.

Exercise 4. Complete the following summary on variables using the given words.

affects collecting controlled data dependent independent

How much a variable (1)_____ a relationship can be discovered by (2)_____ experimental (3)_____ on changes to the relationship as the variable is changed. In an experiment, there will be: one (4)_____ variable – this is the feature you are measuring; one or more (5)_____ variables – these are the variables which you change; one or more (6)_____ variables – these are not being tested and so they stay the same.

Exercise 5. Complete the lines below using the extract from the following research paper to help you.

A promising candidate among the different adsorbent materials are activated carbons. Through activation, highly porous materials can be prepared. Due to their high porosity, activated carbon materials are able to adsorb large amounts of hydrogen. Following adsorption, hydrogen molecules can be found at two possible locations: (1) on the surface of the adsorbent, or (2) as a compressed gas in the void space between adsorbent particles. (adapted from *Konowsky et al. 2009*)

Noun

Verb

Adjective

1. **compression**

2. compress

3. _____

4. _____

5. adsorb

6. _____

7. _____

8. activate

9. _____

10. _____

11. _____

Exercise 6. The gapped words below all describe physical or chemical properties of substances. The meaning of each word is given on the right. Complete the words with the correct vowels (a, e, i, o, u)

1. br_ttl_n_ss

how easily something can be broken

2. c_p_c_t_nc_

how well something holds an electrical charge

3. c_nc_ntr_t_n

how much of one substance is found in another

4. c_nd_ct_v_ty

how well something allows heat or electricity to go

5. d_ns_ty

how much mass a given volume of a substance has

6. fl_mm_b_l_ty

how easily something burns

7. l_m_n_nc_

how much light passes through or comes from a substance

8. m_ss	how much matter is in a solid object or in any volume of
9. p_rm__b_l_ty	how easily gases or liquids go through a substance
10. p_r_s_ty	how many small holes are in a substance
11. r__ct_v_ty	how easily a chemical substance reacts
12. s_l_b_l_ty	how easily something can be dissolved to form a solution
13. v_l_c_ty	how quickly an object is travelling
14. v_sc_s_ty	how thick a liquid is
15. v_l_m_	how much space is contained within an object or solid

Exercise 7. Complete the paragraphs from the results section of a paper using the following words and phrases in the box.

**as can be seen in considerably contrast to noticeably thicker resulted in a
longer while**

During the rapid heating, the Ni near the Ni/SiC interface reacted with the SiC, which resulted in carbon atoms moving into the Ni. The carbon atoms then separated onto the surface of the Ni during the cooling procedure, forming graphene layers (1)_____Fig. 1b. In (2) _____the graphene generated using single-crystalline SiC, the graphene synthesised by this process is (3)_____easier to remove from the SiC surface.

A slower heating rate (4) _____ process. As shown in Fig. 4, more carbon atoms were released into the Ni in a long process. Higher carbon concentration in the Ni produced a (5)_____carbon nanofilm on the Ni surface, (6)_____a lower carbon concentration reduced the thickness of the carbon nanofilm and formed graphene.

PP15. Презентация исследовательского проекта. Структура презентации в целом и исследовательского проекта, в частности.

PP15. Presentation of a research project. The structure of the presentation as a whole and the research project in particular.

Exercise 1. Write ‘who’, ‘why’, ‘what’ or ‘how’ next to each phrase. Check any vocabulary you don’t know.

1. On behalf of Mr Keane, may I welcome you to Jackson Inc. My name’s Jo Black and I’m responsible for ...

2. My purpose today is to ...

3. I’m going to develop three main points. First, ... Second, ... Third, ...

4. Let me introduce myself. I am ... I am a ...

5. I’ll pass round copies of my slides so you can make notes as I go through the presentation.

6. Before I continue, let me tell you something about myself.

7. Today I would like to give you a general overview of...

8. I’ve divided my presentation into three main points. I would like to begin with ...

9. So, I’ll be addressing three main points and the first one is going to be ... The second point will be ... And finally the last point is ...

10. I’m going to outline three proposals. Firstly, I’ll ... Then, I’d like to ... and finally ...

11. We can take two or three questions at the end of each point.

12. You don’t need to take notes as we’ll be handing out presentation booklets.

Exercise 2. Match each pair of phrases (1 – 8) from to their correct function (a – f) below.

Note that one of the functions may be expressed with three different pairs of phrases.

1. Good afternoon, everybody. / Welcome, ladies and gentlemen.
2. To start, thank you / I'd like to start by thanking you all for coming to my talk today.
3. I'm Milan Poborski and at present / My name is Milan Poborski and I'm a PhD candidate at Northumbria University.
4. I'm going to talk today / My talk today is about my recent research investigating ...
5. I'll begin by explaining / To start with, I'll explain briefly how T-cell responses...
6. After that, I'll / I'll go on to describe the alternative method I have been investigating ...
7. Finally, I will discuss / I'll conclude by discussing why this method could be useful as a way ...
8. I plan to talk for about 40 minutes, leaving plenty of time for / I will talk for about 40 minutes and then I'll answer any questions at the end of my talk.

- | | |
|---|-----------------------------------|
| a. Give instructions for asking questions | b. Greet the audience. |
| c. Introduce the topic of the presentation | d. Introduce yourself |
| e. Outline the structure of the presentation. | f. Thank the audience for coming. |

PP16. Презентация исследовательского проекта. Анализ различных проектов и обсуждение их сильных и слабых сторон. Написание теста по пройденному разделу.

PP16. Presentation of a research project. Analysis of various projects and discussion of their strengths and weaknesses. Test 2.

Exercise 3. Below are *five sentences* from the main part of a presentation. Match the beginnings (1 – 5) to the endings (a – e).

1. A number of potential vaccine types have been developed and
2. As I have already said,
3. As you can see from this image,
4. Let's begin by looking at the size of the malaria problem.
5. That's all I have to say about the vaccine itself,

- a. counting IFN- γ secreting cells has been the preferred method to date.
- b. using flow cytometry to detect MIG secretion gives us a more accurate way of measuring immune responses.
- c. I will be returning to those shortly.
- d. Malaria kills over one million people every year in 109 countries.
- e. so now I'd like to move on to looking at judging the response of the immune system to the vaccine.

Exercise 4. The underlined phrases in Exercise 3 help speakers to organize their presentation clearly and guide listeners through the information. Write the correct underlined phrase to complete the advice below.

Use:

- a. _____ : to introduce a new part of the talk
- b. _____ : to conclude one part of the talk and then begin another
- c. _____ : to refer back to an earlier part of the talk
- d. _____ : to refer forward to a later part of the talk
- e. _____ : to refer to a visual aid

3. Деловая коммуникация

MODULE 3. Business Communication

ПР17. Межличностные и межкультурные отношения.

ПР17. Interpersonal and Intercultural Contacts

When you observe people from a certain culture, some characteristics – such as dress and the way people greet each other are easy to see. Others are not so easy.

Culture is sometimes compared to an iceberg, some of which is visible, but much of which is difficult to see, or invisible.

Exercise 1. Look at the list of components of national culture, and place each one in one of the three categories:

A things which you can recognize quite easily

B things which you recognize only when you are very familiar with a culture

1. Beliefs
2. Family values
3. Language
4. Expectations
5. Food
6. Manners
7. Holidays and festivals
8. Rules of conduct
9. Greetings
10. Attitudes to the environment
11. Physical gestures
12. Work ethic
13. Roles of males and females
14. Art and architecture
15. Punctuality
16. Humour

Add any other elements which you think are important in defining a national culture.

ПР18. Межличностные и межкультурные отношения.

ПР18. Interpersonal and Intercultural Contacts

When you meet people for the first time, greet them politely and warmly. Use a mix of questions during the first conversation – try to discover what things you have in common. Be sensitive to the cultural background of the other person during the meeting.

Exercise 2. Nigel Hastings is a director of an intercultural consultancy. Listen to him talking about managing first meeting in different cultures and answer the questions.

1. Why does Nigel say it's important to manage first meetings well when working internationally?
 - a) Because it establishes positive relationship with international partners.
 - b) Because it helps to learn a foreign language.
2. What does he describe as the purpose of first meetings in the Arab world?
 - a) To start talking about business as soon as possible.
 - b) To get to know your partner and build trust.
3. When he goes to China, how does he usually start small talk?
 - a) He talks about Shanghai.
 - b) He talks about a person's roots.
4. Why is asking 'open questions' important?
 - a) It's a way to create an emotional connection with your partner.
 - b) It's a way to learn about your partner's background.

Audio transcript

Interviewer: How important is it to handle first meetings well?

Nigel: Very important, and breaking the ice can be more difficult due to language differences, cultural unfamiliarity and so on. But when we have to work with people across the globe we don't see very often, it becomes critical to manage that first moment positively, to get the relationship going.

Interviewer: In terms of cultural differences in first meetings, what have you experienced?

Nigel: Well, being British I do a little small talk about the weather, how I travelled to the place, some of the challenges I faced on the way, and a touch of humour but quite quickly getting down to business because I don't want to take up too much of the time of the individual I'm talking to. In other cultural contexts, I've learned you need to take a different approach. For example, working in the Arab world, I think the concept of 'small' talk is less relevant because those opening social moments are critical in a relationship-orientated culture. And showing respect for the local culture, saying positive things about what you've seen, what you know, why you appreciate where you are, showing and receiving hospitality. There's a slower transition into business, more exchange on the personal level before getting down to the task. And that builds trust. All this can be seen as slow and time-wasting in a UK context.

Interviewer: What kinds of positive things do you say?

Nigel: I think when I go to China I often find myself asking where someone comes from because I know quite a lot about China and it's an opportunity to show a kind of interest which I

have for the country and a sensitivity to that person's roots. So I would say something like 'Oh, you come from Shanghai, I've been there a couple of times and I love it there.' This creates a positive impression. And it's true. I do love it there!

Interviewer: Does asking open questions generally work across all cultures to stimulate conversation?

Nigel: I think it definitely does. The open questions beginning with 'what' and 'how' open people up, give you an opportunity to listen to people speak, understand their interests ... so questions are very important as a way to create common ground.

Interviewer: Is this one of the secrets of success – creating common ground?

Nigel: I think so. You can ask an open question, listen to the response and then connect your experiences. You have the potential to build common ground also at an emotional level in terms of going through some common challenges. However, you can also ask closed questions to show interest in something around you.

Exercise 3. It is important to find things in common when responding to what people tell you. Match each comment to a response.

Comments

1. I've been to Italy a few times.
2. I studied mechanical engineering at uni.
3. I grew up just outside Madrid.
4. So these are the new offices.
5. Business is a bit challenging at the moment.

Responses

- a) I know the city quite well but not the region around it.
- b) Really? My brother did something similar and now works in construction.
- c) Indeed, but I read that things should be improving by the year end.
- d) Impressive. Looks like a nice place to work. Our offices are very different.
- e) Have you? Me too. I love the South.

ПР19. Проведение переговоров
ПР19. Negotiations

Exercise 4. Which of the actions a)-g) correspond to the negotiation stages 1-7?

1. build rapport
2. agree on a procedure
3. make proposals and counter-offers
4. probe with questions
5. enter the bargaining zone
6. resolve any areas of conflict

7. conclude the negotiation

- a) make concessions
- b) find things in common
- c) celebrate the deal!
- d) state your opening position
- e) decide who will speak first
- f) clarify anything you don't understand
- g) call for a time-out

Exercise 5. In a negotiation each team member must play a specific role. Complete the team roles 1-6 below using appropriate pairs of words a)-f).

- 1. Decision-maker: overall strategy and has the final
- 2. Facilitator: and provides of their team's position.
- 3. Number-cruncher: down key figures and does the
- 4. Chief negotiator: the main negotiations and acts as
- 5. Observer: the other team's behaviour and looks for signs of
- 6. Ideas-generator: deadlocks by coming up with creative

- a) breaks + solutions
- b) monitors + movement
- c) formulates + authority
- d) notes + calculations
- e) conciliates + clarification
- f) conducts + spokesperson

ПР20. Проведение переговоров

ПР20. Negotiations

Exercise 6. Match each of the techniques 1-6 from the previous section to a comment a)-f) demonstrating this technique in action.

1 2 3 4 5 6

- a) I changed my mind because I believed what he said – he didn't hide anything from me.
- b) I felt they really understood my needs so I was happy to accept their proposal.
- c) I had to agree. The facts spoke for themselves.
- d) She convinced me to join the project because I enjoy working with her.
- e) He offered to help me out on my project so I agreed to support him at the meeting.
- f) I supported her idea because I could see she really believed in it 100%.

ПР21. Контракты и соглашения

ПР21. Contracts and Agreements

Exercise 7. Match the types of contract 1-6 to their definitions a)-f).

- 1. a lease
- 2. an employment contract
- 3. a contract of sale
- 4. an insurance policy
- 5. a credit agreement
- 6. a software license

- a) a contract between a seller (or vendor) and a buyer (or purchaser)
- b) a contract between an employer and employee
- c) a contract between an insurance company and a person who pays for the insurance
- d) a contract which allows one party (the tenant) to use the land or property of the other party (the landlord) for a specified period of time
- e) a contract which allows someone to use a computer program
- f) a legal contract in which a bank agrees to loan a customer a certain amount of money for a specified amount of time

ПР22. Контракты и соглашения

ПР22. Contracts and Agreements

Exercise 8. Complete the phrases below with the following verbs:

breach complete draw up extend renew sign terminate

1. sign a contract ⇒ to put your signature on a contract to show that you agree to it
2. _____ a contract ⇒ to prepare/write a contract
3. _____ a contract ⇒ to break one or more of the terms of the contract
4. _____ a contract ⇒ to end a contract before the official end date
5. _____ a contract ⇒ to fulfil all the terms of the contract
6. _____ a contract ⇒ to make a contract continue for a longer time
7. _____ a contract ⇒ to sign a new contract (when the previous contract has finished)

Now complete the sentences below:

- a) Our lease expires at the end of the year, but we were able to _____ the contract by another year.
- b) We cannot sell the goods to a third party. If we do this, we will _____ the contract.
- c) We were very happy with the service so after our contract ended, we decided to _____ it.
- d) The construction firm couldn't _____ the contract because of the bad weather.
- e) We can _____ the contract if we notify the other party one month in advance.
- f) In order to make everything legal, the lawyers decided to _____ a contract.

МЕЖДУНАРОДНАЯ ПРОФЕССИОНАЛЬНАЯ КОММУНИКАЦИЯ:

методические рекомендации

по дисциплине

«Международная профессиональная коммуникация»

Н.А. ГУНИНА

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

Полная версия пособия доступна по ссылке

Путилов, А. В. Коммерциализация технологий и промышленные инновации : учебное пособие / А. В. Путилов, Ю. В. Черняховская. — Санкт-Петербург : Лань, 2021. — 324 с. — ISBN 978-5-8114-3371-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/169312> (дата обращения: 28.02.2022). — Режим доступа: для авториз. пользователей.

Логин: ulgtu2019@yandex.ru

Пароль: 778452asd

**А. В. ПУТИЛОВ,
Ю. В. ЧЕРНЯХОВСКАЯ**

КОММЕРЦИАЛИЗАЦИЯ ТЕХНОЛОГИЙ И ПРОМЫШЛЕННЫЕ ИННОВАЦИИ

Учебное пособие



САНКТ-ПЕТЕРБУРГ
МОСКВА
КРАСНОДАР
2021

ББК 65.290я73

П 90

Путилов А. В., Черняховская Ю. В.

П 90 Коммерциализация технологий и промышленные инновации: Учебное пособие. — СПб.: Издательство «Лань», 2021. — 324 с.: ил. — (Учебники для вузов. Специальная литература).

ISBN 978-5-8114-3371-1



Пособие посвящено изучению вопросов промышленной коммерциализации технологий — деятельности, направленной на получение дохода от использования результатов научных исследований и разработок. Приводятся примеры предварительного комплексного изучения финансовых и рыночных перспектив научных разработок и доказательства их будущих экономических и технологических преимуществ, прогнозирование развития рынков, сравнение с лучшим опытом и оценка конкурентных преимуществ разработок и технологий. Настоящее учебное пособие предназначено для студентов магистратуры по направлениям подготовки: «Системный анализ и управление», «Экономика», «Менеджмент», «Бизнес-информатика».

ББК 65.290я73

Рецензенты:

А. И. АГЕЕВ — доктор экономических наук, профессор, генеральный директор Института экономических стратегий РАН; *А. В. ФОМИНА* — доктор экономических наук, профессор, генеральный директор ЦНИИ «Электроника».

Обложка

Е. А. ВЛАСОВА



© Издательство «Лань», 2021

© А. В. Путилов, Ю. В. Черняховская, 2021

© Издательство «Лань»,

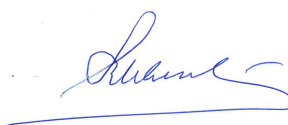
художественное оформление, 2021

ПРЕДИСЛОВИЕ

В последнее время в высокотехнологичных отечественных отраслях, прежде всего в атомной отрасли, происходят существенные экономические сдвиги, характеризующиеся переходом к долгосрочной инновационной политике и внедрению технологий цифровой экономики, связанных с освоением технологий следующего поколения и развитием новых инновационных подходов к созданию принципиально новой продукции и оказанию высокотехнологичных услуг. Книга «Коммерциализация технологий и промышленные инновации» является универсальным пособием, которое может быть использовано при обучении студентов как технических, так и управленческих специальностей, связанных с высокими технологиями, например с созданием ядерных реакторов, производством ядерного топлива, разработкой систем управления АЭС. Актуализация образовательного контента, ярко проявленная в данном издании, — это технология выявления наиболее одаренных учащихся в режиме реального времени, которая позволяет производить отбор из больших коллективов потенциальных кандидатов на «инновационную образовательную спираль»: последовательное чередование новых курсов дисциплин, ориентирующих обучающихся стремиться к инновационным решениям в дальнейшей профессиональной деятельности.

Развитие цифровой экономики является одним из ключевых направлений государственной политики Российской Федерации, обусловленной модернизацией производственных отраслей и проникновением цифровых технологий в экономические процессы на всех уровнях производства. Перед ведущими госкорпорациями «Росатом» и «Ростех», как одними из участников Центра компетенций программы «Цифровая экономика Российской Федерации», стоит задача по интеграции всех предложений в план мероприятий по цифровой трансформации промышленности и внедрению «сквозных» технологий. Например, в атомной отрасли был разработан новый инструментарий управления знаниями на основе больших данных с использованием нейросетей и машинного обучения. На ряде конференций в последнее время было отмечено, что будущее развитие цифровых платформ и цифровых технологий заключается именно в управлении результатами интеллектуальной деятельности и их коммерциализации. Расширение использования цифровых технологий позволит более активно вовлекать интеллектуальную собственность в хозяйственный оборот. Система управления знаниями выходит на новый уровень: постоянно развиваются и адаптируются продукты и технологии под меняющиеся условия рынка и цифровые особенности производства, что делает их еще более востребованными. Материалы данного издания позволят расширить использование новых технологий на новые области знаний и инженерные системы нового поколения.

*Заместитель Президента РАН,
доктор экономических наук,
член-корреспондент РАН*



В. В. Иванов

ВВЕДЕНИЕ В ЦИФРОВУЮ ЭКОНОМИКУ

Переход к цифровой экономике — общемировая тенденция, поэтому существует большое число определений этого нового экономического уклада. По определению Всемирного банка цифровая экономика — система экономических, социальных и культурных отношений, основанных на использовании цифровых информационно-коммуникационных технологий. В отечественной практике чаще всего используется следующее определение, приведенное в указе Президента России от 9 мая 2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы»: «цифровая экономика — хозяйственная деятельность, в которой ключевым фактором производства являются данные в цифровом виде, обработка больших объемов и использование результатов анализа которых по сравнению с традиционными формами хозяйствования позволяют существенно повысить эффективность различных видов производства, технологий, оборудования, хранения, продажи, доставки товаров и услуг». Очевидно, что коммерциализация технологий и реализация промышленных инноваций должны учитывать эти макроэкономические особенности ближайшего будущего.

Наступающую эру можно охарактеризовать как эпоху цифровых платформ, вытесняющих с рынка и из сферы производства неэффективных посредников и заменяющих их эффективными алгоритмами. Существуют разновидности этих платформ, можно различить две большие группы: электронные торговые площадки (например, Uber, Avito, Cainiao и др.) и инструменты автоматической внерыночной координации совместной деятельности (виртуальные офисы, интернет-взаимодействие отдельных производств, а также цифровые инструменты для более крупных хозяйственных единиц), производства товаров, оказания услуг. Существует тезис о том, что подключение к цифровым платформам дает субъектам рынка такие конкурентные преимущества, что по мере захвата этими платформами национальных и мировых рынков в полной мере начнет срабатывать принцип «кто не с нами, тот банкрот». Можно выделить американскую (англосаксонскую) и китайскую (восточноазиатскую) модели построения цифровой экономики и провести сопоставление этих моделей: по способам инвестирования, роли государственного и частного капитала, доле провальных и сверхприбыльных стартапов за последние несколько лет, по динамике роста и суммарному экономическому эффекту от «цифрового» сектора. Сопоставления явно свидетельствуют в пользу китайской модели, а нехитрая экстраполяция показывает, что если ничего не предпринимать в сфере высоких технологий прежде всего, то к середине 30-х гг. нынешнего столетия эти два мировых лидера по цифровым технологическим платформам захватят не менее 90% совокупных рынков планеты. И если не ответить на вызов чем-то «несимметричным», то места для России в дележе этого рыночного пирога может и не остаться.

Можно и нужно продемонстрировать математическую ясность и логику в осмыслении темы цифровых платформ, постоянно искать ответы на несколько принципиальных вопросов «цифровизации» реального сектора экономики. Следует выявлять схематическую модель каждой такой технологической еди-

ницы в «идеальной» полноте необходимых функций и элементов, искать ответ на вопрос об оптимальном размере формируемых цифровых платформ. В некоторых работах считается, что пределом целесообразной масштабируемости является отраслевая платформа, причем отрасль должна рассматриваться в контексте глобального разделения труда. Важен также вопрос о границах охвата цифровыми платформами того или иного сегмента экономики. Уже существующие современные цифровые платформы в состоянии глубоко автоматизировать производства и рынки тех товаров и услуг, характеристики которых поддаются жесткой количественной или качественной параметризации. Неподатливыми для этой платформенной логики пока остаются сектора, продукты которых обладают значимыми «невербализуемыми» свойствами и спрос на которые существенно зависит от персонального вкуса того или иного потребителя. Но по мере развития возможностей искусственного интеллекта новые типы цифровых платформ будут захватывать все новые и новые области «неформализуемой» экономики.

Можно дать высокие оценки наиболее продвинутым современным платформам Alibaba Group (Китай) и Abbyu SmartCAT (Россия): по мнению многих экспертов, они уже сегодня на 70% соответствуют перспективным платформенным моделям. Кроме того, важен вопрос о корпоративных платформах глобальных производителей (например, авиастроительной корпорации Boeing), масштабы которых сильно повышают барьер выхода на рынок платформ собственных отраслевых разработок. Большие данные и результаты их анализа дают основу развития нового экономического уклада, поэтому технологии искусственного интеллекта или создаваемые квантовые компьютеры могут в перспективе дать кардинальные преимущества в этой экономической гонке своим разработчикам.

В настоящее время в нашей стране высказываются идеи относительно путей «платформенной» оптимизации системы госзакупок и государственных информационных систем, предложено несколько конкретных проектных решений. В обсуждениях государственного участия в создании цифровых платформ часто звучит предостережение: рассматривать государство в качестве субъекта интересов, решений и действий — значит самим запутываться в мифах и запутывать других. Главным предметом дискуссий является проблема субъекта «информационного реформирования» государства: как достичь синергии несочетаемых, казалось бы, в одном лице качеств — компетентности, заинтересованности и полномочий.

В ряде работ делаются попытки корректно ввести понятие «платформа», но указывается необходимость более глубокого переопределения этой экономической сущности. Наиболее целесообразно не изобретать велосипед, а «встать на плечи гигантов»: восстановить систему смысловых координат для обсуждения, обратившись к наработкам классиков экономических и общественных наук последних десятилетий. Всем нам нужно осмыслить фундаментальный сдвиг мировой экономической мысли в понимании, что такое хозяйственная деятельность по своей сути. Это позволит разработать понятие технологической платформы в контексте современного понимания. Тогда и в теоретических разработках, и в практических реализациях идеи «цифровых платформ» в нашей стране может появиться реальный шанс возглавить «миро-

вой мейнстрим». Догоняющая стратегия обречена на неудачу, и только понимание глубокой значимости теории дает возможность «ухватить тренд и забежать вперед». Все вышесказанное дает основание утверждать, что мы все находимся в преддверии масштабных экономических изменений и на конкретных примерах следует искать ответы на поставленные жизнью экономические вопросы. Коммерциализация технологий и промышленные инновации — это та сфера, которая наиболее перспективна при рассмотрении ближайшего экономического будущего.

Государственный и корпоративный подход к разработкам информационных систем производственных предприятий в эпоху цифровой экономики

В течение 2017 г. в нашей стране было выпущено несколько документов наивысшего уровня, в которых цифровая экономика была обозначена совершенно конкретно. Это весенние указы Президента Российской Федерации:

- от 9 мая 2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы»;
- от 13 мая 2017 г. № 208 «О Стратегии экономической безопасности Российской Федерации на период до 2030 года».

Наконец, летом 2017 г. программа долгосрочного экономического развития «Цифровая экономика Российской Федерации» была утверждена распоряжением Правительства России от 28 июля 2017 г. № 1632-р.

Основные цели этой программы:

- создание экосистемы цифровой экономики Российской Федерации, в которой данные в цифровой форме являются ключевым фактором производства во всех сферах социально-экономической деятельности;
- создание необходимых и достаточных условий институционального и инфраструктурного характера для создания и (или) развития высокотехнологичных бизнесов;
- повышение конкурентоспособности на глобальном рынке как отдельных отраслей экономики Российской Федерации, так и экономики в целом.

При формировании данной программы подразумевались следующие основные уровни цифровой экономики:

- среда (регуляторика, инфраструктура, кадры, информационная безопасность);
- платформы и технологии (где формируются компетенции для сфер деятельности);
- рынки и отрасли экономики (сферы деятельности), где и осуществляется взаимодействие конкретных экономических субъектов.

Отмеченная в программе «Цифровая экономика Российской Федерации» роль государства — это создание фундамента развития цифровой экономики путем развития компонентов платформ и среды:

- институтов цифровой экономики (нормативное регулирование, кадры и образование, исследовательские компетенции и технологические заделы);
- инфраструктуры цифровой экономики (информационная инфраструктура, информационная безопасность).

По мнению разработчиков программы, которое поддержано на правительственном уровне, рынки и отрасли экономики самоорганизуются в новой среде и вокруг создаваемых и развивающихся цифровых платформ.

Базовые направления развития программы «Цифровая экономика Российской Федерации» (табл. 1) обеспечивают комплексный подход к разработке принципиальных основ цифровой экономики. При реализации программы предусматривается, что каждое из этих направлений развития цифровой среды и ключевых институтов цифровой экономики учитывает как поддержку развития уже существующих условий для возникновения прорывных и перспективных сквозных цифровых платформ и технологий, так и создание условий для возникновения новых платформ и технологий. Введено понятие «сквозные» технологии: основными сквозными цифровыми технологиями, которые входят в рамки реализуемой программы, являются:

- 1) новые производственные технологии;
- 2) большие данные;
- 3) нейротехнологии и искусственный интеллект;
- 4) системы распределенного реестра;
- 5) квантовые технологии;
- 6) промышленный Интернет;
- 7) компоненты робототехники и сенсорики;
- 8) технологии беспроводной связи;
- 9) технологии виртуальной и дополненной реальностей.

Таблица 1

**Программа «Цифровая экономика Российской Федерации»:
базовые направления**

№	Направления	Задачи развития направлений
1	Нормативное регулирование	Формирование регуляторной среды, обеспечивающей благоприятный правовой режим для возникновения, развития и использования цифровых технологий
2	Кадры и образование	Подготовка кадров, модернизация системы образования, создание рынка труда для цифровой экономики
3	Формирование исследовательских компетенций и технических заделов	Создание системы поддержки поисковых, прикладных исследований в области цифровой экономики, обеспечивающей технологическую независимость, национальную безопасность и конкурентоспособность на мировых рынках
4	Информационная инфраструктура	Создание отечественных сетей связи, центров обработки данных (ЦОД), обеспечение доступа к создаваемым цифровым данным и внедрение цифровых платформ
5	Информационная безопасность	Обеспечение безопасности данных цифровой экономики, инфраструктуры, институтов, граждан

Эти девять сквозных технологий должны пронизывать все проектные направления, которые в настоящее время активно формируются в рамках программы. В целях управления развитием цифровой экономики определяются цели и задачи в рамках пяти центров компетенций, целью которых является соз-

дание благоприятной экосистемы развития основных сквозных технологий. Формирование исследовательских компетенций и технических заделов закреплено на правительственном уровне (Постановление Правительства России от 28 августа 2017 г. № 1030 «О системе управления реализацией программы „Цифровая экономика Российской Федерации“») за АО «РВК», которое опирается на госкорпорации «Росатом» и «Ростех». Зоны ответственности этих госкорпораций по развитию «сквозных» технологий распределяются примерно поровну. Госкорпорация «Росатом»: новые производственные технологии, большие данные, квантовые технологии, технологии виртуальной и дополненной реальности. Госкорпорация «Ростех»: промышленный Интернет, компоненты робототехники и сенсорики, нейротехнологии и искусственный интеллект, системы распределенного реестра, технологии беспроводной связи.

Государственные корпорации — база для разработки «сквозных» технологий цифровой экономики

Современная стратегия, направленная на модернизацию отечественной экономики, тесно связана с развитием крупных вертикально-интегрированных государственных компаний и корпораций, которые нацелены на решение как коммерческих, так и специальных государственных задач. Заявленные цели инновационной модернизации требуют интеграции широкого спектра современных инструментов развития, в том числе принципов технологических платформ (ТП), ресурсов «институтов развития» (ИР) на федеральном и региональном уровнях, а также механизмов государственно-частного партнерства (ГЧП). Рациональное комбинирование методик и подходов ГЧП и ТП с вовлечением ведущих федеральных и региональных ИР позволяет достигать синергетического эффекта и реального стимулирования процессов инновационной модернизации высокотехнологической промышленности. Атомная энергетика и развитие ядерных технологий в России в настоящее время являются абсолютной монополией государства. Отечественная атомная энергетика вертикально интегрирована и централизована в Государственной корпорации по атомной энергии «Росатом» (Госкорпорации «Росатом»), которая была преобразована из соответствующего министерства, затем федерального агентства и, по сути, сохранила все признаки государственной структуры, поскольку является государственной компанией и занимает монопольное положение на рынке атомной энергетики в стране. Однако по многим параметрам эта ситуация не отвечает требованиям времени и подлежит определенному реформированию. Необходимо проведение модернизации атомного энергетического комплекса как на институциональном, так и на технологическом уровне. Оба вектора модернизации тесно взаимосвязаны друг с другом, что не позволяет осуществлять реализацию одного вектора без другого. Для дальнейшего развития, сохранения и укрепления конкурентоспособности отечественная атомная энергетика нуждается в масштабной инновационной модернизации. Очевидно, отрасль необходимо «открыть» для частных инвестиций как капитального, так и интеллектуального характера. Отрасли необходима реформа, которая повлечет за собой повышение инвестиционной привлекательности в долгосрочной перспективе, а следовательно, инновационности. Это должно быть одной из ключевых целей в стратегии развития атомной энергетики, стратегической установкой Госкорпорации

«Росатом» как ключевого игрока атомной отрасли. Стоит отметить, что качество стратегического управления в отрасли существенно выросло за последние десять лет с момента начала реорганизации управления атомной отрасли в 2007 г. В настоящее время поддерживается постоянный процесс повышения эффективности и непрерывного качественного развития. Однако недостаточное внимание уделяется внедрению в практическое использование современного инструментария стратегического управления, нет последовательности в реализации принятой стратегии. Очевидна высокая вероятность того, что принятая стратегия так и не будет реализована, «останется на бумаге». Необходимо сфокусировать внимание на повышении инновационности и инвестиционной привлекательности отрасли, в поиске путей запуска межотраслевого мультипликативного эффекта для целей стратегического развития и инновационной модернизации высокотехнологичных отраслевых комплексов. Руководством Госкорпорации «Росатом» разработана и утверждена Программа инновационного развития, что подтверждает ориентацию на формирование технологической компании мирового уровня. Однако зафиксированные в этой программе неэнергетические направления развития (ядерная медицина, облучение материалов, системы безопасности и пр.) пока не вносят достойного вклада в экономические результаты деятельности корпорации, сопоставимого с атомной энергетикой. Коммерциализация неэнергетических ядерных технологий (медицинских, материаловедческих и пр.) позволит внести новый импульс в реализацию промышленных инноваций.

Государственные подходы к развитию цифровой экономики

Правительством Российской Федерации разработан и утвержден (или пока еще находится в процессе утверждения) ряд стратегических документов, определяющих общий долгосрочный вектор модернизации и инновационного развития промышленности России. На правительственном уровне утвержден План мероприятий по реализации Стратегии научно-технологического развития Российской Федерации на 2017–2019 гг. (распоряжение Правительства Российской Федерации от 24 июня 2017 г. № 1325-р), в соответствии с которым будет принят целый комплекс важных документов. Достаточно еще раз упомянуть программу «Цифровая экономика Российской Федерации», утвержденную в июле 2017 г. Принимаемые документы должны обозначить вертикаль стратегического планирования в экономике страны. Ключевую роль в единой системе планирования играют государственные корпорации. В принятых документах роль государственных корпораций обозначена формально и нуждается в конкретизации, увязке с верхними уровнями стратегического планирования, отраслевыми стратегиями, региональными и муниципальными стратегиями, а также стратегиями отдельных корпораций (на паритетных началах и принципах консенсуса в случае частных корпораций). Однако пока российские госкорпорации (доля государства, превышающая 50% уставного капитала, характерна более чем для пятидесяти отечественных корпораций) недостаточно эффективны и прозрачны в реализации целенаправленной стратегии инновационной модернизации. Хотя Госкорпорация «Росатом» и демонстрирует существенные успехи в повышении качества корпоративного управления, эффективности стратегического управления и движения по пути инновационной модернизации, существуют при этом и направления для дальнейшего совершенствования в области

стратегического управления отраслью. Главное в этом направлении развития — систематизации регулярного управления за счет повышения инвестиционной привлекательности отечественной атомной энергетики, цифровая трансформация атомной отрасли в целом. Ключевыми инструментами экономического подхода к совершенствованию управления госкорпорацией являются:

- государственно-частное партнерство — выстраивание механизмов кооперации государства с частными предприятиями, интенсификация инновационного развития и модернизации. В развитии российских высокотехнологических отраслевых комплексов применение инструментов и механизмов ГЧП решает не только проблему привлечения дополнительных инвестиций, но и способствует обновлению управленческих ресурсов, внедрению современных моделей стратегического управления отраслевыми комплексами, проведению совместных научно-исследовательских разработок;

- развитие совместных проектов между государством (в том числе в лице госкорпораций) и частными бизнес-структурами. Эти связи пока затруднены из-за отсутствия нормативно-правовой базы ГЧП на федеральном уровне и фрагментарности правового поля в регионах. Существующие нормативные документы либо носят особый региональный характер (например, для Санкт-Петербурга), либо обладают выраженной отраслевой направленностью (закон о концессионных соглашениях, разработанный Минтрансом России). Необходимо скорейшее принятие федерального закона о ГЧП, регулирующего максимально полный спектр взаимоотношений государства и частного бизнеса в реализации общественно значимых проектов в различных отраслях экономики, в том числе в высокотехнологичных разработках и долгосрочных проектах, например в атомной энергетике.

Для перехода к инновационной модернизации экономики крайне важно четко обозначить права и обязанности сторон в рамках ГЧП. Необходимо минимизировать риски вмешательства государства в развитие нормальных рыночных отношений, снижение конкурентоспособности через создание неравных условий для различных участников рынка. Правила участия в общественно значимых государственных проектах с использованием ГЧП должны быть едины, понятны и прозрачны для всех участников рынка. Очень важно обеспечить неизменность государственной политики в рамках реализации проектов ГЧП в долгосрочной перспективе (15–20 лет). Здесь важную роль играет создание эффективных механизмов оценки, одобрения и контроля осуществления соглашений — органов, уполномоченных на выполнение указанных функций на межведомственном уровне. Среди них можно выделить следующие:

- технологические платформы создаются для более эффективного функционирования госкорпораций. В Госкорпорации «Росатом» наиболее известны технологическая платформа реакторов типа ВВЭР и разрабатываемая в настоящее время технологическая платформа замыкания ядерного топливного цикла с реакторами на быстрых нейтронах (проект «ПРОРЫВ»). Концепция технологических платформ не только позволяет обеспечить выбор стратегических научных направлений и всесторонний анализ рыночного потенциала технологий, но и предполагает учет точек зрения всех заинтересованных сторон: государства, промышленности, научного сообщества, контролирующих орга-

нов, пользователей и потребителей. В конечном итоге, использование инструментов и методов технологических платформ ведет к дополнительной мобилизации общественных и частных источников финансирования;

- цифровые платформы являются системой объединения ресурсов различных участников процесса в рамках решения научной, технической или технологической задачи. В частности, цифровые технологические платформы в атомной энергетике должны решить задачу объединения ресурсов коллективов инженеров и ученых в стране и за рубежом в целях разработки новых прорывных технологий энергогенерации и энергосбережения.

Институты развития как драйвер роста цифровой экономики в реализации промышленных инноваций

В России сложилась достаточно обширная система институтов развития в сфере инноваций, обеспечивающая грантовое и заемное финансирование инновационной сферы. К таким институтам можно отнести Российский фонд фундаментальных исследований (РФФИ, 1992 г.), Фонд содействия инновациям (1994), Российскую венчурную компанию (2006), ОАО «Росинфокоминвест» (2006), РОСНАНО (2007) и ФИОП РОСНАНО (2010), Фонд «Сколково» (2010), Фонд «ВЭБ Инновации» (2011), Фонд развития интернет-инициатив (ФРИИ, 2013 г.), Российский научный фонд (РНФ, 2013 г.), Фонд развития промышленности (с 2014 г., ранее — Российский фонд технологического развития), а также Российский экспортный центр (2015) и ГК Внешэкономбанк (2007). Кроме того, выделяется ряд особых институтов, к числу которых можно, в частности, отнести АНО «Агентство стратегических инициатив по продвижению новых проектов (АСИ)». К нефинансовым инновационным институтам развития также должны быть отнесены: ОАО «Особые экономические зоны», НКО «Фонд развития моногородов», инновационная инфраструктура (индустриальные парки, технопарки, инновационные кластеры и др.), Корпорация развития Дальнего Востока, Фонд развития Дальнего Востока, ОАО «Корпорация развития Северного Кавказа», региональные корпорации развития.

Важным звеном в стратегии цифровой трансформации атомной отрасли и повышения инвестиционной привлекательности этой отрасли, в конечном счете — интенсификации инновационной модернизации атомной энергетики, должны быть институты развития — федеральные и региональные. Необходимо вовлекать в развитие атомной энергетики федеральные и региональные институты развития — Российскую венчурную корпорацию (РВК), «Роснано», Инновационный центр «Сколково», Инвестиционный фонд РФ, Внешэкономбанк (и его дочерние компании — Российский банк развития и Экспортное страховое агентство России), Евразийский банк развития и пр.

Стратегические задачи развития, технологические и научные ресурсы Госкорпорации «Росатом» позволяют ей играть ведущую роль в развитии страны — принять активное участие в формировании системы территориальных научных и производственных центров (инновационных кластеров и индустриальных парков), территорий опережающего развития (ТОР). В этой связи требуется решение следующих задач: законодательные ограничения на привлечение иностранных инвестиций и организацию предприятий на территориях для отраслевых кластеров (большая часть предприятий расположена на закрытых

территориях — ЗАТО), низкое качество транспортной, энергетической и социальной инфраструктуры, недостаточная квалификация специалистов региональных и муниципальных органов власти в вопросах проектного управления и инновационного развития и пр.

Развитие атомной энергетики в России всегда осуществлялось и финансировалось государством как по причине стратегической важности отрасли (она появилась «в дополнение» к военному ядерному комплексу), так и в связи с крайне высокой капиталоемкостью и длительным сроком возврата инвестиций. Тем не менее принятая стратегия развития атомной отрасли (в том числе экономическая экспансия за рубеж) и потребность в скорейшей модернизации существующих генерирующих мощностей ставят перед Госкорпорацией «Росатом» задачу создания условий для привлечения российского и иностранного частного капитала для развития атомной энергетики страны. Сегодня существует ряд ограничений для приватизации «гражданской» части отечественного атомного комплекса. Основное препятствие — интегрированность научно-технической базы отрасли с оборонным ядерным комплексом, международные требования МАГАТЭ и пр. В ведущих ядерных странах мира подобное сращивание гражданской и военной частей атомного комплекса не является типичным, что, соответственно, наносит определенный ущерб глобальной конкурентоспособности российской атомной энергетики и энергомашиностроения. Необходимо проводить целенаправленные управленческие и инженерно-технические действия в целях разделения атомного и оборонного направлений Госкорпорации «Росатом» с целью частичной приватизации атомного энергетического комплекса России в перспективе.

В ближайшем будущем даже частичная приватизация Госкорпорации «Росатом» не может быть эффективно реализована как в силу технико-экономических причин (внутренних — неготовность компании к приватизации из-за значительной интеграции ресурсной базы и системы НИОКР, и внешних — негативный макроэкономический фон в мире), так и в силу специфики функционирования данной отрасли в России. Даже с учетом негативных качеств, присущих организационно-правовой форме государственной корпорации (также во многом и публично-правовой компании), экономический анализ показал, что именно эта форма (при организации должного надзора) является наиболее адекватной профилю и стратегии компании, ее отраслевой специфике.

Стратегические направления развития цифровой экономики и государственно-частное партнерство

Стратегическое управление отраслью атомной энергетики реализуется на сравнительно высоком и качественном уровне. Это подтверждается высокими результатами, достигнутыми российской атомной энергетикой в последние годы, — принят ряд стратегических документов, приняты цели по научным исследованиям, сделана фокусировка на ключевых (прорывных) проектах по НИОКР. Также отрасль проводит активную модернизацию существующих мощностей, строительство новых объектов, реализуется масштабная экспансия на мировом рынке строительства и эксплуатации АЭС, совершен ряд сделок по приобретению активов за рубежом, в первую очередь в целях дальнейшего развития ресурсной базы и минимизации рисков топливного дефицита на среднесрочную и долгосрочную перспективу. Тем не менее важно обратить внимание

на то, что функционал стратегического управления в Госкорпорации «Росатом» (и соответственно в отрасли в целом) серьезно размыт, отсутствует необходимая для эффективной деятельности централизация ответственности и соответствующих полномочий, что могло бы позволить реализовывать политику перманентного повышения инвестиционной привлекательности отрасли. Следует продолжать работу в направлении дальнейшего совершенствования стратегического управления в российской атомной отрасли для формирования новых цифровых платформ, что неизбежно потребует развития ГЧП.

Привлечение инвестиционных ресурсов в развитие атомной энергетики как высокотехнологичной стратегической отрасли зависит от комплекса факторов. Прямолинейное увязывание доступности инвестиционных ресурсов в экономике с инвестиционной привлекательностью и соответственно наличием или отсутствием инвестиций в отрасли оказывается слишком упрощенным, поверхностным и не позволяет эффективно управлять инвестиционным потенциалом в долгосрочной перспективе. В ряде исследований показано, что последовательное стратегическое повышение инвестиционной привлекательности отраслевого комплекса возможно только при создании и управлении системой, включающей управление следующими инфраструктурными блоками:

- нормативно-правовой блок;
- ресурсный блок;
- технологический блок, включая цифровые технологии.

Причем эффективность данной работы по повышению инвестиционной привлекательности непосредственно зависит не только от координации этапов работ по каждому из направлений, но и от конструктивного сотрудничества между этими блоками и отдельными проектными группами, специалистами и экспертами на основе матричного принципа организации процесса. В управлении атомной отраслью стратегически обоснованным будет сосредоточение всей полноты ответственности и полномочий по разработке инвестиционной политики и осуществления деятельности по повышению инвестиционной привлекательности в рамках одного структурного подразделения или блока.

Инвестиции совершенно необходимы в атомной отрасли для реализации перспективных цифровых технологий. В частности, при помощи когнитивных технологий машины поиска информации в Интернете перестанут выдавать абсурдные миллионы ссылок. Они станут сами обрабатывать собранные ссылки, соревнуясь в полноте, достоверности и доступности для восприятия человеком создаваемых ими рефератов. Потребитель самостоятельно найдет производителя, а, учитывая возможность автоматического документооборота, тот сможет напрямую взаимодействовать со всеми своими контрагентами. Так появятся бизнес-модели M2C (manufacturer to customer, производитель — потребителю) и обратный C2M, при котором возможна реализация персонализированного производства, предполагающего производство товара, обладающего необходимыми (или желательными) для данного потребителя оригинальными свойствами. Облачные технологии или облачные вычисления (Cloud Computing) — информационно-технологическая концепция, подразумевающая обеспечение повсеместного и удобного сетевого доступа по требованию к общему объему конфигурируемых вычислительных ресурсов, которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами

или обращениями к провайдеру. Примерами ресурсов могут являться сети передачи данных, серверы, устройства хранения данных, приложения и сервисы — как вместе, так и по отдельности. Иначе говоря, облачные технологии — это технологии обработки данных, в которых компьютерные ресурсы предоставляются интернет-пользователю по запросу (on demand) как онлайн-сервис. Необходимо подчеркнуть, что облачные технологии внесли колоссальный вклад в фундамент зарождающейся цифровой экономики. Этот вклад не ограничивается лишь технологической составляющей, но включает еще экономическую и идеологическую компоненты. Развитие облачных технологий, например, привело к появлению таких понятий, как производство по требованию (production on demand), программное обеспечение как услуга (software as a service) и многих других, которые станут лейтмотивом большинства бизнес-моделей будущего и принципом большинства экономических взаимодействий.

Перспективные интернет-технологии и информационная среда цифровой экономики

Промышленный Интернет или Интернет вещей — это концепция, объединяющая множество технологий, подразумевающая оснащенность датчиками и подключение к Интернету всех приборов (и вообще вещей), что позволяет реализовать удаленный мониторинг, контроль и управление процессами в реальном времени (в том числе в автоматическом режиме). Сегодня сформированы два крупных направления: Интернет вещей (IoT — Internet of Things) и промышленный Интернет вещей (IIoT — Industrial Internet of Things). Инструментально данные технологии очень похожи, ключевая разница в предназначении: если основная задача Интернета вещей — это сбор всевозможных данных (которые будут приоритетно использоваться для построения моделей и прогнозов), то предназначение промышленного Интернета вещей состоит в автоматизации производства (за счет удаленного управления ресурсами и мощностями по показаниям датчиков).

Технологии цифровой трансформации экономики прежде всего подразумевают массовое использование интернет-технологий. Использование интернет-систем в этом случае должно базироваться на стандартах, которые подразумевают пятый (следующий за действующим сейчас четвертым уровнем) уровень стандартизации или стандарты 5G. Целевые характеристики 5G предусматривают резкий рост производительности и скорости реализации интернет-протокола:

- скорость передачи данных должна обеспечить рост производительности интернет-связи в 10–100 раз в расчете на абонента — до 10 Гбит/с (DL) и до 5 Гбит/с (UL);
- потребляемый трафик абонентов должен обеспечить рост в 1000 раз — до 500 Гб на пользователя в месяц;
- предусматривается увеличение количества подключаемых абонентских устройств в интернет-соте в 10–100 раз (до 300 000 на узел). Рост M2M-устройств с 50 до 500 млрд;
- для промышленного интернет-протокола потребуется увеличение в 10 раз времени автономной работы абонентских устройств с небольшим энергопотреблением, таких как сенсоры M2M;

- потребуется сокращение времени задержки в цепочке E2E с 5 до 1 мс и менее;
- новые технологии предусматривают снижение стоимости эксплуатации и энергопотребления сетей 5G до 10% от текущего потребления сетей 4G.

По своим производственным возможностям технология 5G должна составить к 2025 г. альтернативу наземным сетям цифрового телевидения DVB-T. Все эти изменения и многие другие технологические нововведения дадут возможность в массовом масштабе создавать цифровые платформы.

Платформа цифровой экономики — это цифровая среда (программно-аппаратный комплекс) с набором функций и сервисов, обеспечивающая потребности потребителей и производителей, а также реализующая возможности прямого взаимодействия между ними. Ценность цифровой платформы — в предоставлении самой возможности прямой коммуникации и облегчении процедуры взаимодействия между участниками бизнес-процессов. Платформы снижают издержки и предоставляют дополнительный функционал как для поставщиков, так и для потребителей. Также они предполагают обмен информацией между действующими лицами, что должно существенно улучшать сотрудничество и способствовать созданию инновационных продуктов и решений. Цифровая платформа как бизнес-модель существует давно. Простым примером может служить классический рынок, на котором продавцы и покупатели (производители и потребители) находят друг друга. В современном мире можно привести много активно растущих компаний, в основе которых функционируют принципы «платформенной бизнес-модели», и самые яркие — это Uber и Airbnb. Для сферы образования интернет-технологии (рис. 1) дают возможность не только дистанционного обучения, но и создания новых образовательных платформ.



Рис. 1

Основные направления использования интернет-технологий в образовании

Организация педагогических сообществ в интернет-среде может позволить быстро и качественно формировать новое содержание образовательных программ, которые пока сильно отстают от потребностей реальной жизни. А ведение международной учебно-проектной деятельности станет новым стратегическим направлением, позволяющим объединить консалтинг, обучение, повышение квалификации и формирование «интеркультурных» творческих коллективов для решения крупных международных проблем.

Указ Президента России от 09.05.2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» декларирует, что развитие цифровой экономики является стратегически важным вопросом для России в целом, определяющим ее конкурентоспособность на мировой арене. Необходимо признать, что в России сегодня нет условий для стихийного формирования зрелой цифровой экономики за приемлемый период времени — в первую очередь из-за технологического отставания и отсутствия критической массы экономических субъектов. Это значит, что государству необходимо стимулировать и направлять развитие цифровой экономики. Важной отличительной особенностью российской экономики является тот факт, что львиная доля ВВП создается государственными корпорациями (или компаниями со значительной долей государственного участия). Во многих отраслях производства игроки с государственным участием могут составлять до 80% рынка. В таких условиях наиболее рациональным шагом представляется создание ряда индустриальных цифровых платформ под руководством профильных министерств или госкорпораций. Такие платформы создадут необходимый инфраструктурный базис для максимально быстрого развития цифровой экономики и распространения сопутствующих технологий. При построении платформ цифровой экономики необходимо фокусировать усилия на ключевых направлениях: транспорт, телекоммуникации, энергетика, обработка данных. Развитие именно этих областей позволит создать инфраструктурный и технологический базис, тиражируя который на другие области Россия сможет максимально быстро развить зрелую цифровую экономику.

Теоретические подходы к обоснованию методов развития цифровой экономики

Фундаментальная экономическая теория пока отстает от практики. На сегодняшний день не существует теоретической базы не только для будущей цифровой экономики, но даже для современной экономики услуг, основанной во многом на цифровых платформах (системы такси, интернет-торговли и пр.). После первой четверти XX в. базовая экономическая теория практически не имела развития — некоторые успехи были достигнуты лишь в рассмотрении отдельных вопросов. Начиная со второй половины XX в. официальные социально-экономические науки беззащитно обслуживали интересы финансово-олигархических «элит». Важнейшие вопросы системных свойств капитализма, на которые обращали внимание классики, старательно игнорировались. Все основные экономические законы и метрики (в том числе ВВП) были введены и сформулированы еще в XIX — первой половине XX в. и хорошо описывают реальный сектор (производящую промышленные товары экономику). Со второй половины XX в. сектор услуг и нематериального производства по-

лучил значительное развитие и со временем стал основным сектором экономики в наступившем XXI в. Свойства производства и потребления в нематериальной сфере значительно отличаются, но человечество пока не создало соответствующей теоретической базы для корректного описания новой экономики. Вместо этого создавались и постоянно пересматривались методики, позволявшие «привести» нематериальную сферу к уже имевшимся метрикам и показателям, чтобы иметь возможность включить ее в уже привычные формы описания экономики. До определенного момента эти попытки давали приемлемые результаты, но только до тех пор, пока нематериальный сектор производства услуг не стал превышать реальный сектор экономики по производству товаров. Еще одним отягчающим обстоятельством является политизированность и предвзятость современной экономической науки, что приводит к спекулятивным и преднамеренным искажениям общей картины (например, из-за практики постоянного пересмотра методик подсчета ВВП). Одним из ключевых вопросов при формировании новой теории является выбор адекватных интегральных параметров и формирование новых метрик. В нашем мире уже есть несколько подходящих устойчивых тенденций, учет которых может помочь в формировании необходимого базиса обновленной экономической теории:

- информация постепенно становится товаром и происходит конвергенция и взаимообогащение программной продукции и аппаратных средств;
- благосостояние общества коррелирует с удельным потреблением энергии;
- социальный статус постепенно вытесняется социальным авторитетом.

Возможно, валютой будущего может стать субстанция, которую можно охарактеризовать как «количество произведенной вами полезной информации на количество потребленной энергии», где «полезность» измеряется в сетевых «лайках». Сегодня такое предсказание кажется чрезмерно футуристическим, но прогресс экономического развития постоянно ускоряется и подобная перспектива может ожидать нас совсем не за горами, а уже через 15–20 лет. Достаточно вспомнить, что такому важному технологическому изобретению для человечества, как ткацкий станок, понадобилось 120 лет, чтобы покинуть пределы Европы, а Интернету понадобилось всего 10 лет, чтобы охватить всю планету.

Цифровая экономика предлагает широкие возможности для развития системы государственного управления. Современные технологии позволяют в ближайшем будущем создать среду высокотехнологичной цифровой платформы государственного управления, которая обеспечит минимизацию человеческого фактора, сопутствующей ему коррупции и ошибок, автоматизирует сбор статистической, налоговой и иной отчетности, обеспечит принятие решений на основе анализа реальной ситуации. Оказание государственных услуг будет строиться на базе единой цифровой облачной платформы, имеющей открытые интерфейсы межмашинного взаимодействия и позволяющей расширять возможности взаимодействия граждан с государством путем создания ими собственных приложений, работающих на базе этой платформы (с обязательной сертификацией по безопасности и соблюдению законодательных норм). В атомной отрасли информационные системы и комплексы должны учитывать все описанные выше ас-

пекты, давая возможность на производственных предприятиях получать экономические выгоды от создания и использования цифровых платформ.

Цифровая трансформация традиционных общественных отношений

Современный период развития реального сектора экономики зачастую отождествляют с технологической революцией, оказывающей все большее влияние на общество в целом. Следствием широкого распространения цифровых решений и прорывных технологий являются изменения не только в производстве, но и на рынках труда. Во всех сферах задаются жесткие требования к скорости и качеству информации, являющейся основой государственных и бизнес-решений. По многим традиционно «неторгуемым» (т. е. непереносимым в географическом пространстве) товарам и услугам кратно снижаются издержки перемещения. Новые форматы товаров и услуг из неторгуемых становятся торгуемыми. В первую очередь это касается сфер образования, здравоохранения, безопасности, а также других услуг, традиционно предоставляемых государством:

- практически весь спектр государственных услуг для граждан становится электронным;
- контроль юридически значимых действий и финансовых операций с помощью интеллектуальных роботизированных систем начинает осуществляться по цифровому следу;
- беспрепятственный доступ в Интернет увеличивает число безналичных финансовых расчетов;
- получают широкое распространение цифровые сервисы в социальной сфере, связанные прежде всего с контролем в режиме реального времени самых разнообразных объектов (от использования ресурсов в «умном» доме до мониторинга состояния здоровья с помощью подключенных цифровых устройств);
- развиваются компьютерные форматы образования, в том числе дистанционного.

Количественные характеристики новых образовательных форматов можно проиллюстрировать тем, что в конце 2016 г. цифровая образовательная платформа Coursera объявила, что число ее студентов достигло 18 млн человек, число пройденных курсов превысило 3,6 млн, а совокупный трафик видеопросмотров курсов составил 17 тыс. лет. Количество учащихся, зарегистрированных на цифровой образовательной платформе edX, достигло 7 млн человек. В то же время испанская МООС-платформа Miriada X достигла показателя в 2 млн студентов. Всеиндийский совет по техническому образованию (АICTE) принял решение о том, что 10% учебного плана в 10,8 тыс. технических институтов страны должно базироваться на платформе МООС. Под давлением технологической революции быстро перестраивается качество услуг в социальной сфере в целом:

- профессиональное образование во всем мире сталкивается с растущей конкуренцией со стороны неакадемической сферы. Квалификации выпускников устаревают быстрее, чем успевает среагировать традиционная система образования. Происходит переключение с проектно-ориентированного образования на экспериментально-ориентированное, а передача практических навыков обеспечивается не только за счет стажировок, но и за счет распространения специально созданных учебных заводов и учебных лабораторий. Использование в образовании цифровых технологий, в том числе больших данных, искусственного

интеллекта, разного рода нейротехнологий, претендует на изменение природы познания и предоставление образовательных услуг в целом;

- на базе цифровых и биологических технологий продолжается развитие медицины «4П»: превентивной/предупредительной, прогностической, пациент-ориентированной, персонифицированной. В последние годы закладываются основы «5П-медицины», которая дополняется решениями на базе цифровых платформ, основанных на использовании математических моделей состояния здоровья или связанных с данными моделями комплексов методов его контроля и лечения;

- в науке и научной деятельности происходят преобразования, связанные прежде всего с использованием больших данных, искусственного интеллекта и цифровизации исследований: аналоговые системы фиксации экспериментальных данных вытесняются цифровыми, происходит замена натуральных экспериментов цифровым моделированием, растет использование цифровых средств сбора, обработки и хранения информации.

Цифровая экономика и цифровая трансформация производственных комплексов вызовут значительные изменения в ближайшее время на рынке труда, особенно в традиционных отраслях и профессиях:

- изменение структуры рынка труда в сторону роботизации производств будет иметь последствия для занятости в сегменте рабочих профессий: по прогнозам, к 2035 г. в развитых странах роботы заместят работу, выполняемую людьми, в 25–30% видах профессиональной деятельности;

- изменение структуры занятости в сторону использования искусственного интеллекта будет иметь последствия для «белых воротничков» — менеджеров, аналитиков и пр., а в некоторых случаях и для высших управленческих кадров. Ожидается, что к 2025 г. до 30% корпоративных аудиторских проверок будет осуществляться с использованием технологий искусственного интеллекта. Кроме того, полная роботизация возможна в отдельных сегментах банковской деятельности, юридических услугах, бухгалтерском учете, сложной аналитике;

- повсеместный отказ от систем пожизненного найма и быстрая смена квалификационных требований к работникам приводят к изменению ими своего отношения к профессиональной карьере и выбору занятий: работники все больше должны стремиться сами создавать себе работу и заботиться о ее рентабельности;

- одновременное формирование на рынке труда огромной потребности в новых занятиях и профессиях, связанных с использованием передовых производственных технологий, интеллектуализацией, роботизацией производства и т. п., формирует новые требования к персоналу.

Новые требования к работникам будут постоянно изменяться, трансформируясь совместно с изменениями цифровых платформ в различных сферах. При этом движущей силой этих изменений будут новые подходы к формированию потоков в цифровой экономике: информационных, энергетических, материальных, финансовых® и пр. Изменения внешней среды будут воздействовать на структуру производственных комплексов, а внутри этих комплексов новые потоковые конфигурации последовательно определяют новые необходимые компетенции работающих в каждой конкретной сфере (рис. 2).

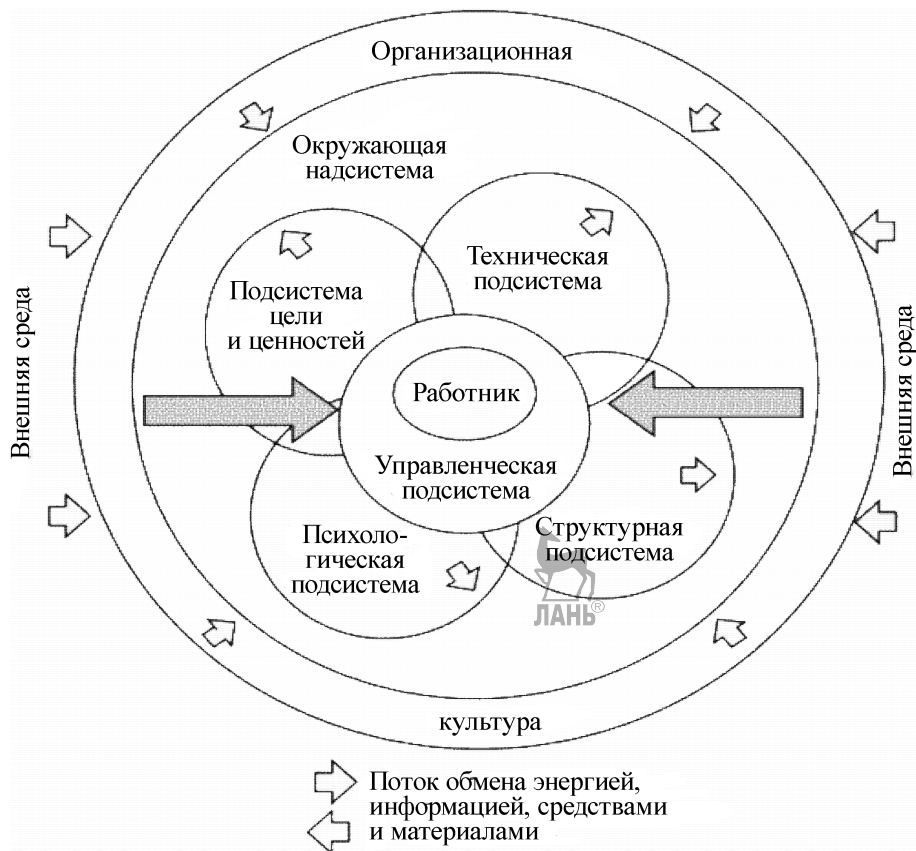


Рис. 2

Подсистемы производственных комплексов, в центре развития которых формируются новые требования к персоналу

Анализ и сопоставление зарубежного опыта цифровой трансформации

В результате проходящей в настоящее время «цифровой революции» государственная политика во всем мире претерпела в последнее время резкие изменения. Сформировать ответ на вызов сохранения конкурентоспособности и достижения высоких темпов производительности в настоящее время призвана проводимая государствами научно-технологическая и инновационная политика. Ее цель — стимулировать разработку и внедрение передовых технологий, которые отличаются высокой производительностью и могут обеспечить наибольший вклад в технологический и экономический рост. Индустриально развитые страны (США, Германия, Великобритания, Япония, Китай, Южная Корея и др.) приняли решение о развертывании новой технологической революции в виде государственной политики. Кроме того, эти государства хотят сосредоточить у себя ключевые универсальные (цифровые) платформы, агрегирующие так называемые стратегические данные и алгоритмы их обработки. Страны-лидеры уже сегодня реализуют целый пакет больших государственных программ в сфере передовых технологий в промышленности и непромышленных секторах экономики, рассчитанных на запуск новой технологической револю-

ции и радикальное укрепление конкурентных позиций на глобальных рынках. Так, в Германии в 2012 г. была инициирована промышленная стратегия «Индустрия 4.0» (Industrie 4.0) как один из десяти «проектов будущего» в рамках «Плана действий по реализации обновленной федеральной Стратегии в области высоких технологий». В США приняты «Стратегия инновационного развития», «Национальный стратегический план развития передовых промышленных технологий США», а также реализуется ряд профильных межведомственных инициатив, таких как «Инициатива генома материалов», национальные инициативы в сфере робототехники и т. п. Великобритания реализует собственный план развития передовых производств, а также программу развития «Восемь великих технологий». В 2013 г. Франция запустила программу «Новая промышленная Франция», в рамках которой реализуются проекты по 10 перспективным технологическим направлениям развития индустрий и технологий будущего. В Японии запущен уже 5-й пятилетний план развития науки, технологий и инноваций (2016–2020). Китай с 2015 г. реализует программы «Сделано в Китае — 2025» и «Интернет+». Кроме того, в июле 2017 г. в КНР был утвержден «Национальный план стимулирования технологических разработок в сфере искусственного интеллекта».

Практически все развитые страны с конца нулевых годов нового века, помимо традиционной активизации промышленной и технологической политики, увеличивают инвестиции в научные исследования — источник прорывных технологий. Особенностью данного инвестиционного цикла является то, что при сокращении государственных бюджетов на НИОКР растут частные инвестиции в исследования и разработки. Это связано с тем, что критическим условием успешности в разворачивающейся технологической гонке является ранний доступ к прорывным технологиям на том этапе, когда они еще не доведены до прототипа, а являются научной гипотезой и ранней технологической идеей. Это заставляет всех, кто включился в данную гонку, переоценить риски, связанные с созданием продукта, переписать ключевые составляющие формулы времени продвижения на рынке (time-to-market). Инвестиции в науку и исследования, увеличение пространства для экспериментов, различные способы ускорения исследований и доведения научных гипотез до полезных продуктов на рынке (R&D fast track programs), коллаборативные механизмы объединения усилий и разделения ответственности (исследовательские и технологические консорциумы) — это широко применяемые в настоящее время инструменты хеджирования рисков.

Учет российских реалий при вступлении в эпоху цифрового бизнеса

Анализируя тематику разного рода конференций и публикаций о трендах развития информационных технологий (ИТ) в 2017 г., нетрудно заметить нарастающую волну направления «цифрового предприятия» (Digital Enterprise). Особенно четко это проявилось в попытке преобразования проекта DOCFLOW с двадцатилетней историей из выставки-конференции по проблемам развития ИТ в мероприятие DigEn, посвященное Digital Enterprise. Данная конференция помимо воли ее организаторов наглядно отразила специфику текущего момента в плане трансформации бизнеса от его «доцифровой» стадии развития к «цифровой». По сути мероприятие состояло из двух мало связанных между собой

частей, в первой из которых шел сугубо теоретический разговор о «цифровом будущем», а во второй — традиционная демонстрация ИТ-решений для нашего «сегодня». Тем не менее, хотя эпоха «цифровой экономики» еще не наступила, уже понятно, что она не за горами и к ней нужно готовиться. И на DigEn были не только представлены «розовые» перспективы цифрового будущего, но и показаны имеющиеся трудности на пути к нему. Сегодня есть общее понимание того, что речь идет о трансформации не просто ИТ-рынка, но всей мировой экономики под влиянием новых возможностей ИТ. Разумеется, степень такой цифровой трансформации зависит от уровня экономического развития и специфики разных стран, а также от вертикальных отраслей. Однако технологии — условие необходимое, но недостаточное.³⁰ Для реализации Digital Enterprise нужны новые модели организации бизнеса, позволяющие создавать такие предприятия, вся деятельность которых будет базироваться на обработке и анализе данных. И необходим также третий обязательный компонент — стратегия развития, подразумевающая возможность не только постоянного оперативного изменения предприятия в меняющихся условиях окружающего мира, но и заблаговременную подготовку к будущим изменениям на основе точных прогнозов, обеспечиваемых как раз новыми технологиями.

Конечно, цифровая трансформация компаний, производств, отраслей — очень непростое дело: существует классический пример компании «Кодак», которая, став одним из инициаторов цифровой революции в фотографии, как раз в результате начатого преобразования отрасли потеряла свое лидерство. Разумеется, темпы цифровой трансформации различаются в отраслевом разрезе. Лидерами являются ритейл, банковский сектор, телекоммуникационная отрасль, т. е. те, кто в наибольшей степени работает в массовом потребительском секторе, кому нужно быстро реагировать на изменения в спросе и на внедрение новейших технологий. В физическом производстве темпы будут медленнее, и там объектом анализа станет улучшение внутренних производственных процессов.

Одним из активных пропагандистов идей «цифрового предприятия» в России выступает Сбербанк, который не только расширяет использование ИТ в своей внутренней деятельности, но все чаще выступает и как ИТ-провайдер. Например, руководство компании «Сбербанк-Технологии» (дочерняя структура Сбербанка) совсем недавно предложило такое определение цифрового предприятия: «Это организация, способная быстро адаптироваться к быстро меняющимся окружающим условиям, в том числе на основе проактивного прогноза развития ситуации в будущем. Жизнь — это движение, таков её лозунг». Одна из главных характеристик современного мира — высокая скорость его изменения, которая, по мнению экспертов, характеризуется несколькими трендами:

- активы (любые — финансовые, материальные, интеллектуальные) в цифровую эпоху становится легче получать;
- связность общества и его технологического обеспечения повышается: «люди и вещи» постоянно на связи;
- возникают новые производственные системы: появляются возможности использования умных материалов, аддитивных технологий и пр.;
- сложившиеся жесткие организационные иерархии постепенно сменяются гибкими сетями;

-
- массовые решения и услуги (в здравоохранении, ЖКХ и пр.) становятся персонализированными;
 - относительно несовершенные человеческие отношения постепенно заменяются машинными алгоритмами, которые все больше управляют бизнесом и обществом.

В цифровом бизнесе исчезает былое деление специалистов по роду деятельности: «информационщики» должны быть глубоко погружены в деловые бизнес-процессы, а управленцы — отлично знать возможности ИТ и тренды их развития. Например, все ведущие менеджеры Сбербанка не только читают книги по бизнес-аналитике, большим данным и нейронным сетям, но и сдают зачеты по этим дисциплинам. Хотя, конечно, организационная трансформация требует и иных базовых технологий. В частности, Сбербанк сейчас ведет работы по созданию новой технологической ИТ-платформы, которая использует механизмы in-memory, ориентируется на применение аппаратных средств стандартной low-end-архитектуры и идей Open Source. На одну важную проблему следует также обратить внимание: трудность трансформации заключается также во внешних условиях для бизнеса, а именно в нормативно-законодательной базе, которая по-прежнему нацелена на поддержку старой жесткой иерархической системы управления. Существуют примеры трудностей работы внешних поставщиков информационных решений с тем же Сбербанком, который, как государственная компания, должен работать по 223-ФЗ, фактически не позволяющему вносить какие-то изменения в исходное техзадание. А современные командные и итеративные технологии создания ИТ-решений вступают с этими обстоятельствами в противоречие: при этом нормативном подходе невозможно применять Agile-методы. Именно поэтому Сбербанк развивает внутреннее ИТ-направление (в том числе в виде выделенных «дочек»), поскольку оно позволяет взаимодействовать с командами разработчиков в рамках гибких моделей управления.

Следует признать, что сейчас тема «цифрового предприятия» находится пока в стадии Нуре («много шума, маркетинга, рекламы») и задачей сегодняшних дискуссий во многом является отделение зерна от плевел, обсуждение реальной содержательной сущности каждого тренда и перспектив его развития. Но уже сейчас ясно, что речь идет не о мыльном пузыре, а о реальных изменениях в экономике, связанных с ИТ: можно сравнить сегодняшний этап зрелости данной темы с разговорами об облаках и облачных сервисах в 2009–2010 гг. А это означает, что цифровое предприятие, основанное на информационных технологиях, должно отвечать модели рисков, причем весьма значительных. Нужно быть готовым к тому, что придется провести десять пилотных проектов, в то время как реального результата достигнет лишь пара из ваших начинаний. Главное достижение и одновременно главная проблема современных ИТ-систем — это их сложность. Дальнейшее повышение их производительности и адаптивности даже при неизменном составе решаемых бизнес-задач невозможно без радикального пересмотра бизнес-архитектуры и организационного обеспечения. Это можно сделать на базе концепции иерархической эмерджентной стратификации систем при организации разработки и внедрения ИТ, реализуя отдельные рабочие места в виде своеобразных «инфороботов», которые смогут на основе имеющихся данных выполнять часть процессов в автономном режиме. Многие руководители

ИТ-компаний сегодня постулируют: если Сбербанк и другие традиционные компании становятся ИТ-поставщиками, то и собственно ИТ-бизнес должен заниматься другими видами деятельности, используя свои преимущества, в том числе в плане умения эффективно управлять данными. ИТ-специалисты должны погружаться в бизнес-схемы не только на уровне отношений внутри отдельного предприятия, но и на уровне самостоятельного бизнеса. Собственно, известный пример компании Uber как раз об этом и говорит.

Тренды цифровой трансформации, связи с общественностью и рекламный рынок

Рассматривая проблему перехода к цифровой экономике, невозможно не замечать ту сторону больших данных, которые относятся к СМИ, связям с общественностью и просто рекламному бизнесу. Международные компании в своих информационных кампаниях все чаще пользуются цифровыми носителями. Например, недавний отчет Digital PR and Communications Report, авторитетной британской ассоциации консультантов по связям с общественностью PRCA, отмечает, что средний процент от маркетингового бюджета компаний, потраченного на цифровые социальные медиа, к осени 2016 г. составил 25% по сравнению с 16% в предыдущем. При этом участники исследования ожидают роста digital-затрат в ближайшие год-полтора до 62%. Соответственно, чтобы не просто удержаться на плаву, а достичь преимущества, компаниям имеет смысл уже сейчас активнее интегрировать новые технологические инструменты коммуникации в продвижение своих продуктов и услуг.

Структура групп, на которые планируются воздействия при формировании связей с общественностью (рис. 3), определяет и выбор цифровых технологий, которые могут быть использованы в перспективном развитии этого направления деятельности компаний. Следует рассмотреть, какие ключевые тренды в перспективной области digital-продвижения сейчас наиболее актуальны.



Рис. 3

Структура групп, на которые планируются воздействия при формировании связей с общественностью в любой компании

Тренд № 1: рост использования Интернета вещей (IoT)

Об Интернете вещей говорят уже давно и много, но по-прежнему далеко не все понимают, как правильно использовать его возможности в коммуникациях с клиентами. Сейчас в России IoT применяют в основном в промышленном секторе, на производствах, а также в инфраструктурных объектах в рамках концепции «Умный город» (камеры видеонаблюдения, датчики энергопотребления, расхода тепла и т. д.). В то же время на Западе большую популярность приобрело применение интеллектуальных устройств на персональных продуктах, прежде всего потребительской электроники. Особенно актуально это для сферы розничной торговли, где уже предлагаются соответствующие решения. К примеру, компания Intel предлагает платформу Intel Retail Sensor, которая использует специальную аналитическую систему на базе RFID-меток (радиочастотная идентификация), размещенных на продаваемых в магазине товарах, в комплексе с камерами видеонаблюдения и интерактивными экранами Digital Signage (видеореклама на мониторах). Система отслеживает действия покупателя, определяет его социально-демографический портрет и на основе полученных данных делает персональное предложение. Например, покупатель берет рубашку в руки, а на экране рядом приятный женский голос предлагает приобрести к ней галстук со скидкой. Частично эта технология Intel уже применяются в магазинах Levi Strauss & Co., Brooks Brothers и Nordstrom. Из отечественных ритейлеров в начале 2017 г. сеть Media Markt внедрила в своих магазинах электронные ценники, которые в режиме реального времени отображают цены и информацию по ассортименту. Они позволяют не только вывести на экран всю необходимую информацию о продукте, но и использовать устройства как медианосители. В скором времени с помощью NFC-технологий и анализа покупательского поведения клиентам будут предлагаться индивидуальные условия покупки. Это позволит привлечь дополнительное внимание покупателей и, как следствие, увеличить продажи.

Тренд № 2: использование виртуальной реальности (VR) и дополненной реальности (AR)

Технологии виртуальной и дополненной реальности стремительно из области развлечений переходят в средство коммуникации с клиентами для продвижения продуктов компаний. Доступность контента, очков и смартфонов делает VR-технологии удобным каналом коммуникаций с рядовыми потребителями. Первые шаги к массовому вовлечению пользователей с помощью VR-технологий уже сделал Facebook, анонсировав в 2017 г. бета-версию продукта Spaces. Пользователи сервиса будут подключаться через свои аккаунты в Facebook и погружаться в виртуальное пространство, где общение осуществляется с помощью голоса и движения тела. Там можно общаться, рисовать, смотреть сферические ролики, совершать видеозвонки через Facebook Messenger и создавать VR-селфи. В таком пространстве можно будет предлагать пользователям как виртуальные, так и физические товары, рекламировать услуги, создавать сообщества по интересам, обсуждать важные темы. Что касается традиционного маркетинга, там уже сейчас набирает популярность демонстрация клиентам продуктов и услуг с помощью VR- и AR-технологий, из актуальных направлений — VR-туры по недвижимости. Отечественные девелоперы элитной недвижимости предлагает своим потенциальным клиентам с помощью технологий виртуальной реальности (специально разработанной программы, уста-

навливаемой в виде приложения на смартфон, и VR-очков), не вставая с дивана, познакомится с жилым комплексом, изучит планировку квартир, интерьер, испытать эмоции, сравнимые с реальным присутствием на объекте. Необычное применение AR-технологий можно наблюдать на примере «умной фермы» от компании DeLaval. Здесь клиенту, чтобы увидеть работу молочной фермы, не обязательно ехать в коровник, достаточно воспользоваться специальным голографическим столом и очками, которые позволят в деталях рассмотреть, например, процесс дойки коров. О финансовых эффектах подобных инструментов судить пока сложно: это выставочный инструмент, его задача продемонстрировать масштабное оборудование на обычном столе.

Тренд № 3: массовое использование мессенджеров (Telegram-каналы)

Только в 2016 г. за счет стремительно растущего числа каналов аудитории активных пользователей мессенджера Павла Дурова выросла в 3 раза до 6 млн человек, зацепив самую продвинутую аудиторию российского Интернета, так называемых «трендсеттеров»: журналистов, блогеров, пиарщиков, а актуализировалась только сегодня. Как по теоретической «диффузной» модели коммуникации Э. Роджерса — от «новаторов» к «ранним последователям». Поэтому именно сейчас самое время заводить канал для распространения информации, если нацеливаться на завтрашнее «раннее большинство». Особенно это актуально для компаний, чей бизнес связан с Интернетом. На сегодняшний день свои успешные каналы уже имеют мобильные операторы, банки, продавцы онлайн-услуг. Уже вышло немало статей на тему роста количества «полезных» телеграм-каналов. При этом необходимо осознавать, что в свете последних событий есть вероятность «антитеррористической» блокировки таких систем. Некоторые каналы уже предложили пользователям подписаться на аналогичные страницы в Facebook. Несмотря на существующую угрозу, скорее всего компромисс в том или ином виде будет найден противоборствующими сторонами, что будет играть на руку специалистам по PR.

Тренд № 4: широкое использование возможностей YouTube

Система YouTube в последнее время обретает новое дыхание, не раз уже хоронили эту сеть, но с каждым годом она наполняется все новыми смыслами. Новый виток развития связан с «цифровым поколением Z» (возраст 13–24 лет), потребляющим информацию исключительно визуально. Это подтверждает недавнее исследование Google, в котором утверждается, в частности, что четверть этих пользователей используют YouTube для поиска релевантных ответов на многие жизненные вопросы. То есть видеоконтент становится для них главным источником не только развлечений, но и быстрых ответов на возникающие вопросы. Фактически происходит замещение традиционных поисковиков типа Google и Yandex. При этом, несмотря на то что молодые пользователи часто не обладают финансовыми возможностями для совершения покупки, их мнение влияет на выбор старшего поколения, особенно в вопросах гаджетов и трендовых продуктов. Помимо этого, сеть с конца 2016 г. начала внедрять в интерфейс каналов новый инструмент общения «Сообщество». Теперь создатели каналов смогут размещать свои сообщения, не обязательно привязывая их к видео, они могут содержать только текст, картинку или просто ссылку. Инструмент уже доступен для некоторых западных блогеров и в ближайшее время появится и у отечественных.

Если раньше ключевая функция PR-специалиста сводилась к общению с традиционными или гибридными СМИ (Media relations), а для работы в социальных сетях привлекался SMM-специалист (существовало разделение на офлайн и онлайн), то в условиях digital-трансформации две эти компетенции окончательно слились воедино. В настоящее время ключевой компетенцией PR-специалиста стало управление всеми каналами, при этом модель односторонней коммуникации изменилась на двустороннюю, где во главе угла — вовлечение аудитории как более эффективный способ транслирования информации.

Подводя итог обзору трендов и тенденций, следует отметить, что цифровая трансформация — это реалии сегодняшнего дня, которые необходимо учитывать. Если компания намерена оставаться конкурентоспособной или повышать эффективность, она должна меняться в соответствии с требованиями современности.

Этапы перехода к использованию цифровых технологий в жизненном цикле сложных систем

Реализацию конкретных мер по формированию цифрового экономического пространства и внедрению «сквозных» технологий программы «Цифровая экономика Российской Федерации» по приоритетным направлениям на период до 2024 г. можно условно разбить на два этапа.

На первом этапе (2021–2020) основные действия должны быть направлены на запуск организационных изменений и пилотных проектов: этот этап может объединить все мероприятия, связанные с нормативным регулированием приоритетных сфер научно-технологической политики, организационными изменениями, выработкой новых стандартов и механизмов финансирования предлагаемых изменений. Второй этап (2021–2024) должен быть связан с переходом к действию в регулярном режиме в масштабе всей экономики: вступление в силу и применение правовых норм и организационных изменений, совершенных на предшествующем этапе, масштабирование содержательных мер, предусмотренных к реализации по каждому из направлений. Целесообразно отметить, что для обеспечения дальнейших изменений по итогам двух этапов реализации программы может быть подготовлен новый стратегический пакет действий на перспективу примерно до 2035 г., который будет отталкиваться от уже достигнутых результатов включения России в новую глобальную технологическую реальность, актуальных мировых тенденций технологического развития и позиции России на глобальных высокотехнологичных рынках. В этот период на базе передовых производственных технологий может быть совершена смена моделей развития ряда ключевых секторов российской экономики:

- в сфере энергетики — переход к цифровой и интеллектуальной энергетике, энергетическим мультиагентным системам, развитие технологий так называемой «постуглеродной энергетики», включая масштабное развитие атомной энергетики и ВИЭ;
- в сфере природных ресурсов — переход к предельно рациональному использованию природных ресурсов России, а в отдаленной перспективе — к использованию природоподобных технологий;
- в сфере здравоохранения — переход к «4П»-медицине (предсказательной, превентивной, персонализированной, партисипативной, т. е. предполагающей активное управление здоровьем самим пациентом), а затем и к «5П»-

медицине (так называемой платформенной медицине, т. е. базирующейся на общих программно-информационных и продуктовых платформах);

- в сфере финансовых технологий и управления — переход к мобильным расчетам и электронным финансам, в перспективе — к расчетам на базе технологий распределенных реестров (блокчейн) и управлению на основе алгоритмического регулирования и так называемых «умных» контрактов (автоматических и роботизированных сделок, большая часть которых осуществляется программными агентами);

- в сфере агропромышленного производства — создание новой отрасли «продовольственной системы», включающей в себя не только часть сельского хозяйства (высокоточное земледелие, «Интернет вещей» в АПК, широчайшее распространение биотехнологий, включая генную инженерию и пр.), но и требования к логистическим и торговым системам (короткие цепочки поставок, обеспечение качества продовольствия, рациональное использование последнего и сокращение отходов), к индивидуальным диетам, функциональным продуктам.

Цифровые оценки промышленной коммерциализации технологий

Промышленная коммерциализация технологий — деятельность, направленная на получение дохода от использования результатов научных исследований. Это самое простое определение понятия, учебное пособие по которому вы держите в руках.

Рыночные реформы в России изменили лицо российской науки, как следствие — произошло изменение в подходах в формировании отечественных производственных технологий, цифровая трансформация затронула целый ряд отраслей. Эти изменения проявились прежде всего в децентрализации управления и изменении схем финансирования научных исследований. Однако российская наука по-прежнему остается по сути государственной: подавляющее большинство научных учреждений, экспериментальное оборудование и опытные производства принадлежат государству, а ученые получают заработную плату в основном из бюджетных средств. В этом заключается основное отличие нашей науки от западной, где существенная доля исследовательских работ выполняется в частных лабораториях и научных центрах. В то же время на протяжении последних лет наша наука финансируется на уровне нескольких десятков процентов от требуемого объема средств.

Многие экономисты и политики понимают необходимость серьезной реорганизации системы финансирования науки, что означает прежде всего более широкое привлечение частных инвестиций. Однако частные инвесторы вкладывают средства в научные изыскания только в расчете на будущую прибыль. Это осуществляется путем закрепления прав инвесторов на результаты научных исследований и использования их в дальнейшем для производства новых товаров или для последующей перепродажи. Фактически, частные инвестиции в науку означают приобретение опциона на специфический товар — интеллектуальную собственность, — способный в дальнейшем приносить прибыль. Для того чтобы стать таким товаром, научные исследования должны быть облечены в соответствующую «упаковку», что означает как минимум проведение предварительного комплексного изучения финансовых и рыночных перспектив научных разработок и доказательства их будущих экономических и технологических преимуществ.

Описанный выше процесс инвестирования в перспективные научные разработки носит название *коммерциализации технологий*. Коммерциализация предполагает прогнозирование рынков, поиск, экспертизу и отбор разработок для финансирования, привлечение инвестиций, распределение и юридическое закрепление прав на будущую интеллектуальную собственность между всеми участвующими в процессе сторонами, управление научным проектом, внедрение результатов в производство, дальнейшую модификацию и сопровождение интеллектуального продукта. Желательно, чтобы все эти мероприятия выполнялись в рамках цифровой платформы, где практически автоматически можно проводить расчеты вариантов.

К сожалению, в силу оторванности науки от бизнеса развитие технологий в России происходит по своим собственным законам, без учета конкретных потребностей промышленного производства, и в большинстве случаев — без использования цифровых технологий. Чаще всего процесс коммерциализации начинается в результате счастливой встречи ученого или инженера — носителя некоторой передовой идеи — и предпринимателя, способного эту идею оценить и поддержать в финансовом плане. То, что такие встречи происходят редко, и мы имеем на сегодняшний день не так много примеров успешной коммерциализации технологий, объясняется прежде всего тем, что у российских участников этого процесса отсутствует понимание механизмов функционирования современного рынка интеллектуальной собственности и правил игры на нем. Особенно эта тенденция проявляется у ученых, которые воспринимают коммерциализацию своих идей прежде всего как проблему поиска финансовых средств для продолжения исследований. Многие ученые достаточно поверхностно относятся к вопросам оценки окупаемости своих разработок и необходимости учета интересов инвесторов. Все эти обстоятельства, безусловно, осложняют работу по поиску перспективных для коммерческого использования технологий в России.

Важно отметить также отличие коммерциализации технологий и понятия «трансфер технологий». Понятие «трансфер технологий» появилось в русскоязычной научной литературе сравнительно недавно и напрямую связано с переориентацией на рыночные отношения в большинстве сфер человеческой деятельности. Часто его употребляют в связке с другим понятием — «коммерциализация технологий», хотя смысловое содержание этих понятий неодинаково. Англоязычное слово «трансфер» успешно заменило насильственный термин «внедрение», которым административно-командная система наградила процесс претворения в жизнь инновационного предложения. Однако это не простое замещение, а существенное преобразование смысла процесса. Вместо насильственного «внедрения» (предполагающего активное или пассивное сопротивление среды, в которую производится это «внедрение» чего-то инородного) «трансфер» предполагает не только передачу информации о новшестве, но и ее освоение при активном позитивном участии и источника этой информации (например, автора изобретения), и реципиента, приемника и реализатора на практике информации о новой технологии, и конечного пользователя продукта, производимого с помощью этой технологии. Поэтому, кстати, основной акцент при трансфере технологии делается не столько на технологии как таковой, сколько на субъектах — участниках этого процесса.

Понятие «коммерциализация технологии» предполагает обязательное коммерческое использование информации о технологии, т. е. использование с обязательным извлечением выгоды. Чаще всего эта выгода непосредственно измеряется

в конкретных денежных единицах, гораздо реже — в тех же единицах, но опосредованно, через, например, увеличение эффективности другой технологии. Но деньги в этих расчетах присутствуют всегда и являются определяющим критерием успешности процесса. В то же время вопрос о том, кто, какой субъект осуществляет непосредственное использование технологии, при коммерциализации не является первостепенным, и в частности коммерциализацией нередко пытается заняться сам автор, первоисточник новой технологии (физическое лицо или организация).

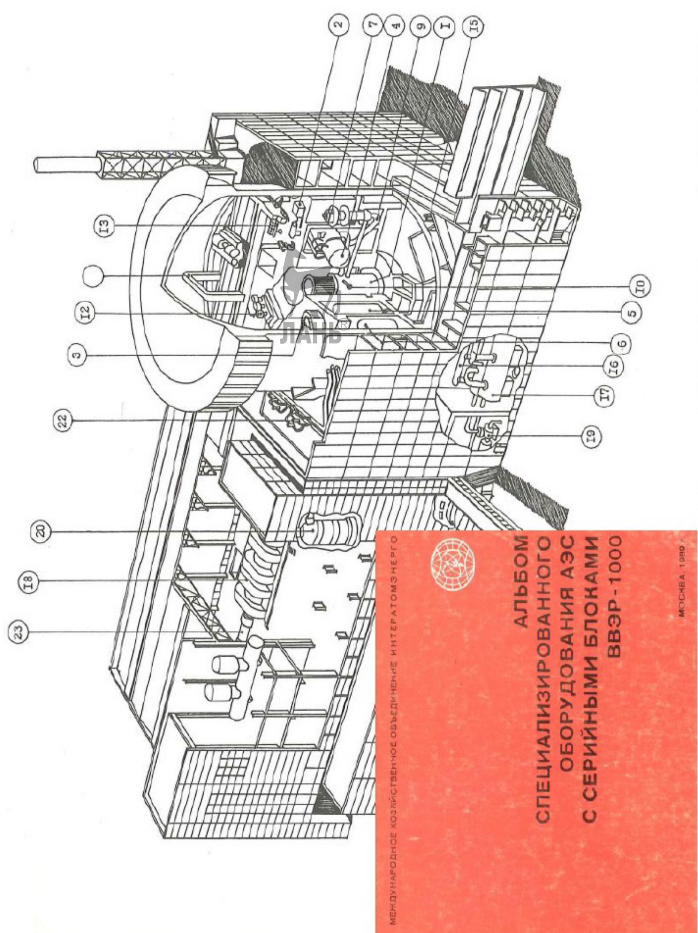
Исторический пример масштабной коммерциализации промышленных технологий

Успешным и в своем роде уникальным опытом такого рода в области распространения передовых промышленных технологий можно назвать в эпоху деятельности административно-командной системы программу Совета экономической взаимопомощи (СЭВ) в области атомной энергетики в 1970-х гг. Уникальность и значимость Программы заключалась в том, что за достаточно короткий отрезок времени в странах — членах СЭВ был создан мощный производственно-промышленный потенциал для изготовления специализированного оборудования для АЭС, сооружаемых в этих странах (локализация), что позволило обеспечить непрерывное и поточное строительство АЭС в Чехословакии, а также сооружение АЭС «Пакш-2» в Венгрии и Ровенской АЭС в СССР (ныне — Украина) и др. Были определены предприятия — участники системы кооперации, на которые комплексно и последовательно был осуществлен трансфер технологий изготовления оборудования для АЭС с ядерными реакторами типа ВВЭР. Это создавало экономико-технический базис сотрудничества и дало значительный рывок в развитии атомного энергомашиностроения стран-участниц.

На рисунке 4 представлена схема из документа Международного хозяйственного объединения (МХО) «Интератомэнерго», показывающая распределение обязанностей между социалистическими странами при производстве специализированного оборудования АЭС с серийными блоками ВВЭР-1000. Этот пример свидетельствует о том, что в любой общественно-экономической формации находят инструменты и механизмы, которые способствуют распространению передовых технологий за рубежи отдельных государств, фактически реализуя подходы, которые в современную эпоху характеризуются термином «глобализация». Таким образом, экономические принципы и закономерности являются объективной реальностью, приспособление которых к определенной политической ситуации является скорее искусством, чем наукой. В то же время международное разделение труда, параметры которого в современную эпоху постоянно изменяются и модифицируются, является устойчивым трендом, и все инструменты промышленной коммерциализации технологий и реализации крупных промышленных инноваций, описанные в настоящем пособии, — это живая экономическая реальность. Российская промышленность, базовые параметры которой закладывались еще в бывшем СССР, должна развиваться и совершенствоваться, перенимая в зарубежном опыте самое передовое, но не растрачивая и положительного потенциала, накопленного десятилетиями упорного труда наших предшественников. В последующих разделах пособия будут приведены отдельные примеры и зарубежного опыта (Южная Корея, КНР) трансфера ядерных технологий, что свидетельствует об объективном характере процессов глобализации развития наукоемких технологий и локализации производственных возможностей их осуществления.

**СПЕЦИАЛИЗАЦИЯ СТРАН-ЧЛЕНОВ СЭВ И СФРЮ
В ПРОИЗВОДСТВЕ И ПОСТАВКАХ ОБОРУДОВАНИЯ
ДЛЯ АЭС С РЕАКТОРАМИ ТИПА ВВЭР-1000**

СССР	СЭВ	СФРЮ	И	И	И	И
1 Контрольно-измерительная аппаратура в комплексе						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
2 Станки для ремонта реакторов и аппаратура						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
3 Тяловины						
СССР	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
4 Парогенераторы						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
5 Компрессоры (двухвал) в сборе						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
6 Тягловозы						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
7 Тяговые дизельные насосы						
СССР	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
9 Тяговые турбогенераторы						
СССР	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
10 Турбогенераторы для выработки электроэнергии						



СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
24 Контрольно-измерительная аппаратура в комплексе						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
25 Тягловые насосы						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
22 Аппаратура Оланда Вилле						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
21 Тягловые насосы						
СССР	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
20 Станки для ремонта парогенераторов						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
19 Насосы для АЭС						
СССР	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
18 Турбогенераторы						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
17 Оборудование системы защиты						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
16 Тягловое оборудование						

СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
13 Тягловое оборудование						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
15 Оборудование выработки электроэнергии						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
12 Парогенераторы						
СФРЮ	СЭВ	СФРЮ	СЭВ	СФРЮ	СФРЮ	СФРЮ
11 Крыша, лестница (включая другие)						

Рис. 4

Описание специализации стран — членов СЭВ и СФРЮ в производстве и поставках оборудования для АЭС с реакторами ВВЭР-1000 в эпоху деятельности СССР и стран социалистического лагеря

Западный бизнесмен прекрасно понимает, что одним-единственным удачным изобретением, доведенным до массового производства, он может обеспечить безбедное будущее себе и своим потомкам, поэтому он активно ищет такое изобретение и вкладывает деньги в изобретение и в изобретателя (разумеется, за соответствующую долю в будущем предприятии). Однако он понимает и то, что это вложение средств — долгосрочное, это не челночный торговый бизнес с длительностью «производственного цикла» в недели, а многостадийный и тщательно планируемый процесс, сочетающий прогнозирование потребностей и научное исследование технологии, и ее масштабирование и оптимизацию, и маркетинг, и заботы о сбытовой сети, и многое-многое другое.

Менталитет же отечественных бизнесменов в большинстве случаев пока еще далек от такого скрупулезного подхода и поиска революционных технологий или продуктов, да и с финальной частью проблемы (маркетинг и сбыт) у нас пока отношения отнюдь не блестящие. Поэтому мы имеем достаточно сложную картину: по широким российским научным и технологическим полям рыщут акулы развитого капитализма и скупают на корню самые экономически перспективные инновационные разработки. Причем они бы и рады организовать производство по этим разработкам здесь же (хотя бы из-за дешевизны рабочей силы), но, с одной стороны, легионы «доброжелателей», воспитанные в известном духе, ставят столько препятствий, что никакого терпения все это понять и пережить не хватает, а с другой стороны, наличная производственная база и ее владельцы настолько озабочены перманентным форс-мажором, что найти с ними общий язык преуспевающему западному бизнесмену непросто.

Еще одна особенность словоупотребления в этом курсе состоит в несколько необычном употреблении слова «технология». В действительности «видимым» предметом трансфера может быть и конкретный объект предметного типа (например, материал, система, микропроцессор нового типа и пр.), который сам по себе технологией не является. Однако и в этом случае с инновационным предметом обязательно связано то или иное умение, искусство, та или иная технология (либо технология производства, либо технология применения или осуществления), поэтому такие словосочетания, как «коммерциализация технологий», так и «трансфер технологии», всегда несут вполне определенный реальный смысл. Этот смысл становится еще более понятным, если учесть, что наиболее надежным способом защиты любого объекта и любой технологии как интеллектуальной собственности является не патент, а так называемое «нераскрываемое ноу-хау».

Ведь почему наши ведомства, занимающиеся экспортом технологий, так настойчиво выясняют, есть ли в предлагаемой на экспорт технологии это самое «нераскрываемое ноу-хау»: именно потому, что обойти практически любой патент для современных мастеров этого дела не составляет никакого труда, если патентный документ содержит информацию, достаточную для запуска технологии. А доказывать нарушение патентных прав — это такая дорогостоящая и сомнительная (в смысле успешности) операция, которая не только нам, но и более богатым странам представляется крайне нежелательной.

Так что в настоящем пособии смысл понятия «технология» ближе всего не к устоявшемуся производственному смыслу, а к его прямому переводу (tehne —

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего
образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ ПРАКТИЧЕСКИХ
РАБОТ**

Дисциплина (модуль)	Управление проектами в области искусственного интеллекта <i>наименование дисциплины (модуля)</i>
Уровень образования	магистратура <i>(СПО/бакалавриат/магистратура/специалитет/подготовка кадров высшей квалификации)</i>
Квалификация	Магистр <i>Техник/Бакалавр/Магистр/Инженер/ Исследователь. Преподаватель-исследователь</i>

г. Ульяновск, 2021

Практическая работа 1

Понятие инновационного проекта. Определение тематики проекта в области ИИ

1. Проанализировать актуальные известные действующие проекты в области искусственного интеллекта (ИИ).
2. Обсудить проекты с преподавателем. Определить инновационную составляющую известных проектов.
3. Сгенерировать собственную тематику возможного проекта в области искусственного интеллекта.
4. В отчет по практической работе включить обзор актуальных известных действующих проектов с описанием инновационной составляющей, предложение по тематике возможного проекта в области искусственного интеллекта с описанием инновационного эффекта.
5. Подготовить отчет по практической работе.
7. Также необходимо быть готовым устно ответить на контрольные вопросы.

Методические указания

Практическая работа может быть выполнена с помощью поиска информации в сети Интернет

Контрольные вопросы

1. Содержание понятие «инновационный проект». Системное представление проекта.
2. Примеры инновационных проектов в области искусственного интеллекта.

Практическая работа № 2-3.

Планирование проекта. Организационная структура проекта

Преподаватель анализирует сформированные студентами предлагаемые темы инновационного проекта в области ИИ. В аудитории проводится голосование / он-лайн голосование по оценке привлекательности предлагаемых тематик проектов. По результатам голосования отбирается несколько проектов, рекомендуемых для разработки в рамках данной дисциплины. Авторы проектных идей, набравшие наибольшее число голосов, получают возможность сформировать проектную команду из числа студентов, изучающих данную дисциплину (до 4 человек в одной команде в зависимости от числа студентов). По определенной тематике разрабатываемого проекта в области искусственного интеллекта разрабатывается организационная структура проекта.

Что такое организационная структура проекта

Организационная структура проекта – это временная организационная структура, созданная для повышения качества управления и взаимодействия в проекте путем определения и визуализации процессов взаимодействия как между внутренними, так и с внешними участниками проекта.

условные типы организационных структур проекта:

Организационная структура управления проектом. Предназначена для определения уровней принятия решений

Организационная структура выполнения проекта предназначена для организации взаимодействия между командами, вовлеченными в проект (архитектура, тестирование, разработка, анализ и проч.).

Организационная структура работы с подрядчиком или подрядчиками в проекте. Согласуется на уровне ответственных за проект от каждой вовлеченной стороны для определения процесса работы и точек принятия решений.

Организационная структура программы проектов. Согласуется на уровне руководителя программы и ее возможного спонсора (в случае наличия спонсоров, готовых профинансировать открытие стартапа) для определения процесса взаимодействия между проектами (и, конечно, руководителями проектов), включенными в программу.

Понять, кто вообще будет вовлечен в проект. Понять, достаточно ли вам будет одной структуры или необходимо построить несколько, и для чего вообще вы ее строите. Например, организационная структура управления проектом, которую вы будете согласовывать на уровне управляющего комитета будет отличаться от организационной структуры выполнения проекта для организации взаимодействия между командами или от организационной структуры, которую вы делаете, чтобы четко определить процесс взаимодействия с подрядчиками в этом проекте.

Разработать слайд презентации, например, в MS visio, mindmap или в любом другом инструменте список всех участников.

Определить, какую информацию помимо ролей вам необходимо видеть. Обычно это как минимум должности и подчиненность, а как максимум – уровень принимаемых решений, конкретные имена, регулярность встреч и проч. Пытаться вставить туда все я не

рекомендую – для этого есть план коммуникаций, а картинку с оргструктурой лучше этим не перегружать.

Прорисовать подчиненность/иерархию и направления коммуникации.

Эффектно оформить организационную структуру, избавившись от всей лишней информации, «потерявшихся» людей и стрелок и проч. Организационная структура проекта – один из основополагающих документов и должен выглядеть эффектно, чтобы его воспринимали всерьез.

Показать получившуюся оргструктуру проекта кому-нибудь, не входящему в нее, но понимающему контекст. Этот человек сможет вам подсказать, что в ней непонятно, и, возможно, обратит внимание на какие-то логические или политические несоответствия, т.к. в процессе разработки взгляд все-таки замыливается.

В случае наличия возможного спонсора согласовать построенную организационную структуру со спонсором проекта или с другими заинтересованными лицами, чем мнение неплохо бы получить до обнародования проекта.

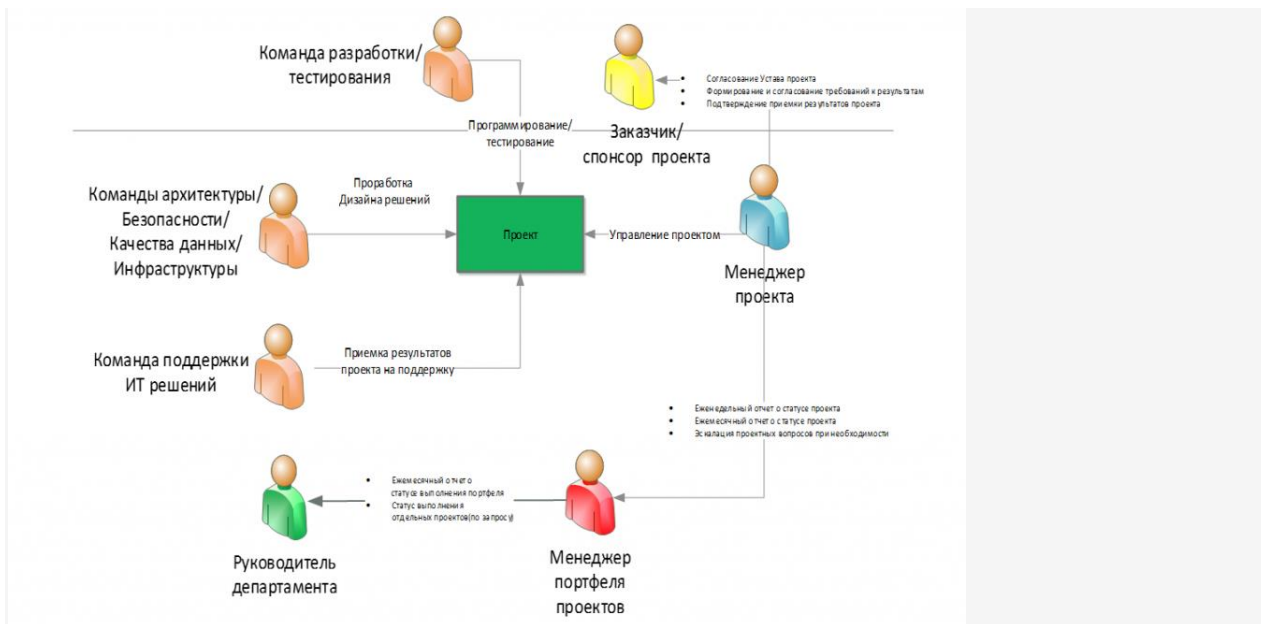


Рис 1. Организационная структура выполнения проекта

Примечание. Это пример реального проекта, для студенческой работы схема будет более общей.

Практическая работа №4-5.

Подбор персонала проекта. Распределение обязанностей и активностей проекта.

На данном этапе детально должны быть прописаны роли всех членов команды, а также возможные роли среди иных специалистов, не присутствующих среди студентов, например, инженеров по знаниям, веб-дизайнеров, программистов, бухгалтеров и пр. Роли прописываются и оформляются в виде таблицы 1.

Таблица 1.

Распределение функционала проектной команды

№пп	Наименование должности в проекте	ФИО	Число ставок
1	Руководитель	Степанов Илья Сергеевич	1
		

Практическая работа №6

Процессы и функции управления проектами. Основные и вспомогательные процессы в управлении проектами.

Студенты в командах продолжают работать над проектом.

1. По сформированным проектным идеям необходимо расписать процессы управления проектами.
2. Подготовить презентацию по процессам управления проектом

Управление проектами — это приложение знаний, опыта, методов и средств к работам проекта для удовлетворения требований, предъявляемых к проекту, и ожиданий участников проекта. Чтобы удовлетворить этим требованиям и ожиданиям, необходимо найти оптимальное сочетание между целями, сроками, затратами, качеством и другими характеристиками проекта.

Управление проектами подчиняется четкой логике, которая связывает между собой различные области знаний и процессы управления проектами.

Прежде всего у проекта обязательно имеются одна или несколько целей. Под целями мы будем далее понимать не только конечные результаты проекта, но и выбранные пути достижения этих результатов (например, применяемые в проекте технологии, система управления проектом).

Достижение целей проекта может быть реализовано различными способами. Для сравнения этих способов необходимы критерии успешности достижения поставленных целей. Обычно в число основных критериев оценки различных вариантов исполнения проекта входят сроки и стоимость достижения результатов. При этом запланированные цели и качество обычно служат основными ограничениями при рассмотрении и оценке различных вариантов. Конечно, возможно использование и других критериев и ограничений, в частности ресурсных.

Для управления проектами необходимы рычаги. Влиять на пути достижения результатов проекта, цели, качество, сроки и стоимость исполнения работ можно, выбирая применяемые технологии, состав, характеристики и назначения ресурсов на выполнение тех или иных работ. Таким образом, применяемые технологии и ресурсы проекта можно отнести к основным рычагам управления проектами. Кроме этих основных существуют и вспомогательные средства, предназначенные для управления основными. К таким вспомогательным рычагам управления можно отнести, например, контракты, которые позволяют привлечь нужные ресурсы в нужные сроки. Кроме того, для управления ресурсами необходимо обеспечить эффективную организацию работ. Это касается структуры управления проектом, организации информационного взаимодействия участников проекта, управления персоналом.

Информация, используемая в управлении проектами, обычно не бывает стопроцентно достоверной. Учет неопределенности исходной информации необходим и при планировании проекта, и для грамотного заключения контрактов. Анализ и учет неопределенностей посвящен анализ рисков.

Любой проект в процессе своей реализации проходит различные стадии, называемые в совокупности жизненным циклом проекта. Для реализации различных функций управления проектом необходимы действия, которые в дальнейшем именуется процессами управления проектами.

Процессы управления проектами

Управление проектами — интегрированный процесс. Действия (или их отсутствие) в одном направлении обычно влияют и на остальные направления. Такая взаимосвязь заставляет балансировать между задачами проекта — часто улучшение в одной области может быть достигнуто лишь за счет ухудшения в другой. Для лучшего понимания интегрированной

природы управления проектами опишем его через процессы, из которых оно состоит, и их взаимосвязи.

Проект состоит из процессов. Процесс — это совокупность действий, приносящая результат. Процессы проекта обычно выполняются людьми и распадаются на две основные группы:

- процессы управления проектами — касающиеся организации и описания работ проекта (которые будут подробно описаны далее);
- процессы, ориентированные на продукт — касающиеся спецификации и производства продукта. Эти процессы определяются жизненным циклом проекта и зависят от области приложения.

В проектах процессы управления проектами и процессы, ориентированные на продукт, накладываются и взаимодействуют. Например, цели проекта не могут быть определены при отсутствии понимания того, как создать продукт.

Процессы управления проектами могут быть разбиты на шесть основных групп, реализующих различные функции управления:

- процессы инициации — принятие решения о начале выполнения проекта;
- процессы планирования — определение целей и критериев успеха проекта и разработка рабочих схем их достижения;
- процессы исполнения — координация людей и других ресурсов для выполнения плана;
- процессы анализа — определение соответствия плана и исполнения проекта поставленным целям и критериям успеха и принятие решений о необходимости применения корректирующих воздействий;
- процессы управления — определение необходимых корректирующих воздействий, их согласование, утверждение и применение;
- процессы завершения — формализация выполнения проекта и подведение его к упорядоченному финалу.

Процессы управления проектами накладываются друг на друга и происходят с разной интенсивностью на всех стадиях проекта, как показано на рис. 1.

Кроме того, процессы управления проектами связаны своими результатами — результат выполнения одного становится исходной информацией для другого. Эти взаимосвязи проиллюстрированы на рис. 2.

И наконец, имеются взаимосвязи групп процессов различных фаз проекта. Например, закрытие одной фазы может являться входом для инициации следующей фазы (пример: завершение фазы проектирования требует одобрения заказчиком проектной документации, которая необходима для начала реализации).

В реальном проекте фазы могут не только предшествовать друг другу, но и накладываться.

Повторение инициации на разных фазах проекта помогает контролировать актуальность выполнения проекта. Если необходимость его осуществления отпала, очередная инициация позволяет вовремя это установить и избежать излишних затрат.

Взаимосвязи процессов

Процессы инициации

Инициация включает единственный подпроцесс — авторизацию, то есть решение начать следующую фазу проекта.

Процессы планирования

Планирование имеет большое значение для проекта, поскольку проект содержит то, что ранее не выполнялось. Естественно, что планирование включает сравнительно много процессов. Однако не следует считать, что управление проектами — это в основном



Рис. 2.
Взаимосвязи групп процессов управления проектами в фазе

планирование. Усилия, прилагаемые для планирования, следует соизмерять с целями проекта и полезностью полученной информации.

Напомним, что следует различать цели проекта и цели продукта проекта, под которым понимается продукция (или услуги), созданная или произведенная в результате исполнения проекта.

- Цели продукта — это свойства и функции, которыми должна обладать продукция проекта.
- Цели проекта — это работа, которую нужно выполнить для производства продукта с заданными свойствами.



Рис. 3. Взаимосвязи процессов планирования

Взаимосвязи между процессами планирования представлены на рис. 3.

В ходе исполнения проекта эти процессы многократно повторяются. Изменениям могут подвергнуться цели проекта, его бюджет, ресурсы и т. д. Кроме того, планирование проекта — это не точная наука. Различные команды проекта могут разработать различные планы для одного и того же проекта. А пакеты управления проектами могут составить различные расписания выполнения работ при одних и тех же исходных данных.

Некоторые из процессов планирования имеют четкие логические и информационные взаимосвязи и выполняются в одном порядке практически во всех проектах. Так, например, сначала следует определить, из каких работ состоит проект, а уж затем рассчитывать сроки выполнения и стоимость проекта. Эти основные процессы выполняются по несколько раз на протяжении каждой фазы проекта.

Кроме перечисленных основных процессов планирования имеется ряд вспомогательных

процессов, необходимость в использовании которых сильно зависит от природы конкретного проекта:

- планирование качества — определение того, какие стандарты качества использовать в проекте, и того, как этих стандартов достичь;
- планирование организации — определение, документирование и назначение ролей, ответственности и взаимоотношений отчетности в организации;
- назначение персонала — назначение человеческих ресурсов на выполнение работ проекта;
- планирование взаимодействия — определение потоков информации и способов взаимодействия, необходимых для участников проекта;
- идентификация риска — определение и документирование событий риска, которые могут повлиять на проект;
- оценка риска — оценка вероятностей наступления событий риска, их характеристик и влияния на проект;
- разработка реагирования — определение необходимых действий для предупреждения рисков и реакции на угрожающие события;
- планирование поставок — определение того, что, как и когда должно быть поставлено;
- подготовка условий — выработка требований к поставкам и определение потенциальных поставщиков.

Взаимосвязи между вспомогательными подпроцессами, как и само их наличие, в большой мере зависят от природы проекта.

Процессы исполнения и контроля

Под исполнением подразумеваются процессы реализации составленного плана. Исполнение проекта должно регулярно измеряться и анализироваться для того, чтобы выявить отклонения от намеченного плана и оценить их влияние на проект. Регулярное измерение параметров проекта и идентификация возникающих отклонений далее также относится к процессам исполнения и именуется контролем исполнения. Контроль исполнения следует проводить по всем параметрам, входящим в план проекта.



Рис. 4. Взаимосвязи процессов исполнения

Как и в планировании, процессы исполнения (рис. 4) можно подразделить на основные и вспомогательные.

К основным можно отнести сам процесс исполнения плана проекта.

Среди вспомогательных процессов отметим:

- учет исполнения — подготовка и распределение необходимой для участников проекта информации с требуемой периодичностью;
- подтверждение качества — регулярная оценка исполнения проекта с целью подтверждения соответствия принятым стандартам качества;
- подготовка предложений — сбор рекомендаций, отзывов, предложений, заявок и т. д.;
- выбор поставщиков — оценка предложений, выбор поставщиков и подрядчиков и заключение контрактов;
- контроль контрактов — контроль исполнения контрактов поставщиками и подрядчиками;
- развитие команды проекта — повышение квалификации участников команды проекта.

Процессы анализа

Процессы анализа включают как анализ плана, так и анализ исполнения проекта.

Анализ плана означает определение того, удовлетворяет ли составленный план исполнения проекта предъявляемым к проекту требованиям и ожиданиям участников проекта. Он выражается в оценке показателей плана командой и другими участниками проекта. На стадии планирования результатом анализа плана может быть принятие решения о необходимости изменения начальных условий и составления новой версии плана либо принятие разработанной версии в качестве базового плана проекта, который в дальнейшем служит основой для измерения исполнения. В дальнейшем изложении анализ плана не выделяется в качестве отдельной группы процессов, а включается в группу процессов планирования, делая эту группу процессов по своей природе итеративной. Таким образом,



Рис. 5. Взаимосвязи процессов анализа

под процессами анализа в дальнейшем понимаются процессы анализа исполнения.

Процессы анализа исполнения предназначены для оценки состояния и прогноза успешности исполнения проекта согласно критериям и ограничениям, определенным на стадии планирования. В силу уникальности проектов эти критерии не являются универсальными, но для большинства проектов в число основных ограничений и критериев успеха входят цели, сроки, качество и стоимость работ проекта. При отрицательном прогнозе принимается решение о необходимости корректирующих воздействий,

выбор которых осуществляется в процессах управления изменениями.

Процессы анализа также можно подразделить на основные и вспомогательные.

К основным относятся те процессы анализа, которые непосредственно связаны с целями проекта и показателями, характеризующими успешность исполнения проекта:

- анализ сроков — определение соответствия фактических и прогнозных сроков исполнения операций проекта директивным или запланированным;
- анализ стоимости — определение соответствия фактической и прогнозной стоимости операций и фаз проекта директивным или запланированным;
- анализ качества — мониторинг результатов с целью их проверки на соответствие принятым стандартам качества и определения путей устранения причин нежелательных результатов исполнения качества проекта;
- подтверждение целей — процесс формальной приемки результатов проекта его участниками (инвесторами, потребителями и т. д.).

Вспомогательные процессы анализа связаны с анализом факторов, влияющих на цели и критерии успеха проекта. Эти процессы включают:

- оценку исполнения — анализ результатов работы и распределение проектной информации с целью снабжения участников проекта данными о том, как используются ресурсы для достижения целей проекта;
- анализ ресурсов — определение соответствия фактической и прогнозной загрузки и производительности ресурсов запланированным, а также анализ соответствия фактического расхода материалов плановым значениям.

В число процессов анализа не включены анализ взаимодействия с целью оптимизации процедур обработки информации, анализ исполнения контрактов с целью своевременного внесения изменений и предотвращения споров и ряд других процессов, которые не носят регулярного характера (как анализ взаимодействия) либо составляют часть включенных процессов (как анализ контрактов).

В результате анализа либо принимается решение о продолжении исполнения проекта по намеченному ранее плану, либо определяется необходимость применения корректирующих воздействий.

Процессы управления

Управление исполнением проекта — это определение и применение необходимых управляющих воздействий с целью успешной реализации проекта. Если исполнение проекта происходит в соответствии с намеченным планом, то управление фактически сводится к исполнению — доведению до участников проекта плановых заданий и контролю их реализации. Эти процессы нами включены в процессы исполнения.

Другое дело, если в процессе реализации возникли отклонения, анализ которых показал, что необходимо определение и применение корректирующих воздействий. В этом случае требуется найти оптимальные корректирующие воздействия, скорректировать план оставшихся работ и согласовать намеченные изменения со всеми участниками проекта. Итак, процессы управления предназначаются для определения, согласования и внесения необходимых изменений в план проекта. Такие процессы управления часто называются управлением изменениями и иницируются процессами анализа (рис. 6).



Рис. 6. Взаимосвязи процессов управления

К основным процессам управления, встречающимся практически в каждом проекте, относятся:

- общее управление изменениями — определение, согласование, утверждение и принятие к исполнению корректирующих воздействий и координация изменений по всему проекту;

- управление ресурсами — внесение изменений в состав и назначения ресурсов на работы проекта;
- управление целями — корректировка целей проекта по результатам процессов анализа;
- управление качеством — разработка мероприятий по устранению причин неудовлетворительного исполнения.

Среди вспомогательных процессов управления отметим:

- управление рисками — реагирование на события и изменение рисков в процессе исполнения проекта;
- управление контрактами — координация работы (суб)подрядчиков, корректировка контрактов, разрешение конфликтов.



Рис. 7. Взаимосвязи процессов завершения

Процессы завершения

Завершение проекта сопровождается следующими процессами (рис. 7):

- закрытие контрактов — завершение и закрытие контрактов, включая разрешение всех возникших споров;
- административное завершение — подготовка, сбор и распределение информации, необходимой для формального завершения проекта.

Методы и технологии реализации перечисленных процессов, их интеграция составляют сущность управления проектами.

Обратите внимание, что все перечисленные процессы приложимы к проектам любой природы — и к строительным, и к информационным, и к любым другим. Однако имеются и существенные отличия в управлении проектами различных типов. Следует также отметить, что успешное внедрение системы управления проектами связано с определенной организационной перестройкой и с внедрением специализированных программных средств. Перечисленные вопросы, а также специализированные методы решения отдельных задач управления проектами, технология, опыт и проблемы внедрения будут раскрыты в последующих публикациях.

Практическая работа №7.

Понятие инициации, планирования, выполнения, контроля и закрытия проекта.

Студенческие команды определяют основные процессы управления проектом и готовят презентации по следующим 5 стандартным процессам управления проектом: инициации, планирования, выполнения, контроля и закрытия проекта.

Стандартная схема жизненного цикла проекта состоит из пяти фаз:

- Инициация
- Планирование
- Исполнение
- Контроль
- Завершение

Обычно, когда начинается новый проект, эти пять фаз следуют друг за другом по порядку, но так бывает не всегда. Например, в случае каких-то изменений в ходе выполнения проекта вы можете вернуться к фазе планирования, чтобы учесть изменения, но не станете снова повторять весь цикл, начиная с фазы инициации.

Такой уровень гибкости упрощает процессы управления изменениями в ходе жизненного цикла проекта.

В чем важность жизненного цикла проекта? Эта последовательность фаз может показаться вам формальностью, но на самом деле она очень помогает. Она позволяет применять организованный подход к управлению проектом. Это значит, что у вас будут более четкие роли и обязанности, улучшенное взаимодействие и стабильные результаты (а упущений и забытых задач станет намного меньше!).

Фаза 1. Инициация

Это подготовительный этап, когда вам нужно убедиться, что проект действительно можно осуществить, прежде чем вкладывать силы в планирование и последующие задачи.

В контексте нашей работы этот этап включает в себя описание проекта, создание экономического обоснования, выявление ключевых участников и утверждение проекта соответствующими сторонами.

Не нужно недооценивать важность фазы инициации. На этом этапе закладывается фундамент для успешного выполнения проекта.

Когда вы займетесь задачами и сроками, будет очень легко забыть о том, в чем заключается важность проекта. На самом деле четко представляют себе бизнес-цели проекта только 55% участников, преимущественно руководители команд и менеджеры проектов.

При описании проекта нужно обязательно указать конкретные преимущества для бизнеса, связанные с вашим проектом. Это поможет вам и участникам вашей команды не только нацелиться на достижение целей, но и заранее убедиться, что время не будет потрачено зря.

Составление устава проекта поможет вам предоставить руководителям конкретные факты и получить их одобрение.

Фаза 2. Планирование

На этапе планирования вы составляете смету и назначаете дату, когда проект будет запущен в опытную эксплуатацию.

Трудно переоценить важность тщательного планирования, но очень многие команды и компании пропускают этот этап, чтобы сразу приступить к работе. Таким образом, фаза планирования стоит того, чтобы вложить в нее время и силы. Будьте уверены, это сэкономит вам кучу времени и нервов на последующих этапах.

Фаза планирования требует тщательного обдумывания множества элементов, связанных с вашим проектом. В числе прочего это могут быть следующие элементы:

- Ключевые роли и сферы ответственности
- Показатели успеха
- Возможные риски и препятствия, снижающие эффективность
- Ожидания в отношении внутрикомандного взаимодействия

- Календарный план проекта

Все эти элементы должны быть описаны в подробном плане проекта, который вы представите команде на организационном совещании, а затем станете ссылаться на него на протяжении работы над проектом.

Кстати, об организационном совещании. Оно заслуживает особого внимания, поскольку является одним из важных этапов планирования.

В ходе организационного совещания вы и участники проектной группы должны будете обсудить несколько важнейших элементов плана, в том числе сферы ответственности, рекомендации по организации взаимодействия и критерии успешности проекта. Также можно выяснить, какие программные решения для управления жизненным циклом проекта или другие инструменты вы будете использовать.

Помните, что это совещание не просто дает вам возможность донести до всех свою идею и раздать указания. Выделите время для вопросов и отзывов участников проектной группы, чтобы заранее разрешить любые сомнения и избежать недоразумений. В результате у вас должны сформироваться общие представления о том, как будет выполняться проект.

Фаза 3. Исполнение

Во время этой фазы разработанные планы начинают выполняться. Как ни странно, это именно та фаза, когда все может пойти не так, как планировалось.. Планировать действия гораздо проще, чем эти действия выполнять.

Но помните о том, что ничто не случится само по себе только лишь потому, что у вас есть план. Необходимо убедиться, что каждый участник придерживается этого плана и выполняет свою часть работы.

Сразу же после организационного совещания займитесь распределением дополнительных ресурсов, раздайте указания к действию, ссылки и план проекта, чтобы участники могли сверяться с ним. Воспользуйтесь моментом и не теряйте времени, а сразу приступайте к делу.

Фаза 4. Контроль

Эта фаза обычно совпадает по времени с этапом выполнения проекта. Пока разработчики работают, например, над программным обеспечением, вы тщательно следите за ходом работ и заодно сверяетесь со сметой и графиком, чтобы удостовериться, что все идет по плану.

Учтите, что жизненный цикл управления проектом требует определенной гибкости.

Во время фазы контроля могут возникать ситуации, требующие корректировки первоначального плана. Это нормально — именно для этого фаза контроля и нужна!

Организационные совещания необходимы для запуска проектного механизма в действие, но кроме него периодически следует проводить контрольные совещания для мониторинга хода работ и корректировки курса. Кроме того, такие встречи — отличная возможность для оперативного сбора отзывов, чтобы не запускать проблемные или конфликтные ситуации.

Использование платформы управления проектами обеспечивает прозрачность данных о процессе. Этот инструмент позволит вам получать общее представление о ходе выполнения проекта, а также отслеживать отдельные задачи и рабочую загрузку исполнителей, обходясь без лишних совещаний.

Следя за ходом работ, обязательно уделяйте внимание ресурсам (в число которых всходит абсолютно все — от времени до материалов). Выясняйте, выделяются ли они в нормальном объеме или их не хватает. При управлении ресурсами очень важно действовать на опережение, особенно если учесть, что по результатам исследования, проведенного Институтом управления проектами (PMI) в 2018 году, в 21% случаев проекты срываются из-за недостающих или ограниченных ресурсов.

Фаза 5. Завершение

Вот и все. Проект закончен.

Фаза завершения — это финал проекта, когда проще не подчищать хвосты, а оставить все как есть. Но последнее впечатление не менее важно, чем первое, и для завершения необходимо выполнить еще несколько действий.

Вместо того, чтобы считать проект «законченным», во время этой последней фазы стоит позаботиться о паре финальных задач:

- Сдача проекта клиенту или команде, которая будет работать с ним дальше
- Размещение всех документов и материалов в централизованном хранилище (чтобы обратиться к ним в будущем, если потребуется)
- «Разбор полетов», чтобы извлечь полезные уроки из успехов и неудач

Практическая работа №7. Часть 2.

Целеполагание. Формулировка целей.

На предыдущих этапах проектной работы цели проекта уже были сформированы. На данном этапе сформированные цели уточняются и окончательно прописываются в разрабатываемой проектной отчетности. Для целеполагания рекомендуется использовать технологию SMART.

Целеполагание — это процесс определения целей в жизни, бизнесе, творчестве и любой другой деятельности. Протекает осознанно или стихийно, бессистемно.

При осознанном целеполагании постановка целей осуществляется поэтапно и основывается на жизненной/деловой миссии, анализе потребностей и ресурсов.

При стихийном подходе цели рождаются спонтанно, без участия самого субъекта, под напором внешних обстоятельств. Например, человек хочет поступить в экономический ВУЗ, потому что настояли родители. Компания начинает выпуск продукции, потому что появилось дешёвое сырьё. Такая схема не приводит к высоким результатам, поскольку цели выбираются хаотично, а в их основе лежат не реальные потребности, а случайные обстоятельства.

Осознанное целеполагание базируется на системе истинных потребностей и ценностей. Оно предполагает:

- Осознание собственных потребностей – того, в чём человек нуждается больше всего: в любви, внимании, развитии, образовании, власти, и т.д.
- Определение ценностей, миссии компании или человека, которые раскрывают смысл существования.
- Выбор приоритетов – основных и второстепенных целей.
- Построение дерева целей – системы целей, распределённых во времени или по структурным подразделениям (в случае компании).

SMART — прекрасный инструмент. Один из пунктов SMART гласит, что любая цель должна иметь количественный показатель, то есть результат по ней должен быть оцифрован. Я убеждена, что каждая цель может быть оцифрована. Любой процесс, любое решение в компании влияет на ее финансовый результат в той или иной степени и соответственно его можно оценить. Это вопрос широты мышления, способности увидеть весь процесс целиком и практики.



Рис.8 Описание технологии SMART

SMART является набором критериев, по которым пишется формулировка цели. Аббревиатура SMART сформирована от 5 английских слов:

- S** **Specific** — конкретный.
- M** **Measurable** — измеримый.
- A** **Achievable** — достижимый.
- R** **Relevant** — соответствующий.
- T** **Time bounded** — ограниченный во времени.

Эффективность цели также повысится, если она будет про конкретный результат, а не процесс. Не работают цели с формулировками «принимать участие», «формировать», «обеспечивать». Данные глаголы не предусматривают конкретного завершения какого-либо действия и соответственно трудно измеримы. Уверена, что вы понимаете почему. Используйте глаголы совершенного вида прошедшего времени, то есть «выпустить», «создать», «обеспечить». Если вам необходимо поставить цель, касающуюся длительного проекта, но в рамках более короткого, чем предусмотрено проектом, периода, то рекомендую разбить эту цель на несколько промежуточных и сформулировать их по каждому этапу отдельно.

Целеполагание — это процесс определения целей в жизни, бизнесе, творчестве и любой другой деятельности. Протекает осознанно или стихийно, бессистемно.

При осознанном целеполагании постановка целей осуществляется поэтапно и основывается на жизненной/деловой миссии, анализе потребностей и ресурсов.

При стихийном подходе цели рождаются спонтанно, без участия самого субъекта, под напором внешних обстоятельств. Например, человек хочет поступить в экономический ВУЗ, потому что настояли родители. Компания начинает выпуск продукции, потому что появилось дешёвое сырьё. Такая схема не приводит к высоким результатам, поскольку цели

выбираются хаотично, а в их основе лежат не реальные потребности, а случайные обстоятельства.

Осознанное целеполагание базируется на системе истинных потребностей и ценностей. Оно предполагает:

- Осознание собственных потребностей – того, в чём человек нуждается больше всего: в любви, внимании, развитии, образовании, власти, и т.д.
- Определение ценностей, миссии компании или человека, которые раскрывают смысл существования.
- Выбор приоритетов – основных и второстепенных целей.
- Построение дерева целей – системы целей, распределенных во времени или по структурным подразделениям (в случае компании).

Отчетность по практической работе.

Описать все цели проекта по технологии SMART.

Практическая работа №8-9.

Календарное планирование и организация системы контроля проекта. Структурная декомпозиция работ.

Необходимо составить календарный план работ по проекту и сформировать систему контроля проекта. Для уточнения работ нужно провести структурную декомпозицию работ.

Отчетность по практической работе:

Календарный план работ по проекту и структурная декомпозиция работ. Студенческие команды представляют также в виде презентации разработанные материалы.

Структурная декомпозиция работ (СДР или WBS - Work Breakdown Structure) – это представление проекта в виде иерархической структуры работ, полученной путем последовательной декомпозиции. СДР предназначена для детального планирования, оценки стоимости и обеспечения персональной ответственности исполнителей.

Структурная декомпозиция работ (СДР) имеет следующие характеристики:

описывает с необходимой точностью содержание работ по проекту;

определяет весь объем работ по проекту;

формируется в виде иерархической структуры;

имеет измеримый результат, который рассматривается как результат совокупности работ.

СДР является средством для разделения всех работ по проекту на управляемые пакеты работ, позволяющие достичь такого уровня детализации информации, который соответствует потребностям руководства проекта для осуществления контроля.

Разработка СДР имеет две основные цели:

обеспечение планирования всех необходимых работ проекта,

обеспечение отсутствия работ, не связанных с реализацией проекта.

Для руководителя проекта важны обе эти цели. Если в плане отсутствуют все необходимые работы, проект будет задержан, бюджет, скорее всего, будет превышен. Если выполняются работы, не относящиеся к данному проекту – деньги заказчика тратятся нецелевым образом. Если СДР не объединяет обе эти цели, проект может потерпеть неудачу.

Проект разворачивается вокруг СДР. СДР является основным «стержнем» для четырех основных и одного вспомогательного процессов:

определения работ,

планирования ресурсов,

оценки стоимости,

бюджетирования,

определения рисков.

Структурная декомпозиция работ может разрабатываться «с нуля» либо с использованием компонентов уже созданных СДР.

Прежде всего, нужно ответить на целый ряд вопросов:

что нужно сделать (определить продукты проекта);

как это нужно будет делать (определить технологические этапы проекта);

кто это будет делать (определить исполнителей, соисполнителей, субподрядчиков);

кто и в какой форме будет оплачивать работы (определить, какие и с кем будут заключены контракты).

Правила разработки СДР

При разработке СДР необходимо принимать во внимание следующие основные правила:

каждый элемент СДР должен обеспечивать достижение измеримого результата;

каждый элемент СДР должен являться агрегатом всех подчиненных элементов, перечисленных непосредственно под ним;

результаты должны логически декомпозироваться до уровня, на котором можно определить, как они будут достигаться (проектирование, поставки, заключение договоров, производство). Результаты проекта от верхнего до нижнего уровня декомпозиции должны быть логически связаны;

результаты пакетов работ должны быть уникальными и отличаться от результатов других пакетов работ того же уровня;

пакеты работ должны декомпозироваться до уровня детализации, обеспечивающей успешное планирование, координацию и контроль работ, связанных с достижением поставленных целей;

при изменении содержания проекта СДР должна быть откорректирована;

для всех важных событий, связанных с отчетностью (например, ежемесячные отчеты, отчеты о проведении испытаний и т.д.) должны быть определены соответствующие пакеты работ;

все пакеты работ должны быть совместимы с организационной структурой и структурой затрат;

результаты должны быть четко определены так, чтобы исключить дублирование объемов работ внутри элементов СДР, в целом по организации или отдельными ответственными за выполнение работ;

результаты должны иметь размер, достаточный для эффективного управления, но не настолько малый, чтобы сделать затраты на контроль чрезмерными.

Этапы разработки СДР

СДР разрабатывается путем итерационного рассмотрения целей и результатов проекта, критериев планирования объема работ, реализации технических требований. Верхние уровни СДР могут быть разработаны на ранней, концептуальной стадии проекта.

Дальнейшая детализация СДР возможна, как только будет определен проект и подготовлены спецификации.

Очень важно понять, что первоочередная задача составления СДР – разделить проект на подпроекты до той степени детализации, когда появится возможность распределить элементарные работы.

Основной процесс разработки СДР состоит из следующих шагов:

Определение конечных результатов проекта – что должно быть произведено для обеспечения успешного завершения проекта. В качестве руководства рекомендуется проанализировать, рассмотреть документы, описывающие общий объем работ по проекту.

Определение основных пакетов работ, необходимых для получения продукта проекта. Часто такими основными пакетами работ являются результаты, необходимые для создания продукта проекта, но вместе с тем, сами по себе они не являются целями проекта (например, технические требования к разработке ИС).

Приведение в соответствие элементов дополнительных уровней детализации и внутренней системы управления и контроля. Такие элементы обычно связаны с четким и раздельным определением отдельных результатов (продуктов) проекта.

Пересмотр (анализ) и усовершенствование СДР до тех пор, пока все участники проекта не будут согласны, что планирование проекта может быть успешно завершено, и можно будет успешно управлять, контролировать и регулировать получаемые результаты.

Подготовку структурной декомпозиции работ можно считать законченной, когда определены мелкие индивидуальные (элементарные) работы. Ответственность за каждую элементарную работу должна быть поручена одному и только одному члену команды проекта. Если этот человек (или группа) собираются выполнять работу, а не руководить ее выполнением, этот уровень может быть признан самым нижним уровнем СДР.

Календарное планирование в управлении проектами – это ключевой и важный процесс, результатом которого является утвержденный руководством компании календарный план проекта (часто его называют еще **планом-графиком**, календарным графиком, планом управления проектом). Цель календарного планирования – получить точное и полное расписание проекта с учетом работ, их длительностей, необходимых ресурсов, которое служит основой для исполнения проекта.

Календарное планирование включает в себя:

планирование содержания (scope) проекта и построение СДР - структурной декомпозиции работ, или WBS (Work Breakdown Structure);

определение последовательности работ и построение сетевого графика;

планирование сроков, длительностей и логических связей работ и построение диаграммы Ганта;

определение потребности в ресурсах (люди, машины и механизмы, материалы и т.д.) и составление ресурсного плана проекта;

расчет затрат и трудозатрат по проекту.

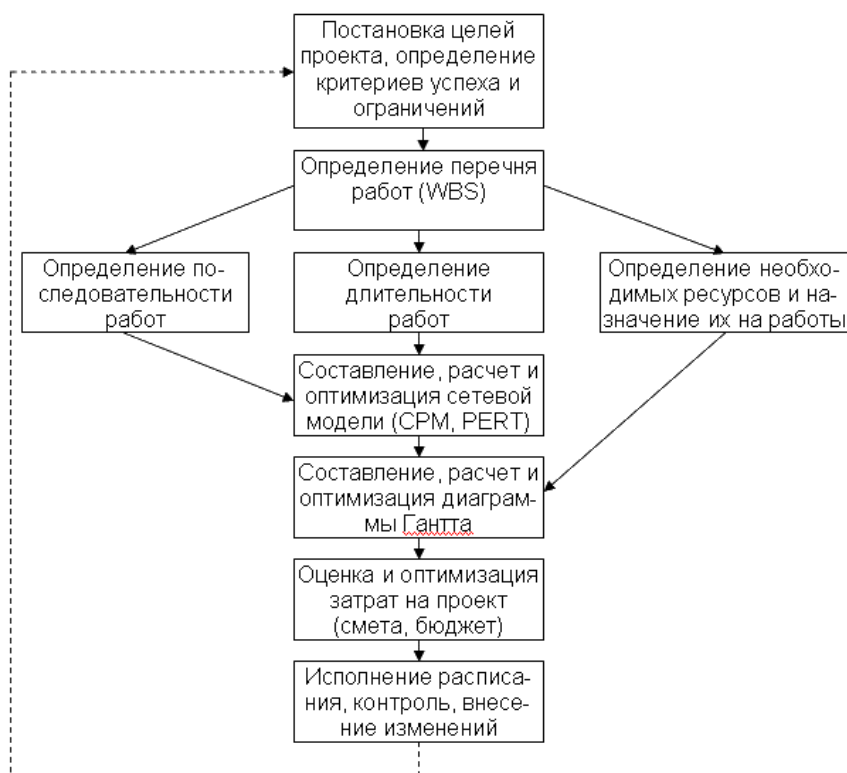


Рис.10. Алгоритм планирования проекта

Составление календарного плана-графика проекта включает в себя несколько аспектов. Мы должны спланировать сроки и длительности работ, определить их последовательность и взаимосвязи, подумать о необходимых ресурсах, учесть стоимость этих работ и ресурсов. В дальнейшем, когда проект перейдет на стадию исполнения, то есть практической реализации запланированных действий, именно по этому плану-графику мы отслеживаем ход выполнения работ. И, если что-то в проекте пойдет не так, можно, сверив с первоначальным планом проекта, внести соответствующие изменения.

Как правило, план-график проекта разрабатывается менеджером проекта с привлечением людей, которые являются экспертами в той или иной области. Например, содержание строительных работ лучше всего знает специалист по строительству; а мероприятия по продвижению продукта, скорее всего, спланирует маркетолог. В результате мы получаем полный перечень работ, структурированный по иерархическому признаку, то есть структурную декомпозицию работ (СДР).

Следующий шаг по созданию календарного плана проекта – это определение длительностей работ и их взаимосвязей. Например, какие-то работы в нашем списке могут выполняться строго последовательно, а какие-то – параллельно друг с другом во времени. Для того чтобы «увязать» сроки работ по проекту, их продолжительность и зависимости, сегодня во всем мире менеджеры проектов используют простой и вместе с тем полезный инструмент календарного планирования – диаграмму Ганта (иногда пишется «диаграмма Ганта»). Диаграмма Ганта – это наглядное представление календарного плана-графика проекта, в котором слева расположен иерархический перечень всех работ проекта (СДР), и справа – календарь с конкретными датами. Работы обозначены полосками, связи между работами - стрелками.

Практическая работа №10.

Управление рисками проекта. Мониторинг и контроль рисков.

В практической работе студенческие команды должны определить риски своего проекта, а также сформулировать мероприятия по мониторингу рисков.

Отчетность по практической работе:

Описание возможных рисков проекта и плана мониторинга рисков, доклад студенческой команды на практическом занятии.

Основными задачами мониторинга и управления рисками являются отслеживание идентифицированных рисков, поиск и получение информации для выявления новых рисков и планирование управления ими, обеспечение своевременной реализации плана мероприятий по управлению рисками, а также оценка их эффективности.

Получаемая в процессе мониторинга информация используется для добавления в реестр рисков не выявленных ранее угроз или возможностей, а также для повторной качественной и количественной оценки рисков и корректировки плана мероприятий по управлению рисками. Пересмотр рисков основывается на сопоставлении получаемой информации о рисках с установленными на этапе планирования и идентификации индикаторами и триггерами риска. По результатам мониторинга принимаются решения о реализации соответствующих мероприятий по управлению.

В рамках мониторинга осуществляется аудит рисков, цель которого — сбор и документирование информации о реализовавшихся рисках, принятых мерах по управлению рисками, а также проводится общая оценка эффективности управления рисками проекта. Результаты аудита рисков используются для совершенствования и повышения эффективности системы управления рисками проекта, а отчетные документы передаются в архив.

Процесс мониторинга и управления рисками осуществляется непрерывно на протяжении всего жизненного цикла проекта и тесно взаимосвязан с процессом коммуникаций и консультаций в области управления рисками. На начальных этапах проекта коммуникации со всеми участниками проекта необходимы для определения внешних и внутренних условий, а также четкой постановки целей и критериев успешности управления рисками проекта. На последующих стадиях процесса управления рисками проекта, от идентификации до планирования мероприятий по управлению рисками, коммуникации необходимы для получения информации о рисках и их анализа, для определения доступных альтернатив и инструментов управления рисками и формирования плана мероприятий по управлению рисками, а также для согласованного определения ответственных за их выполнение. Непосредственно в процессе мониторинга коммуникации направлены не только на получение новой информации, но, возможно, в первую очередь на осведомление основных участников проекта о существенных событиях, произошедших в проекте, о реализовавшихся и потенциальных рисках и их влиянии на достижение целей проекта. Открытые и регулярные коммуникации позволяют не только обеспечить прозрачность управления рисками проекта, но и повысить осведомленность и вовлеченность участников в данный процесс.

Практическая работа №11.

Управление персоналом в проекте. Подбор экспертов для формирования баз знаний.

В данной работе проводится детальный анализ персонала, имеющегося в проекте, в числе студенческой команды, а также функционал привлекаемого персонала. Описываются основные требования к специалисту-эксперту, используемого для формирования базы знаний проекта в сфере ИИ.

Форма отчетности:

Детальное описание функционала персонала проекта, способы руководства персоналом.

Управление персоналом проекта включает организационное планирование, кадровое обеспечение проекта, создание команды проекта, а также реализует функции контроля и мотивации трудовых ресурсов проекта для эффективного выполнения работ и успешного завершения проекта. Целью при этом являются руководство и координация деятельности команды проекта. Для достижения цели используются различные стили руководства, разнообразные административные методы и методы мотивации, повышение квалификации кадров на всех этапах жизненного цикла проекта.

Иногда в качестве синонима к термину «управление персоналом» приравнивают и более узкое, но в рамках управления проектами наиболее важное понятие — «управление командой проекта». Следует оговориться, что управление командой проекта хотя и является основной частью управления персоналом в рамках управления проектами, но далеко не исчерпывает основную его специфику. Управление персоналом проекта помимо управления командой также еще включает и проблемы, свойственные управлению персоналом в рамках общего менеджмента.

Управление персоналом проекта включает процессы, необходимые для того, чтобы сделать использование персонала, вовлеченного в проект, наиболее эффективным.

Организационное планирование (Organizational Planning) — определение, документирование и распределение ролей в проекте, ответственности и отчетности.

Подбор персонала (Staff Acquisition) — подбор персонала для работ над проектом.

Развитие команды (Team Development) — развитие индивидуальных и групповых навыков для улучшения качества работы над проектом.

Основные проблемы, свойственные управлению персоналом любого проекта:

- 1) управление командой проекта (образование команды проекта, ее развитие, проблемы расформирования команды);
- 2) разрешение конфликтов, возникающих в связи с использованием проектно-матричных организационных структур управления проектами;
- 3) проблемы общего управления, связанные со взаимодействием участников проекта с другими членами организаций.

При использовании проектно-матричных организационных структур управления проектами возникает ставшая уже классической проблема управления персоналом, связанная с двойным подчинением сотрудников функциональных подразделений, участвующих в проекте. Так как двойное подчинение изначально закладывается в механизм организационной структуры, то управление конфликтами — одна из главных задач управления человеческими ресурсами проекта.

Одним из основных отличий управления персоналом в рамках управления проектом является использование особого образования — команды.

Команда проекта — одно из главных понятий управления проектами. Это группа сотрудников, непосредственно работающих над осуществлением проекта и подчиненных руководителю проекта; основной элемент структуры проекта, так как именно команда проекта обеспечивает реализацию замысла проекта.

Команду характеризует размытость, нечеткость структуры и формального разделения полномочий. Команда — не простое структурное подразделение, деятельность которого, как правило, формируется по функциональному признаку. Деятельность команды является целевой, она нацелена на конкретный результат (result-driven), а не ориентирована на выполнение некоторой деятельности (activity-oriented), будь то функция или процесс. Это существенным образом повышает эффективность использования человеческих ресурсов, но ввиду нетрадиционности вызывает определенные трудности в формировании команды и управлении ею.

Основные проблемы, возникающие при управлении командой проекта:

- как быстро и эффективно сформировать команду из некогда незнакомых людей с разными характерами, с разными уровнями знаний и опыта, а главное — с разными областями профессиональной деятельности;
- каким образом быстро наладить между ними эффективное взаимодействие и ориентировать их деятельность на достижение общего результата;
- как обеспечить эффективность управления командой и ее взаимодействие с другими участниками проекта;
- как управлять командой в условиях, когда достижение цели проекта означает роспуск команды (какие средства мотивации применять, как управлять судьбой и карьерой членов команды).

3. Вовлеченность

Это минимум два человека, которые занимаются проектом фултайм. Варианты с постоянной занятостью в другой компании и работой над проектом по вечерам не рассматриваются. Бывает, что в Акселератор проходят стартапы с одним человеком на фултайм, но это скорее исключение. Двум членам команды проще распределить задачи, поэтому вовлеченность определяет скорость

4. Мотивация

Для нас идеальный кейс — когда двум-трем участникам, костяку команды принадлежит если не 100%, то хотя бы 70-80%. Это гарантия того, что команда будет заниматься проектом долгосрочно. Как правило, руководитель владеет более

50%. Сомнителен кейс, когда 100% принадлежит руководителю, а команда при этом никаким образом не замотивирована. В этом случае мы спрашиваем, как основатель планирует в ближайшей перспективе мотивировать сотрудников: выделять опционы, перераспределять часть своей доли сотрудникам, делать их соучредителями. Бывают кейсы, когда внутри команды есть договоренность о распределении долей, которая не закреплена юридически. Это понятно членам команды, но для инвестора надежнее, когда у участников есть реальные доли. Опцион — довольно индивидуальная вещь. На самой ранней стадии, когда есть команда, базовый продукт и только начинаются продажи, идеально, если у всех ключевых членов команды есть доли. Если же уже есть объем продаж, проект какое-то время функционирует — могут быть КРІ и завязанные на них функционалы. Обычно это все же КРІ, который определяет основатель проекта.

-

Практическая работа №13-14.

Мотивация участников проекта. Распределение ролей в команде.

В данной работе происходит дальнейшее развитие способов управления персоналом для достижения результатов проекта и окончательно определяются роли в проектной команде.

Форма отчетности по практической работе:

Отчет по распределению ролей в команде и способов мотивации участников проекта.

1. Опыт лидера проекта

У студенческого руководителя проекта должны быть, как минимум, лидерские качества, желателен опыт работы в индустрии, в которой он разрабатывается проект — экспертиза, а также опыт ведения собственных проектов.

2. Компетенции команды и баланс в распределении ролей

Идеальное распределение ролей — когда лидер проекта параллельно обладает какой-то второй компетенцией (маркетолог, продажи, разработка, веб-продвижение). Второй человек дополняет его компетенцию. В идеале — должны быть закрыты техническая часть и маркетинговая. В совокупности у них должна быть синергия, это позволит им быстрее развивать свой проект и масштабироваться. От технической части в ИТ-проекте команде должен быть тим-лид, разработка же может оставаться на аутсорсе.

Практическая работа №15.

Управление коммуникациями в проекте. Распределение проектной информации, представление отчетности. Разработка плана управления коммуникациями проекта

В данной работе необходимо разработать план управления коммуникациями в проекте.

Форма отчетности по практической работе:

план управления коммуникациями в проекте, представляемый студенческой командой.

Управление информацией и коммуникациями проекта (Project Communications Management)

Обеспечение участников и процессов проекта информацией включает каналы связи, накопление данных, обмен и актуализацию данных, ведение баз данных, распределение информации по потребителям. Управление информацией обеспечивает предоставление, оценку, переработку, мониторинг, анализ информации, информационных потоков в течение жизненного цикла проекта.

Коммуникации и сопутствующая им информация — своего рода фундамент для координации действий участников проекта.

Под информацией понимают собранные, обработанные и распределенные данные. Чтобы быть полезной для принятия решений, информация должна предоставляться своевременно, по назначению и в удобной форме. Это решается использованием современных информационных технологий в рамках системы управления проектом.

Управление коммуникациями проекта (управление взаимодействием, информационными связями) — управленческая функция, направленная на обеспечение своевременного сбора, генерации, распределения и хранения необходимой проектной информации.

Управление коммуникациями проекта включает процессы:

Планирование коммуникаций (Communications Planning) — определяет информационные и коммуникационные нужды команды проекта (кому, когда и какая необходима информация).

Распределение информации (Information Distribution) — своевременное предоставление необходимой информации участникам проекта.

Отчетность об исполнении (Performance Reporting) — сбор и распространение информации о ходе выполнения проекта.

Административное завершение (Administrative Closure) — подготовка, сбор и распространение информации и материалов для официального завершения фазы или проекта.

Основными потребителями информации проекта выступают:

Менеджер проекта — для анализа расхождений фактических показателей выполнения работ от запланированных и принятия решений по проекту.

Заказчик — для осведомленности о ходе выполнения проектных работ.

Поставщики — при возникновении потребности в материалах, оборудовании и т.п., необходимых для выполнения работ.

Проектировщики — при необходимости внесения изменений в проектную документацию.

Непосредственные исполнители работ.

Содержание управления коммуникациями проекта. Управление коммуникациями проекта обеспечивает поддержку системы связи (взаимодействий) между участниками проекта, передачу управленческой и отчетной информации, направленной на обеспечение достижения целей проекта. Каждый участник проекта должен быть подготовлен к взаимодействию в рамках проекта в соответствии с его функциональными обязанностями.

Управление коммуникациями включает следующие процессы.

Планирование системы коммуникаций — определение информационных потребностей участников проекта (определение состава информации, сроков и способов ее доставки).

Сбор и распространение информации — процессы регулярного сбора и своевременной доставки необходимой информации участникам проекта.

Подготовка отчетов о ходе выполнения проекта — обработка фактических результатов состояния работ проекта, сравнение их с плановыми показателями, анализ тенденций, прогнозирование.

Документирование хода работ — сбор, обработка и организация хранения документации по проекту.

Рассмотрим перечисленные процессы подробнее.

Планирование системы коммуникаций. План коммуникаций — составная часть плана проекта. Он включает:

план сбора информации, в котором определяются источники информации и методы ее получения;

план распределения информации, в котором определяются потребители информации и способы ее доставки;

детальное описание каждого документа, который должен быть получен или передан, включая формат, содержание, уровень детальности и используемые определения;

план ввода в действие тех или иных видов коммуникаций;

методы обновления и совершенствования плана коммуникаций.

План коммуникаций формализуется и детализируется в зависимости от потребностей проекта.

Сбор и распространение информации. В рамках проекта существует потребность в осуществлении различных видов коммуникаций:

- • внутренних (внутри команды проекта) и внешних (с руководством компании, заказчиком, внешними организациями и т. д.);
- • формальных (отчеты, запросы, совещания) и неформальных (напоминания, обсуждения);
- • письменных и устных;
- • вертикальных и горизонтальных.

Отчетность о ходе выполнения проекта. Процессы сбора и обработки данных о фактических результатах и отображение информации о состоянии работ в отчетах обеспечивают основу для координации работ, оперативного планирования и управления. Отчетность о ходе выполнения работ включает:

- • информацию о текущем состоянии проекта в целом и в разрезе отдельных показателей;
- • информацию об отклонениях от базовых планов;
- • прогнозирование будущего состояния проекта.

Системы сбора и распределения информации должны обеспечивать потребности различных видов коммуникаций. Для этих целей могут использоваться автоматизированные и неавтоматизированные методы сбора, обработки и передачи информации.

Неавтоматизированные методы включают сбор и передачу данных на бумажных носителях, проведение совещаний.

Автоматизированные методы предусматривают использование компьютерных технологий и современных средств связи для повышения эффективности взаимодействия: электронную почту, системы документооборота и архивирования данных.

Документирование хода работ. Основные промежуточные результаты хода работ должны быть формально задокументированы. Документирование результатов хода работ включает:

- • сбор и верификацию окончательных данных;
- • анализ и выводы о степени достижения результатов проекта и эффективности выполненных работ;
- • архивирование результатов с целью дальнейшего использования.

Компьютерные системы ведения электронных архивов позволяют автоматизировать процессы хранения и индексации текстовых и графических документов, значительно облегчить доступ к архивной информации.

Практическая работа №16.

Информационное обеспечение управления проектами: состав, структура, характеристики. Программные средства для управления проектами. Характеристика состояния рынка программных продуктов по управлению проектами. Планирование проекта с использованием MS Project.

В данной работе используется программное обеспечение для управления проектами MS Project.

Форма отчетности:

План выполнения проекта, созданный в MS Project. Диаграмма Ганта.

Процесс планирования в MS Project Online – это процесс, в рамках которого проводится оценка ключевых параметров проекта, формируется перечень задач и оцениваются объемы финансирования и необходимых ресурсов. Этот процесс включает в себя предварительную оценку всех элементов проекта. В результате реализации данного этапа формируется план-график проекта. Если данный план-график устраивает основные заинтересованные стороны проекта, сохраняется базовый план проекта и начинается его реализация.

Удобная форма представления плана работ – диаграмма Ганта.

Диаграмма Ганта — это тип столбчатых диаграмм, который используется для иллюстрации плана, графика работ по какому-либо проекту. Является одним из методов планирования проектов. Используется в приложениях по управлению проектами. Первый формат диаграммы был разработан Генри Л. Ганттом в 1910 году.

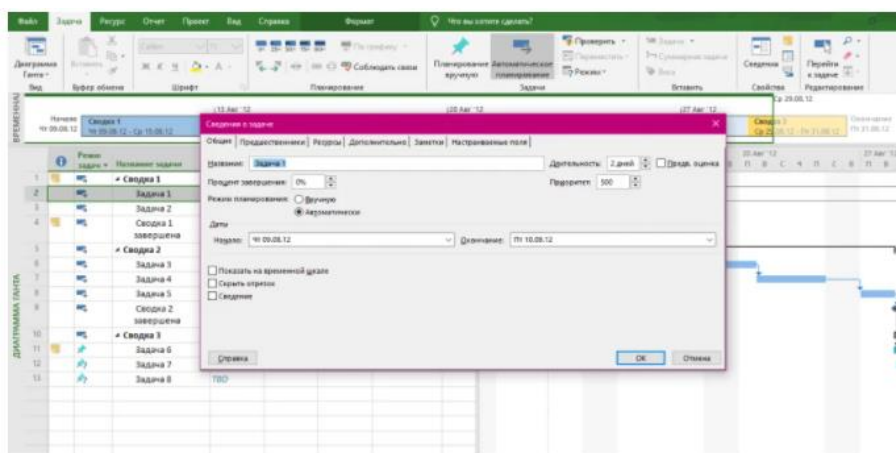


Рис.11 Формирование проектных заданий

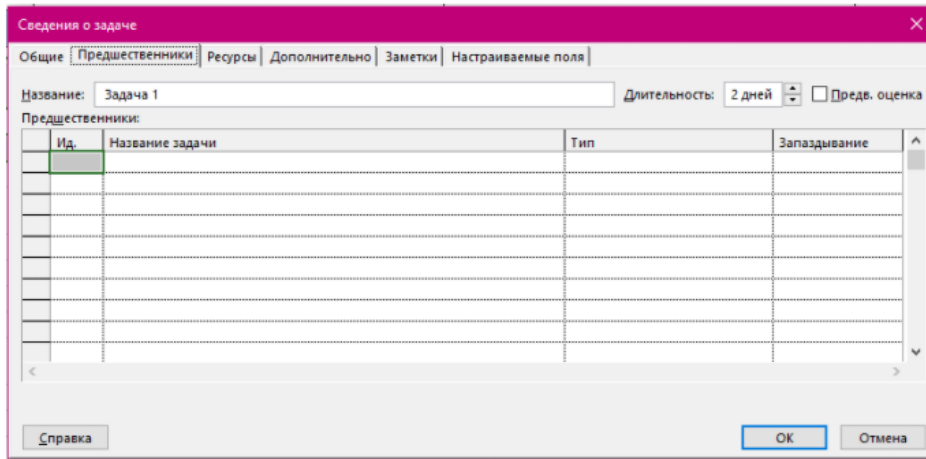


Рис.12 Формирование задач и указание предшествующих задач

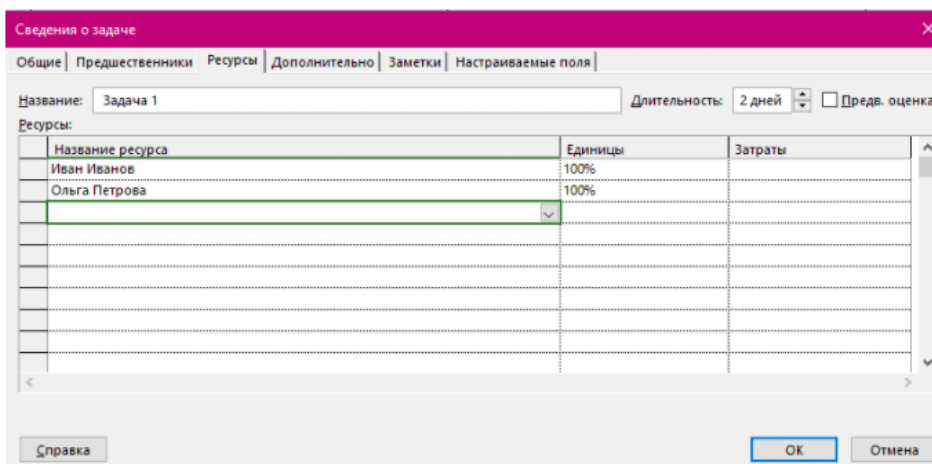


Рис.13 Определение ресурсов при выполнении проекта

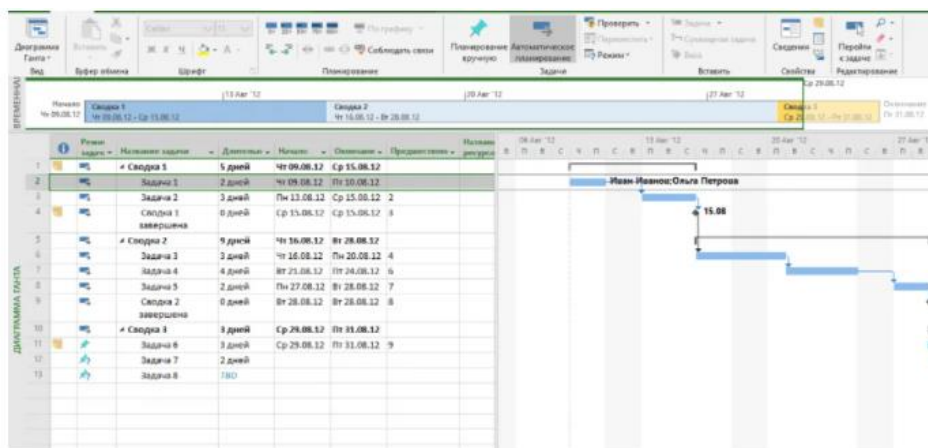


Рис.14 Пример формирования диаграммы Ганта

Список используемой литературы:

- Баранчеев, В. П. Управление инновациями в 2 т : учебник для академического бакалавриата [Текст] / В. П. Баранчеев, Н. П. Масленникова, В. М. Мишин. – 3-е изд., перераб. и доп. – М. : Юрайт, 2015. – 782 с.
- Зуб, А. Т. Управление проектами : учебник и практикум для академического бакалавриата [Текст] / А. Т. Зуб. : МГУ им. М.В. Ломоносова. – М. : Юрайт, 2017. – 422 с.
- Первушин, В.А. Практика управления инновационными проектами : учебное пособие [Текст] / В. А. Первушин ; РАНХиГС – М. : Дело, 2015. – 208 с.
- Поляков, Н. А. Управление инновационными проектами : учебник и практикум для академического бакалавриата [Текст] / Н. А. Поляков, О. В. Мотовилов, Н. В. Лукашов. — М. : Юрайт, 2017. – 330 с.
- Первушин, В.А.. Практика управления инновационными проектами : [учеб. пособие] / В. А. Первушин; – М.: ИД «Дело» РАНХиГС, 2013. – 208 с.
- Попов, В.Л. Управление инновационными проектами : учебное пособие [Текст] / В. Л. Попов и др. ; под ред. В. Л. Попова. – М.: Инфра-М, 2015. – 336.
- Туккель, И.Л., Сурина, А.В., Культин, Н.Б. Управление инновационными проектами: учеб. для студентов вузов [Текст] / И.Л. Туккель, А.В. Сурина, Н.Б. Культин; под общ. ред. И. Л. Туккеля – СПб. : БХВ-Петербург, 2011. – 416 с.

Дополнительная литература:

- Алексеева, М. Б. Анализ инновационной деятельности : учебник и практикум для бакалавриата и магистратуры [Текст] / М. Б. Алексеева, П. П. Ветренко. — М. : Издательство Юрайт, 2017. – 303 с.
- Гончаренко, Л. П. Инновационный менеджмент : учебник для академического бакалавриата [Текст] / Л. П. Гончаренко, Б. Т. Кузнецов, Т. С. Булышева, В. М. Захарова ; под общ. ред. Л. П. Гончаренко. — 2-е изд., перераб. и доп. – М. : Юрайт, 2016. – 487 с.
- Друкер, П.Ф. Менеджмент. Вызовы XXI века [Текст] / П.Ф. Друкер ; пер. с англ. Н. Макарова. – М.: Манн, Иванов и Фербер, 2012. – 256 с.
- Кремер, Н. Ш. Математическая статистика : учебник и практикум для академического бакалавриата [Текст] / Н. Ш. Кремер. – М. : Юрайт, 2017. – 259 с.
- Тарасенко, Ф.П. Прикладной системный анализ. Учебное пособие [Текст] / Ф.П. Тарасенко. – М.: КноРус, 2010. – 224 с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Методы анализа данных в бизнес-аналитике

Профиль подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

ЛЕКЦИИ

Лекция 1. Введение в бизнес-аналитику. Предварительный анализ данных

Такая разная аналитика

Бизнес-аналитика. Бизнес-аналитика — процесс анализа данных, позволяющий руководителям, менеджерам и другим заинтересованным сторонам принимать обоснованные бизнес-решения.

- Бизнес-аналитик
- Продуктовый аналитик
- Маркетинговый аналитик
- Системный аналитик
- UX-аналитик
- Веб-аналитик
- **Аналитик данных**
 - BI-аналитик
 - Аналитик Big Data (Data Scientist)

Данные собирают все — от магазинов и ресторанов до компаний-монополистов и приложений с миллионной аудиторией. Аналитик данных помогает сделать так, чтобы собранная информация приносила пользу бизнесу.

Аналитик данных (или дата-аналитик) — это специалист, который собирает, обрабатывает, изучает и интерпретирует данные. Его работа помогает принимать решения в бизнесе, управлении и науке.

Хороший аналитик данных — не просто математик с навыками программиста. Он понимает бизнес-процессы и хорошо знает продукт. Такой специалист разбирается, на чем зарабатывает конкретный бизнес. В результате его работы компания может получать больше прибыли и делать своих пользователей счастливее. Сильный аналитик данных прежде, чем взяться за работу всегда спрашивает руководителя о том, какую задачу хочет решить бизнес.

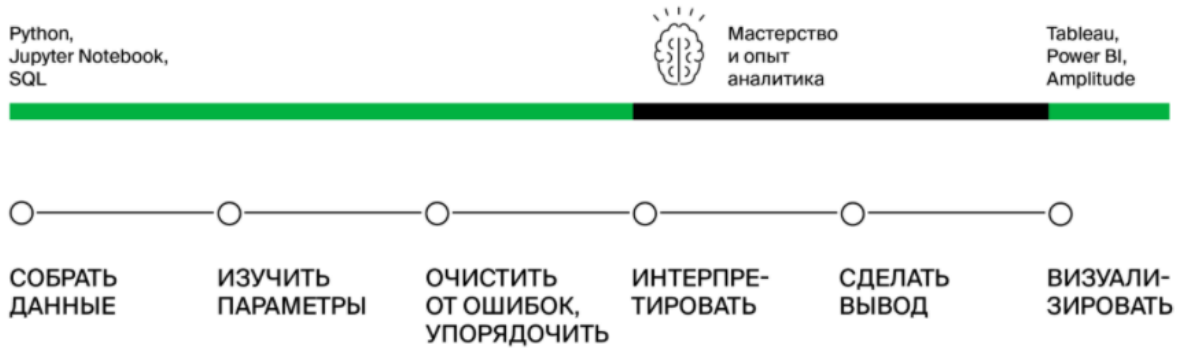
Аналитики данных берут данные, предоставленные инженерами данных, анализируют их и дают рекомендации. Они создают визуализации для отображения своих результатов в информационных панелях и презентациях.

В отличие от исследователей данных, аналитики данных обычно не создают прогностические модели, основанные на алгоритмах машинного обучения.

Знания и навыки

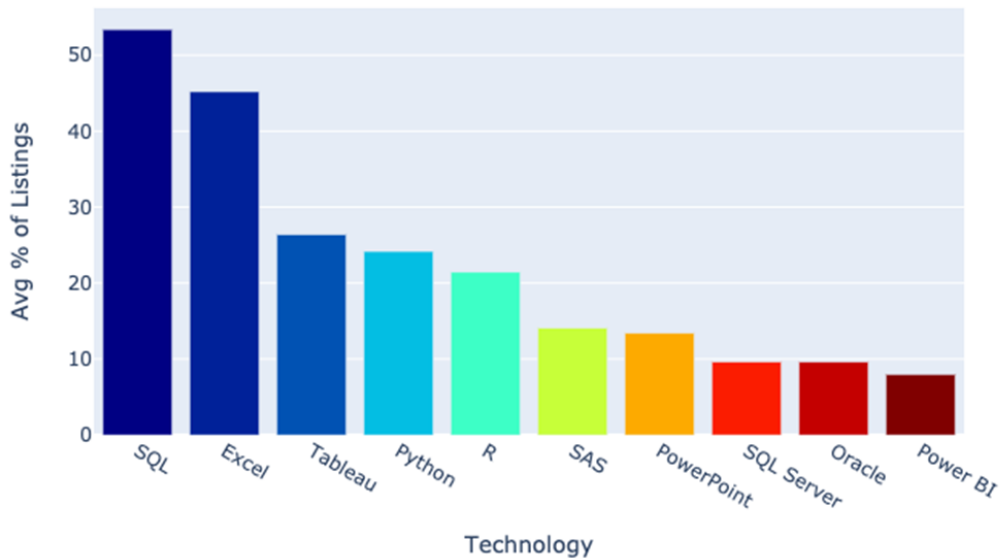


Стадии работы и инструменты



10 технологий из списков вакансий аналитика данных по состоянию на январь 2020 года

Technologies in Data Analyst Job Listings 2020



Аналитик данных, или Data Analyst, становится одной из самых востребованных ИТ-специальностей. По данным исследования рынка аналитиков, спрос на профессионалов значительно превышает предложение.

Сегодня создана целая наука, посвященная обработке данных, которая так и называется data science - наука о данных. Однако некоторые эксперты полагают, что это определение ошибочно, потому что data science – не "наука о данных", как написано в русскоязычной Википедии. Данные не являются предметом этой науки, поэтому называть data science синонимом предложенной Петером Науром науки datalogy ошибочно. Термин data science на русский язык, возможно, стоило бы переводить как "наука работы с данными" или "**научные методы работы с данными**". Задача, решаемая теми, кто занимается data science, состоит в извлечении знаний с использованием методов, объединенных под общим названием data mining, в объединении статистики и других методов анализа данных с целью понимания того, что содержат в себе данные.

Четвертая парадигма науки

Обладатель премии Тьюринга Джим Грей и астроном и футуролог Алекс Шалаи разделили научное прошлое человечества на три периода использования данных и дополнили его современным четвертым.

1. **Античные времена** — описание наблюдаемых феноменов и логические выводы, сделанные на основе наблюдений.
2. **XVII век** — создание теорий с использованием для доказательства их истинности аналитических моделей.
3. **XX век** — использование методов численного моделирования, ставшее возможным благодаря появлению компьютеров.
4. **XXI век** — **использование методов, основанных на анализе данных**; применение для работы с огромными объемами данных статистических и других методов извлечения полезной информации.

Анализ данных

Анализ данных — область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных; процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решений. **Анализ данных имеет множество аспектов и подходов, охватывает разные методы в различных областях науки и деятельности.**

Business Intelligence и бизнес-аналитика

Business intelligence (BI) — обозначение компьютерных методов и инструментов для организаций, обеспечивающих перевод транзакционной деловой информации в человекочитаемую форму, а также средства для массовой работы с такой обработанной информацией.

Цель BI — интерпретировать большое количество данных, заостряя внимание лишь на ключевых факторах эффективности, моделируя исход различных вариантов действий, отслеживая результаты принятия решений.

Традиционно BI-решения используются для статических отчетов о текущем или прошлом состоянии бизнеса. Они отвечают на такие вопросы, как: «Какая динамика объема продаж в прошедшем квартале? За счет чего произошел рост или падение продаж? Какой тип продукции произвели больше всего за месяц?». Это так называемый дескриптивный или описательный анализ. Кроме того, BI системы работают со структурированными данными, извлеченными из хранилищ данных, и представляют результат анализа в виде интерактивных информационных панелей — дашбордов или отчетов.

- Термины BI и «бизнес-аналитика» зачастую используются как синонимы, но между ними есть разница. Бизнес-аналитика (в узком понимании), в отличие от BI, имеет дело с уже очищенными, подготовленными для анализа данными, использует статистические и количественные инструменты для оценки текущей ситуации и прогнозирования, поэтому ее все чаще называют «углубленная аналитика».
- Business Intelligence, BI вначале занимается очисткой, консолидацией данных, преобразованием их в удобный для анализа формат, следующие задачи — интерпретировать большое количество данных, заостряя внимание лишь на ключевых факторах, влияющих на эффективность, моделировать исход различных вариантов действий, отслеживать результаты принятия решений. **Основное назначение BI — это именно принятие решений для бизнеса.**
- Основу методов data mining составляют всевозможные методы классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования, ассоциативной памяти, нечёткой логики. К методам data mining нередко относят статистические методы (дескриптивный анализ, **корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов**, анализ выживаемости, анализ связей). Такие методы, однако, предполагают некоторые априорные представления об анализируемых данных, что несколько расходится с целями data mining (обнаружение ранее неизвестных нетривиальных и практически полезных знаний).

BI-системы развиваются по четырем основным направлениям:

1. **Хранение данных.** Данные в хранилище BI-системы (data warehouse, DW) структурируются специальным образом для более эффективного анализа и обработки запросов (в отличие от обычных баз данных, где информация организована таким образом, чтобы оптимизировать время обработки текущих транзакций).

2. **Интеграция данных.** Для формирования и поддержания хранилищ данных используются ETL-средства — инструменты, обеспечивающие извлечение данных (extract), их преобразование (transform), т.е. приведение к необходимому формату, и загрузку (load) данных в хранилище или в другую базу.
3. **Анализ данных.** Для всестороннего анализа данных используются OLAP-инструменты (on-line analytical processing). Они позволяют рассматривать различные срезы данных, выявлять тренды и зависимости (по регионам, продуктам, клиентам и т.п.).
4. **Представление данных.** Для представления данных используются различные графические средства — отчеты, графики, диаграммы. Общепринятым средством визуализации данных являются информационные панели (dashboards), на которых результаты отображаются в виде индикаторов и шкал, позволяющих контролировать текущие значения выбранных показателей, сравнивать их с минимально/максимально допустимыми и таким образом выявлять потенциальные угрозы для бизнеса.

Интеллектуальный анализ данных

Data mining (рус. добыча данных, интеллектуальный анализ данных, глубокий анализ данных) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Английское словосочетание «*data mining*» пока не имеет устоявшегося перевода на русский язык. Более полным и точным является словосочетание «*обнаружение знаний в базах данных*».

Постановка задачи

Первоначально задача ставится следующим образом:

- имеется достаточно крупная база данных;
- предполагается, что в базе данных находятся некие «скрытые знания».

Необходимо разработать методы обнаружения знаний, скрытых в больших объемах исходных «сырых» данных. В текущих условиях глобальной конкуренции именно найденные закономерности (знания) могут быть источником дополнительного конкурентного преимущества.

Что означает «скрытые знания»? Это должны быть обязательно знания:

- ранее неизвестные — то есть такие знания, которые должны быть новыми (а не подтверждающими какие-то ранее полученные сведения);
- нетривиальные — то есть такие, которые нельзя просто так увидеть (при непосредственном визуальном анализе данных или при вычислении простых статистических характеристик);

- практически полезные — то есть такие знания, которые представляют ценность для исследователя или потребителя;
- доступные для интерпретации — то есть такие знания, которые легко представить в наглядной для пользователя форме и легко объяснить в терминах предметной области.

Data mining и искусственный интеллект

Знания, добываемые методами data mining, принято представлять в виде закономерностей (паттернов). В качестве таких выступают:

- *ассоциативные правила;*
- *деревья решений;*
- *кластеры;*
- *математические функции.*

Алгоритмы поиска таких закономерностей находятся на пересечении областей: Искусственный интеллект, Математическая статистика, Математическое программирование, Визуализация, OLAP.

Прикладная статистика

Прикладная статистика — наука о методах обработки статистических данных. Методы прикладной статистики активно применяются в технических исследованиях, экономике, менеджменте, социологии, медицине, геологии, истории и т. д.

Прикладная статистика нацелена на решение реальных задач. Поэтому в ней возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими методами, то есть путём доказательства теорем.

Машинное обучение

Глубинная суть машинного обучения — предсказание: предсказание наших желаний, результатов наших действий, путей достижения целей, изменений мира.

Машинное обучение принимает много разных форм и скрывается под разными именами: **распознавание паттернов, статистическое моделирование, добыча данных, выявление знаний, предсказательная аналитика, наука о данных, адаптивные и самоорганизующиеся системы** и так далее. Все они находят свое применение и имеют разные ассоциации.

Единственным способом заставить компьютер что-то делать — от сложения двух чисел до управления самолетом — было составление некоего алгоритма, скрупулезно объясняющего машине, что именно от нее требуется. Однако алгоритмы машинного обучения — совсем другое дело: они угадывают все сами, делая выводы

на основе данных, и чем больше данных, тем лучше у них получается. Это значит, что **компьютеры не надо программировать: они программируют себя сами**. Машинное обучение — технология, которая строит саму себя. Это новое явление в нашем мире.

Машинное обучение иногда путают с искусственным интеллектом. С формальной точки зрения это действительно подраздел науки об искусственном интеллекте, однако он очень разросся и оказался настолько успешным, что затмил гордого родителя. Цель искусственного интеллекта — научить компьютеры делать то, что люди пока делают лучше, а умение учиться — наверное, самый важный из этих навыков, без которого компьютерам никогда не угнаться за человеком.

Что такое машинное обучение?

Машинное обучение — это способ обработки и анализа данных, который позволяет компьютерам использовать существующие данные **для прогнозирования поведения, результатов и тенденций в будущем**.

С помощью машинного обучения компьютеры учатся сами, вам не приходится специально их программировать.

Прогнозы на основе машинного обучения помогают оптимизировать работу устройств и приложений. На основании ваших предыдущих покупок через Интернет машинное обучение позволяет рекомендовать товары, которые могут вам понравиться. Когда вы расплачиваетесь кредитной картой, машинное обучение сравнивает эту операцию с другими операциями в базе данных и помогает обнаружить мошенничество. Когда робот-пылесос убирается в комнате, машинное обучение помогает ему решить, что пора заканчивать.

Наука о данных

Наука о данных (data science; иногда даталогия — datalogy) — раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме. Объединяет методы по обработке данных в условиях больших объёмов и высокого уровня параллелизма, статистические методы, методы интеллектуального анализа данных и приложения искусственного интеллекта для работы с данными, а также методы проектирования и разработки баз данных.

- Основная практическая цель профессиональной деятельности в науке о данных — **обнаружение закономерностей в данных, извлечение знаний из данных в обобщённой форме**.
- Наука о данных — это концепция объединения статистики, анализа данных, машинного обучения и связанных с ними методов" для понимания и анализа реальных явлений.
- Она использует методы и теории, взятые из многих областей в контексте математики, статистики, информатики и компьютерных наук

- В настоящее время термин data science часто используется взаимозаменяемо с более ранними концепциями, такими как business analytics (бизнес-аналитика), business intelligence (интеллектуальный анализ данных), predictive modeling (прогнозное моделирование) и statistics (статистика). Во многих случаях более ранние подходы и решения теперь просто переименовываются в "науку о данных", чтобы стать более привлекательными.

С начала 2010-х годов считается одной из самых привлекательных, высокооплачиваемых и перспективных профессий!

В сравнении с классической статистикой, на методах которой во многом основывается и наука о данных, в ней подразумевается исследование сверхбольших разнородных массивов цифровой информации и неразрывная связь с информационными технологиями, обеспечивающими их обработку.

В сравнении с деятельностью в области проектирования и работы с базами данных, где предполагается предварительное проектирование модели данных, в науке о данных предполагается опора на аппарат **математической статистики, искусственного интеллекта, машинного обучения**, зачастую без предварительной загрузки данных в модели.

В сравнении с профессией аналитика, основная цель деятельности которого в описании явлений на основе накопленных данных относительно простыми пользовательскими средствами (вроде электронных таблиц или средств класса Business Intelligence), профиль специалиста по науке о данных в меньшей степени требует концентрации на содержании предметных областей, но требует более глубоких знаний в математической статистике, машинном обучении, программировании, и в целом более высокого образовательного уровня.

Основные отличия data science от business intelligence

В обоих случаях решающую роль играют специалисты. Главное различие между двумя специальностями заключается в том, что эксперт в области BI способен предоставить объективную картину от прошлого до текущего момента, в то время как data scientist должен понимать, как и что нужно делать

Непосредственно в курс по науке о данных входят следующие дисциплины:

- машинное обучение;
- системы управления базами данных;
- инженерия программного обеспечения;
- анализ данных (*intelligent data*) и вероятностный вывод (*probabilistic inference*), в описании дисциплины даются ссылки на байесовский вывод и алгоритмические методы моделирования, классификации и дискриминантного анализа данных на его основе;
- вероятностные модели и продвинутая статистика.

В чем разница между наукой о данных, искусственным интеллектом и машинным обучением?

Наука о данных — это широкая область знаний, которая включает в себя предварительную обработку, анализ и визуализацию структурированных и неструктурированных данных. Полученные из данных выводы затем применяются в широком спектре областей применения.

Искусственный интеллект означает обучение машины подражать человеческому поведению. Цели исследования ИИ включают представление знаний, планирование, обучение, рассуждения, обработку естественного языка, восприятие и способность манипулировать объектами.

Машинное обучение — это подмножество ИИ, которое фокусируется на том, как использовать данные и алгоритмы, чтобы имитировать способ обучения людей. Чем больше данных (также называемых обучающими данными) получает модель машинного обучения, тем точнее она делает прогнозы, не будучи явно запрограммированной на это.

ОСНОВНЫЕ ЗАДАЧИ АНАЛИЗА ДАННЫХ

Основная цель любого анализа данных – поиск и обнаружение закономерностей в объеме данных. В бизнес-анализе эта цель становится еще более широкой. Любому руководителю важно не только выявить закономерности, но и найти их причину. Знание причины позволит в будущем влиять на бизнес и дает возможность прогнозировать результаты того или иного действия.

Углубленная аналитика выходит за рамки ответа на вопросы о том, что произошло, когда это произошло и каковы последствия. Она пытается определить причины произошедшего и решить, что можно предпринять в будущем.

Углубленная аналитика подразумевает ряд мероприятий, включая использование сложных запросов SQL, прогностическое моделирование, интеллектуальный анализ данных, прогнозирование, оптимизацию и др.

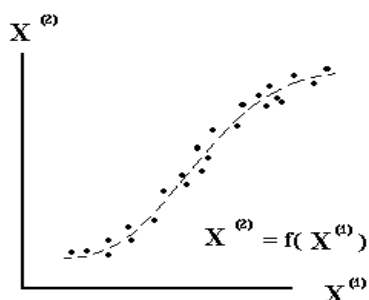
Углубленная аналитика идет дальше, чем базовая, включая в себя процедуры от сложных запросов SQL и прогнозирования до интеллектуального анализа данных и прогностического моделирования.

Углубленная аналитика представляет собой важную часть общей аналитической стратегии организации. Она может помочь вывести организацию на новый уровень. Углубленная аналитика включает в себя очень сложные запросы SQL или манипулирование данными наряду с моделированием, прогнозированием, интеллектуальным анализом данных и другими подобными методами.

1. Элементарные статистические выводы, проверка статистических гипотез

Предположим, что Вы изменили систему оплаты труда или перешли на выпуск новой продукции или использовали новую технологию и т.д. Вам кажется, что это дало положительный эффект, но действительно ли оно так? Может быть это естественная случайность, а завтра Вы можете получить прямо противоположный, но столь же случайный эффект? Для решения этой задачи надо сформировать два набора чисел, каждый из которых содержит значения интересующего Вас показателя эффективности до и после нововведения. Тогда статистические критерии сравнения 2-х выборок покажут Вам, случайны или неслучайны различия этих двух рядов чисел.

2. Исследование зависимостей одного признака от остальных



Формальная постановка: найти функцию зависимости, приближенно описывающую изменение целевого признака при изменении других признаков.

Примеры:

а) прогноз погоды по набору примет – «дым стелется, ласточки летают низко, росы нет, – значит, погода испортится».

б) прогнозирование будущего поведения некоторого временного ряда: изменение курса доллара, цен и спроса на продукцию или сырье и т.п.

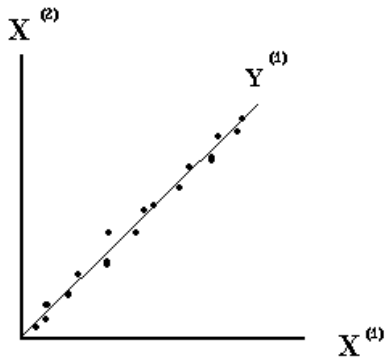
Математический аппарат: корреляционный и регрессионный анализ.

3. Снижение размерности

Формальная постановка: устранить дублирующие друг друга признаки или найти (построить) новые признаки (меньшее число), описывающие данные. При этом определяющие признаки (факторы) могут находиться среди статистически обследованных характеристик, а могут быть латентными или скрытыми, т.е. непосредственно статистически не наблюдаемыми, но восстанавливаемыми по исходным данным (т.е. матрицам X или A).

Примеры: а) гениальный пример практической реализации этого принципа дает нам периодическая система элементов Менделеева: в этом случае роль идеально информативного единственного фактора играет, как известно, заряд атомного ядра.

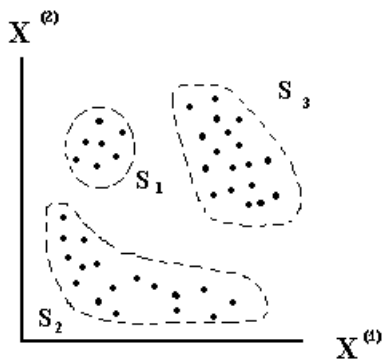
б) нахождение системы признаков «размер - рост- полнота», описывает фигуру человека и определяющий тип размера при изготовлении готовой одежды.



На рисунке изображено двухмерное пространство описания. Структура данных в нем такова, что вместо двух исходных признаков $X^{(1)}$ и $X^{(2)}$ вполне можно обойтись одним $Y^{(1)}$. Если признаков много, то одной проекцией уже не обойдешься, и придется сопоставлять результаты рассматривания структуры данных с различных точек наблюдения.

Математический аппарат: метод главных компонент, факторный анализ, многомерное шкалирование.

5. Классификация объектов (без учителя, когда исследователь не располагает обучающими выборками)



Формальная постановка: обнаружить в пространстве описания компактные скопления точек-объектов.

Примеры:

а) создание Карлом Линнеем классификации растений и животных по классам, отрядам, родам, видам.

б) исследование зависимости интенсивной миграции населения $X^{(p)}$ (профессиональной или территориальной) от ряда социально-экономических и географических факторов $x^{(1)}, x^{(2)}, \dots, x^{(p-1)}$ таких, как заработок, обеспеченность жильем, детские учреждения и т.д. Естественно предположить, что для различных однородных групп людей те же факторы влияют на $x^{(p)}$ в разной степени, а иногда в противоположных направлениях.

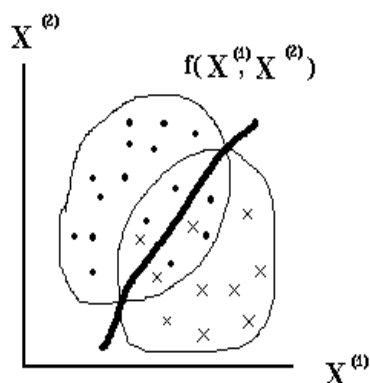
$$X_i = (X^{(1)}_i, X^{(2)}_i, \dots, X^{(p)}_i), \quad i=1, 2, \dots, n$$

на однородные классы и решать далее поставленную задачу отдельно для каждого такого класса.

Здесь классификация выступает как необходимый предварительный этап статистической обработки многомерных данных.

Математический аппарат: кластерный анализ, методы автоматической классификации, методы расщепления смесей вероятностных распределений.

6. Распознавание образов (с учителем, когда исследователь располагает обучающими выборками)



Формальная постановка: найти в пространстве описания поверхность, разделяющую группу точек, соответствующих различным образам и описать ее как функцию исходных признаков; или найти, к какой группе точек (образу) относятся заданные точки-объекты.

Пример: этот пример показывает, как удачно построенный рисунок помог не просто получить информацию об объектах, а дал толчок совершенно новому направлению исследований. С давних пор астрономы знали о различной светимости звезд, т.е. о различной их «истинной яркости». В конце XIX в. были открыты также различные спектральные классы звезд, т.е. различный цвет их излучения (от красного до голубого).

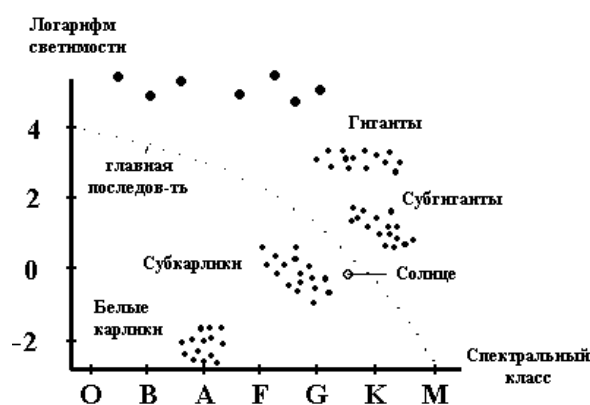


Диаграмма Герцшпрунга-Рессела

До 1913 г. эти характеристики существовали отдельно в представлении ученых, но вот независимо друг от друга датский астроном Герцшпрунт и американец Ресселл сопоставили их между собой и построили двумерную проекцию объектов-звезд на плоскость признаков спектр-светимость. Визуализация здесь помогает понять, различаются ли вообще образы, оценить форму разделяющей поверхности и получить другую важную информацию.

Математический аппарат: Дискриминантный анализ. Сегодня для этого используют нейронные сети.

Основные этапы решения задачи анализа данных

На ранних стадиях развития науки анализировались, прежде всего очевидные, тесно связанные между собой причины и следствия – вспомним легенду о яблоке и Ньютоне. Сегодня исследователь любой ПО пытается найти методами АД причины и следствия очень «далекого диапазона».

Пример: имеется ли связь между жизненным успехом и именем, которое человек получил при рождении? Оказывается, люди с забавными или странными именами в 4 раза больше остальных предрасположены к психическим заболеваниям.

Этап	Содержание
<p>Этап 1. Постановка задачи (исходный анализ исследуемой системы)</p> <p>Этап 1А. составление детального плана сбора исходной статистической информации</p>	<p>1.1. Определение цели исследования</p> <p>1.2. Определение состава данных</p> <p>1.3. Сбор данных</p> <p>1.4. Выбор средств анализа данных</p> <p>1.5. Формализация данных</p>
Этап 2. Ввод данных в обработку	<p>2.1. Ввод данных в память ЭВМ</p> <p>2.2. Работа с архивом данных</p> <p>2.3. Формирование задания обработки</p>
Этап 3. Качественный анализ	<p>3.1. Первичная статистическая обработка данных</p> <p>3.2. Визуализация данных</p> <p>3.3. Анализ структуры данных</p>
Этап 4. Количественное описание данных	<p>4.1. Выбор модели данных</p> <p>4.2. Выполнение обработки</p>
Этап 5. Интерпретация результата	<p>5.1. Анализ результатов</p> <p>5.2. Принятие решения</p>

Этап 1. Постановка задачи

Одна из типичных ошибок исследователей состоит в том, что сначала собираются данные, а затем начинают формулироваться задачи их обработки. В этом случае **цель исследования подменяется** той или иной узкой целью обработки уже собранных конкретных данных. Тогда как заранее собранные данные могут отражать совсем другие характеристики явления, нежели те, которые важны для поставленной цели.

Состав данных — это, прежде всего состав признаков, которые характеризуют объект. Каждый реальный объект имеет бесконечное число различных свойств, отражающих его различные стороны. Выделить наиболее важные признаки – задача специалиста ПО. Необходимо также решать, как представить в ТЭД значения признаков – цифрами, числами, символами.

Определив состав признаков, можно приступить к **сбору данных**. На этой стадии основной вопрос – проблема выбора объектов, т.е. сколько нужно данных.

Следующая стадия – **выбор пакета программ или системы АД**. Решение зависит от следующих факторов: объема данных, числа объектов и признаков, типов признаков, квалификации исследователя.

Пятая стадия 1-го этапа – **формализация собранных данных**. На этой стадии ТЭД необходимо придать такой вид, какой требует от входных данных выбранная пользователем автоматизированная система АД.

Этап 2. Ввод данных в обработку

Собранные данные **вводят в ЭВМ**. Введенные данные попадают в архив. **Работа с архивом** – одна из наиболее трудоемких операций. Прежде всего, необходима проверка входных данных и исправление ошибок. На этой стадии можно выполнить некоторые простейшие подсчеты. Например, выделить объекты, имеющие одинаковые описания по одному или более признакам и т.п.

Последняя стадия второго этапа – **формирование задания обработки** очень ответственна, так как неграмотный выбор метода обработки может привести к одной из ошибок:

- выбор метода, не соответствующего сути задачи;
- выбор метода, неприменимого к имеющимся данным;
- выбор метода, дающего худший результат по сравнению с другими доступными методами.

Опыт показывает, что эти ошибки приносят наибольшее число неправильных или слабых конечных результатов.

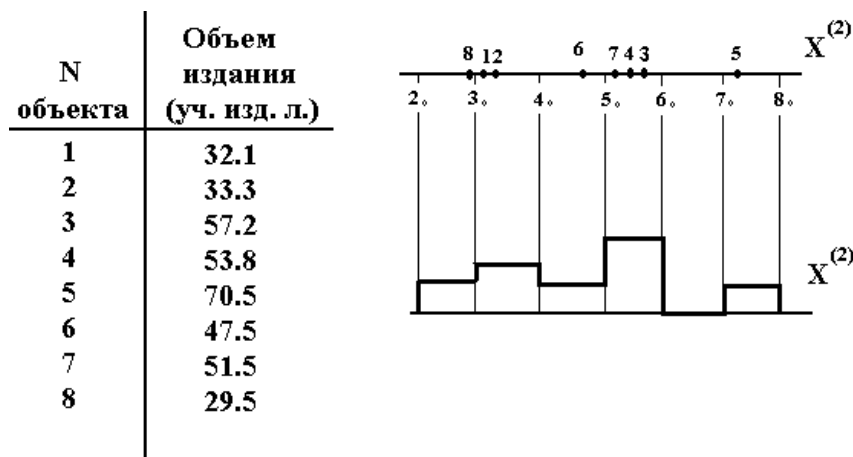
Этап 3. Качественный анализ

Решить задачу на качественном уровне – это, значит, представить собранные данные в **визуальной форме**, для того чтобы увидеть их пригодность для проверки выдвинутых гипотез или достижения поставленной цели. Почему именно увидеть? Зрительный анализатор человека – канал, по которому мозг получает наибольший объем внешней информации. В прикладной статистике этот этап статического анализа принято называть разведочным (РАД – разведочный анализ данных). Цель такого анализа – представить наблюдаемые данные в возможно более компактной и простой форме, позволяющей выявить имеющиеся в них закономерности и связи.

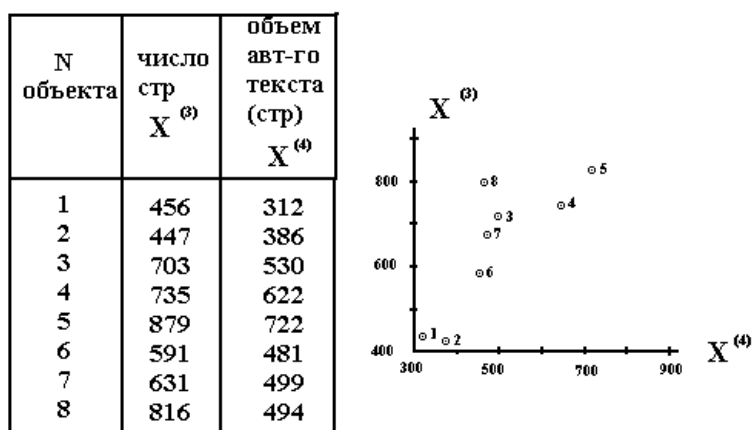
Рассмотрим наиболее важные «слова» этого языка визуальных образов.

Если взять из ТЭД один столбец чисел, то его можно изобразить в виде оси, на которой точками отмечено положение каждого объекта, а около каждой точки может стоять номер соответствующего объекта.

Это изображение называют проекцией данных на признак. Этот же признак можно изобразить иначе, если, разбив всю область его значений на некоторое количество интервалов, построить в каждом интервале столбец, высота которого пропорциональна числу попавших в интервал объектов. Такое изображение называют гистограммой объектов по признаку.



Добавим к первому признаку еще один. Тогда каждый объект можно изобразить точкой, расположенной уже в двумерном пространстве, то есть на плоскости. Такая проекция на плоскость двух признаков называется диаграммой рассеяния (ДР) объектов. Добавив 3-й, 4-й ... признаки получим представление объектов в пространстве 3-х, 4-х и т.д. измерений. Таким образом каждому объекту ТЭД можно сопоставить точку в гипотетическом многомерном пространстве. Тогда подбор таких точек-объектов будет называться структурой данных в пространстве описания.



В качестве отображенных на ДР единиц могут выступать объекты, переменные(признаки), категории переменных (если переменные не количественные). Их называют отображаемыми единицами (ОЕ). Графически же элементы, с

помощью которых ОЕ изображаются на ДР, называются выразительными элементами (ВЭ).

Гипотез, объясняющих явление, может быть множество, следовательно, должен иметься аппарат, помогающий осуществить их проверку. В АД таким аппаратом является вычислительный эксперимент с данными, то есть применение к данным определенного метода машинной обработки, выбранного в соответствии с гипотезой исследователя.

На этом этапе основные гипотезы касаются **структуры данных** – именно ее необходимо исследовать. Поэтому задача исследователя состоит в построении проекций данных на различные пары признаков (на какие - следует определить исходя из выдвинутой гипотезы и смысла задачи); исследование отдельных признаков; поиск дублирующих друг друга или избыточных признаков и т.д.

Этап 4. Количественное описание данных

На этом этапе обычно ведется поиск параметров моделей, найденных на предыдущем этапе. Вычислительный эксперимент дает возможность испытывать различные варианты моделей, т.е. искать различные способы информативного описания данных, а сопоставительный анализ помогает отбирать лучшие варианты (не только на формальном, но и на содержательном уровне. Пуанкаре – формальные результаты теории относительности были получены за 20 лет до Эйнштейна).

Этап 5. Интерпретация результата

Все получаемые на ЭВМ результаты специалист по анализу данных (АД) может интерпретировать, не выходя за рамки понятий АД, в терминах информативных признаков, группировок объектов и т.д. пользователь же переводит полученный результат на привычный язык предметной области (ПО).

Таким образом и в предыдущем этапе выделяют два уровня – формальный и содержательный, причем определяющим является второй.

Однако обработка данных должна заканчиваться не результатом, а рекомендациями по его применению или принятию решения относительно дальнейших действий.

Здесь так же, как и при интерпретации результатов существуют два уровня – в рамках АД и ПО.

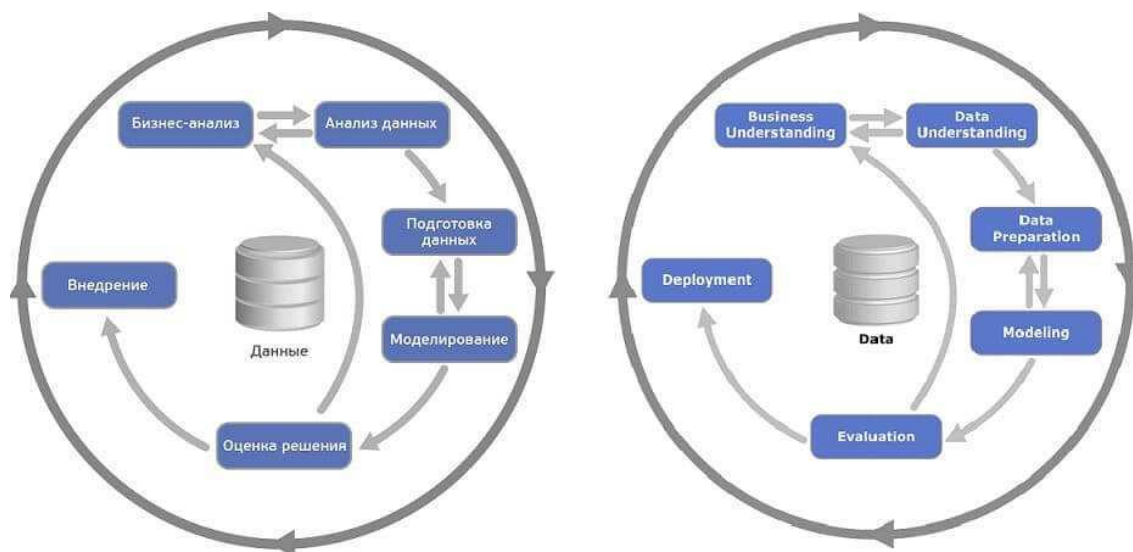
В рамках АД

- 1) Может быть принято решение прекращения дальнейшей обработки, так как поставленные цели достигнуты.
- 2) Продолжение обработки данных с использованием других методов, возможно с коррекцией данных.

CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) — межотраслевой стандартный процесс исследования данных. Это проверенная в промышленности и наиболее распространённая методология, первая версия которой была представлена в Брюсселе в марте 1999 года, а пошаговая инструкция опубликована в 2000 году.

CRISP-DM описывает жизненный цикл исследования данных, состоящий из 6 фаз, от постановки задачи с точки зрения бизнеса до внедрения технического решения.



Фазы CRISP-DM

Последовательность между фазами определена не строго, переходы могут повторяться от итерации к итерации. Все фазы CRISP-DM делятся на задачи, по итогам каждой должен быть достигнут конкретный результат.

Понимание бизнеса

Понимание бизнеса (Business Understanding) – определение целей проекта и требований со стороны бизнеса. Затем эти знания конвертируются в постановку задачи интеллектуального анализа данных и предварительный план достижения целей проекта. Задачи фазы Business Understanding:

- Определить бизнес-цели
- Оценить ситуацию
- Определить цели анализа данных
- Составить план проекта

Начальное изучение данных

Начальное изучение данных (Data Understanding) – сбор данных и знакомство с информацией, выявление проблем с качеством данных (ошибки или пропуски). Необходимо понять, какие сведения имеются, попробовать отыскать интересные

наборы данных или сформировать гипотезы о наличии в них скрытых закономерностей. Задачи фазы Data Understanding:

- *Собрать исходные данные*
- *Описать данные*
- *Исследовать данные*
- *Проверить качество данных*

Подготовка данных

Подготовка данных (Data Preparation) – получение итогового набора данных, которые будут использоваться при моделировании, из исходных разнородных и разноформатных данных. Задачи фазы Data Preparation могут выполняться много раз без какого-то заранее определенного порядка:

- *Отобратить данные (таблицы, записи и атрибуты)*
- *Очистить данные, в т.ч. выполнить их конвертацию и подготовку к моделированию*
- *Сделать производные данные*
- *Объединить данные*
- *Привести данные в нужный формат*

Моделирование

Моделирование (Modeling) – в этой фазе к данным применяются разнообразные методики моделирования, строятся модели и их параметры настраиваются на оптимальные значения. Обычно для решения любой задачи анализа данных существует несколько различных подходов. Некоторые подходы накладывают особые требования на представление данных. Таким образом часто бывает нужен возврат на шаг назад к фазе подготовки данных. Задачи фазы Modeling:

- *Выбрать методику моделирования*
- *Сделать тесты для модели*
- *Построить модель*
- *Оценить модель*

Оценка

Оценка (Evaluation) – анализ количественных характеристик качества модели, подтверждение или опровержение того, что, благодаря построенной модели все бизнес-цели достигнуты. Основной целью этапа является поиск важных бизнес-задач, которым не было уделено должного внимания. Задачи фазы Evaluation:

- *Оценить результаты*
- *Сделать ревью процесса*
- *Определить следующие шаги*

Внедрение

Внедрение (Deployment) – в зависимости от требований фаза развертывания может быть простой (составление финального отчета) или сложной, например, автоматизация процесса анализа данных для решения бизнес-задач. Обычно

развертывание — это внедрение полученных моделей в прикладную сферу. Задачи фазы Deployment:

- *Запланировать развертывание*
- *Запланировать поддержку и мониторинг развернутого решения*
- *Сделать финальный отчет*
- *Сделать ревью проекта*

Первичная обработка данных

- Методы предварительной обработки данных – очистка, нормализация, выделение признаков, преобразование, уплотнение.
- Описательные характеристики и отображение данных.
- Классификация статистических данных по различным критериям

Предварительная обработка данных является важным шагом в процессе интеллектуального анализа данных. Фраза «мусор на входе — мусор на выходе» применима, в частности, и для проектов интеллектуального анализа данных и машинного обучения. Здесь имеется в виду то, что даже самый изощренный анализ не принесет пользы, если за основу взяты сомнительные данные.

Методы сбора данных часто плохо контролируются. Это приводит к появлению недопустимых значений (к примеру: доход, равный -100), комбинаций данных, которые невозможны (к примеру: «мужской пол при наличии беременности»), отсутствию значений и прочее. В результате анализа данных, которые не защищены от такого рода проблем, можно прийти к неверным выводам. Качество данных является первостепенной задачей при проведении анализа.

Предварительная подготовка данных включает в себя:

- *очистку*
- *отбор экземпляров*
- *нормализацию*
- *преобразование данных*
- *выделение признаков*
- *отбор признаков*

Методы

Очистка данных используется для обнаружения, исправления или удаления ошибочных записей в наборе данных;

Нормализация данных используется для стандартизации диапазона значений независимых переменных или признаков данных (например, сведение к интервалам $[0, 1]$ или $[-1, +1]$);

Преобразование данных используется для приведения данных в формат, который ожидает аудитория;

Выделение признаков используется для преобразования входных данных в набор признаков, которые они хорошо представляют;

Уплотнение данных используется для преобразования числовых данных в исправленный, упорядоченный и упрощённый вид. Это помогает уменьшить количество и/или размерность данных.

Описательные характеристики и отображение данных

Описательная статистика занимается получением, систематизацией, обработкой и изучением тех или иных данных с использованием различных числовых характеристик.

Задача описательной статистики, как следует из названия, — дать хорошее описание данных. Она не для предсказаний, выводов или преобразований — только внешняя форма данных, измеренная в показателях.

Основные статистические показатели

Основные статистические показатели можно разделить на две группы:

1. меры среднего уровня
2. меры рассеяния.

Меры среднего уровня

- Среднее значение
- Стандартная ошибка
- Стандартное отклонение
- Эксцесс
- Асимметрия
- Интервал
- Минимум
- Максимум
- Счёт
- Медиана
- Мода
- Квантиль
- Математическое ожидание
- Доверительный интервал

Меры рассеяния

Меры рассеяния показывают, насколько хорошо данные значения представляют данную совокупность.

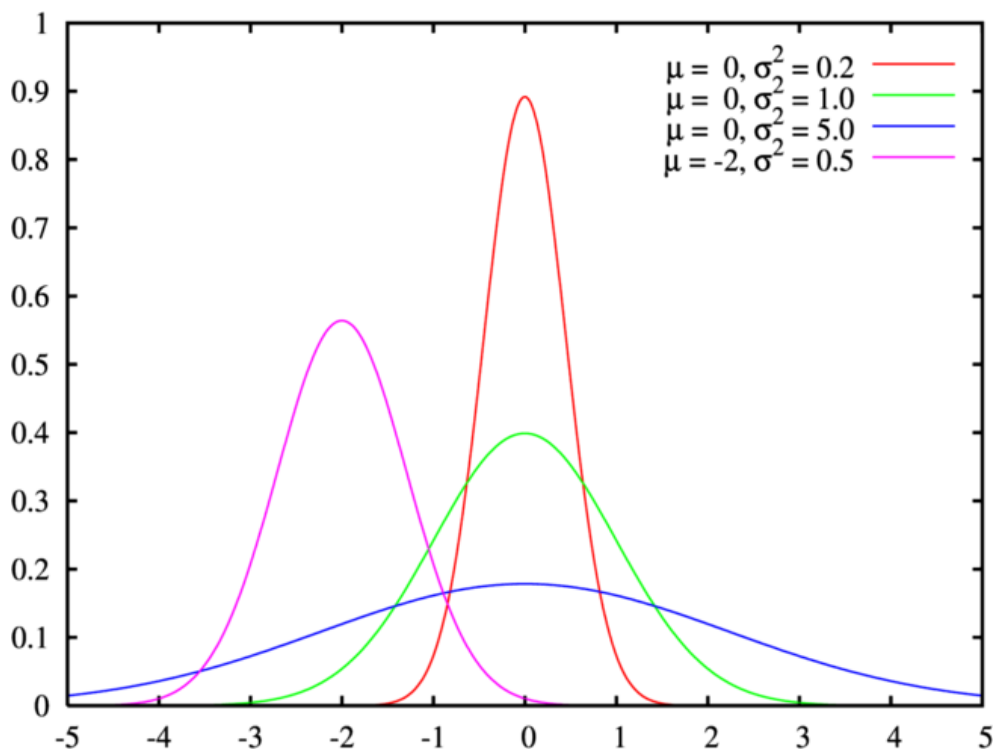
- Дисперсия случайной величины
- Среднеквадратическое отклонение
- Размах вариации
- Интерквартильный размах
- Среднее абсолютное отклонение

Распределения

Внешняя форма данных, выраженная в мерах описательной статистики, даёт нам информацию об их характере. Это как в жизни: по фигуре, походке и одежде человека обычно можно догадаться о его поле, возрасте и даже профессии. В случае числовых данных мы говорим о распределении.

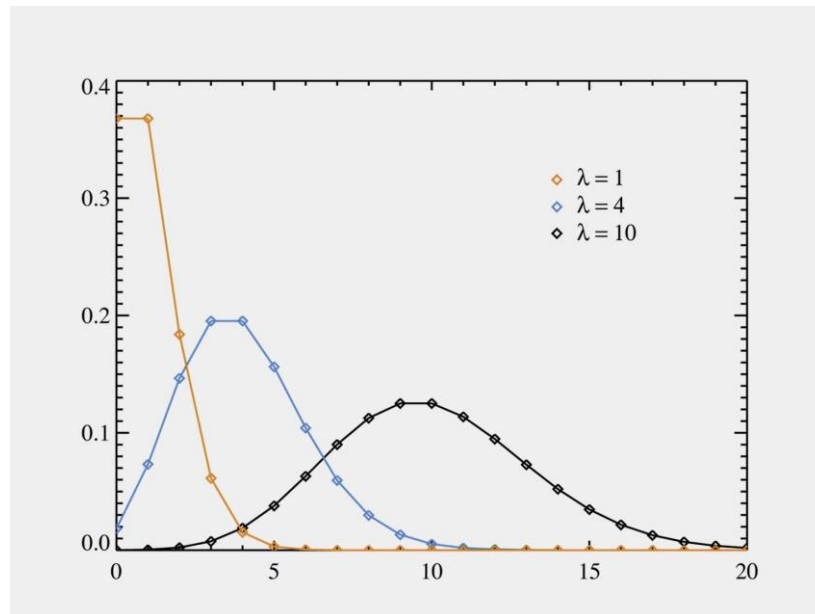
Нормальное распределение

Нормальное распределение описывает процессы, где результат является суммой многих случайных величин, каждая из которых слабо зависит от другой и вносит сравнительно небольшой вклад.

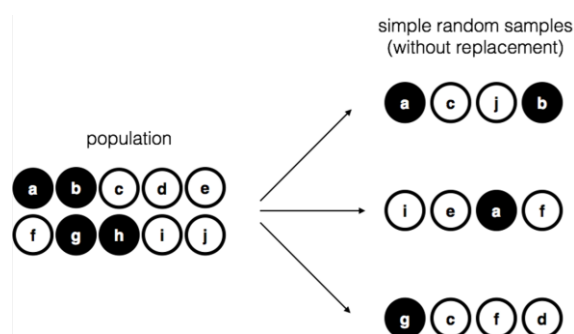


Распределение Пуассона

Распределение Пуассона тоже часто встречается в работе дата-сайентистов и аналитиков: это число событий за какой-то промежуток времени — при условии, что события независимы друг от друга и имеют некоторый порог интенсивности.



Семплирование



Семплирование — это группа статистических методов и приёмов, отвечающих за релевантность выборки. С помощью семплирования мы формируем нашу выборку так, чтобы она наилучшим образом отражала свойства генеральной совокупности.

Визуализация данных

Визуализация данных — это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению. Визуализация данных находит широкое применение в научных и статистических исследованиях (в частности, в прогнозировании, интеллектуальном анализе данных, бизнес-анализе), в педагогическом дизайне для обучения и тестирования, в новостных сводках и аналитических обзорах. Визуализация данных связана с визуализацией информации, инфографикой, визуализацией научных данных, разведочным анализом данных и статистической графикой.

По цели представления данных визуализация делится на презентационную (англ. «presentation», «explanation») и исследовательскую (англ. «exploration»). Презентационная визуализация предназначена для представления данных некоторой

аудитории (например, в рамках научной работы, доклада или аналитического обзора в новостях). Исследовательская визуализация предназначена для анализа и обработки набора данных, например, с целью обнаружения закономерностей в них.

Визуализация как этап анализа данных

Подсистема визуализации данных является важной составной частью качественных систем интеллектуального анализа данных, особенно ориентированных на обработку больших объемов информации. В системах бизнес-аналитики визуализация может использоваться на всех этапах процесса обработки данных:

- Визуализация исходных данных. Этот этап полезен для оценки степени соответствия ожиданиям и пригодности данных к анализу, выдвижения гипотез о закономерностях и необходимых процедурах первичной обработки.
- Визуализация выборки, загруженной в систему обработки.
- Визуализация результатов первичной обработки.
- Визуализация промежуточных результатов.
- Визуализация окончательных результатов.

Классификация переменных

Переменные статистического анализа можно классифицировать:

1. По роли в исследовании

- результирующие (выходные, отклик, прогнозируемые, эндогенные)
- объясняющие (входные, предсказывающие, предикторные, экзогенные)
- скрытые (латентные, остаточные)

2. По вероятностной природе

- детерминированные
- случайные

3. По шкале измерений

- количественные
- порядковые (качественные, ординальные)
- номинальные (классификационные)

В основу классификации по шкале измерений положен объем допустимых операций над числами.

Количественные признаки измеряются в некоторой шкале (длина, вес, скорость) и выражаются целыми или дробными числами.

Порядковые признаки можно естественным образом упорядочить по их значениям, они имеют градации, которые выражаются чаще всего словами, образующими упорядоченный ряд (плохо, удовлетворительно, хорошо, отлично) или целыми числами. Однако с этими значениями в отличие от количественных нельзя производить обычных математических действий: суммировать, умножать, делить и т.п.

Номинальные признаки отражают деление объектов на различные классы (пол, возраст, профессия), но не несут никакой информации об упорядоченности классов или об их количественных соотношениях. Чаще всего признаки задаются словами, но возможно использование также специальных символов и чисел. Но эти символы и числа являются лишь метками, про которые можно сказать только, что они либо одинаковы, либо различны.

Лекция 2. Исследование зависимостей. Корреляционный анализ

ВВЕДЕНИЕ В МНОГОМЕРНЫЙ СТОХАСТИЧЕСКИЙ АНАЛИЗ

Довольно часто невозможно дать оценку какому-либо явлению или состоянию по одному признаку. Например:

- оценить способности выпускника, по одной оценке, необходимо знать все или большинство оценок аттестата или диплома;
- точность выстрела можно оценить только по двум координатам.

При этом исследовать признаки необходимо не изолированно друг от друга, а с учетом характера и структуры их взаимосвязи. Что понимается под многомерной структурой? Речь идет о множестве объектов:

$$\{O_1, O_2, \dots, O_n\},$$

каждый из которых может характеризоваться многомерным вектором наблюдений. Результаты их статистического обследования представляются, как правило, в одной из двух форм:

1. Таблицы (матрицы) «объект-признак» (или «объект-свойство»).
2. Таблица (матрица) попарных сравнений

Таблица «объект-признак»

Примеры признаков: вес, длина, цвет, профессия, цена, пол, наличие и отсутствие симптома и т.д. Примеры объектов: люди, изделия, место рождения и т.д. Таблицы такого вида (табл.1) принято называть таблицей экспериментальных данных (ТЭД). Это название следует трактовать более широко, говоря не только об экспериментальных данных, а о данных научного исследования вообще.

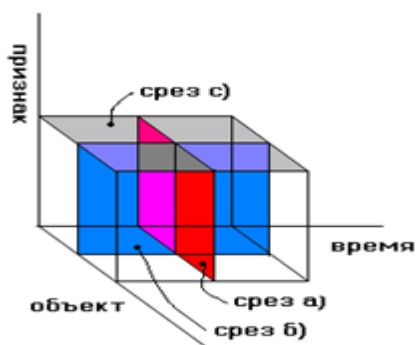
Таблица 1.

Объект	Признаки (свойства, параметры, показатели)					
	$X^{(1)}$	$X^{(2)}$...	$X^{(i)}$...	$X^{(p)}$
Объект 1	$X_1^{(1)}$	$X_1^{(2)}$		$X_1^{(i)}$		$X_1^{(p)}$
...						
Объект N	$X_N^{(1)}$	$X_N^{(2)}$		$X_N^{(i)}$		$X_N^{(p)}$

Если исследователь вводит в анализ еще одну составляющую – время, то мы переходим к кубу данных, то есть к трехмерному пространству (см. рис. 1): «признак - объект – время» (p-o-t).

Фиксируя одну из трех координат, мы получим плоские двухмерные срезы куба данных:

- а) при фиксации времени – матрицу объект-признак.
- б) при фиксации объекта – матрицу признак-время, это временной срез данных.
- в) при фиксации признака – матрицу объект-время, например, фиксируется цена на акции. Тогда объекты – виды акций, строки соответствуют временным интервалам (дни, недели и т.д.).



Представление информации в виде куба является грубым упрощением. В частности, можно предположить, что исследователя, помимо p, o и t интересуют условия получения информации. Условий может быть много, и исследователь может ставить своей задачей изучение изменений в пространстве p-o-t в зависимости от изменения условий.

Рис. 1. Куб данных

Таким образом может быть получен четырехмерный куб данных. Добавьте еще одну координату – способ измерения информации, и вы получите пятимерный куб и т.д.

Введем следующие обозначения:

X – случайная величина;

x_i – i -ая реализация случайной величины;

$\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(p)})$ – случайный p -мерный вектор (многомерная случайная величина, многомерный признак);

$\mathbf{x}_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})$ – i -ая реализация случайного вектора;

$\mathbf{X}^{p \times n}$ – матрица;

$\mathbf{X}^{1 \times n}$ – одна строка таблицы данных.

Пусть $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(p)})$ – случайный p -мерный вектор, где координаты $X^{(1)}, \dots, X^{(p)}$ есть случайные величины (цена, размер, вес, пол и т.п.). Тогда реализация вектора \mathbf{X} (уже не случайная величина) $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})$ и есть i -ая строка ТЭД. После n реализаций вектора \mathbf{X} получим n неслучайных p -мерных векторов $\mathbf{X}^{p \times n} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$, которые называют многомерной выборкой и которые образуют двумерную матрицу данных. Таким образом, матрица данных – это одна выборка случайного многомерного вектора $\mathbf{X} = (X^{(1)}, X^{(2)}, \dots, X^{(p)})$ с объемом выборки, равным n .

Таблица (матрица) попарных сравнений

$$A = \begin{vmatrix} a_{11} & a_{12} & & \\ a_{21} & \dots & & \\ & & & a_{nn} \end{vmatrix}$$

где a_{ij} – результат сопоставления объектов O_i и O_j (мера сходства или различия, мера их связи, геометрические расстояния, отношения предпочтения и т. д.).

Априорно можно предположить, что существует небольшое (по сравнению с “ p ”) число определяющих факторов, с помощью которых могут быть точно описаны как наблюдаемые характеристики анализируемых объектов (т.е. все элементы $x_i^{(k)}$ и a_{ij} матриц \mathbf{X} и \mathbf{A}) и характер связей между ними, так и искомая классификация самих объектов.

ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ

Функционирование многих систем можно задать описанием зависимостей между входными и выходными переменными. В сложной системе учесть влияние на результат всех переменных практически невозможно.

Поэтому выходные переменные всегда будут подвержены случайному неконтролируемому разбросу, обусловленному действию неучтенных факторов.



Рис.2 Модель системы

Представим исследуемую систему (рис.2) в виде «черного ящика», на вход которого воздействуют два вектора: предиктор $X=(x^{(1)}, \dots, x^{(p)})$ и случайный вектор $E=(\varepsilon^{(1)}, \dots, \varepsilon^{(m)})$. Результат представим вектором $Y=(y^{(1)}, \dots, y^{(m)})$. Тогда вместо модели

$$Y=f(X^{(1)}, X^{(2)}, \dots, X^{(k)}),$$

учитывающей действие всех переменных и где k слишком велико для практических целей, рассматривается зависимость с настолько малым числом объясняющих переменных X , что Y можно записать

$$Y=f(X, E).$$

В этой модели суммарный эффект от воздействия остальных факторов отражает векторная переменная E . Для скалярной переменной Y можно записать

$$Y=f(X, \varepsilon).$$

В этих условиях задача исследователя состоит в том, чтобы по результатам N измерений исследуемых переменных X построить такую векторную функцию $f(X)$, которая позволила бы наилучшим образом восстанавливать значения Y по заданным значениям X .

Какова конечная цель статистического исследования зависимостей? С этого вопроса должно начинаться любое статистическое исследование. Можно выделить три основных типа прикладных целей:

1. Выявление причинных связей между Y и X , проникновение в «физический механизм» изучаемых связей.
2. Прогноз неизвестных значений индивидуальных или средних результирующих показателей по заданным значениям X соответствующих переменных.

3. Установление самого факта наличия (или отсутствия) статистической связи между Y и X .

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционным анализом называется анализ структуры и тесноты статистической связи между исследуемыми переменными. Корреляционный анализ (КА) включает в себя следующие этапы:

1. Обоснованный выбор характеристики, измеряющей степень тесноты статистической связи.
2. Оценка числового значения выбранной характеристики.
3. Проверка гипотезы о том, что полученное числовое значение характеристики действительно свидетельствует о наличии статистической связи.
4. Исследование структуры связей между компонентами исследуемого многомерного признака.

Представим все виды характеристик связи классификационной схемой, представленной табл.3.

Зависимости между переменными могут носить как причинный, так и статистический характер (рис.8). Причинная связь – это связь, при которой осуществление одного события является достаточным условием для осуществления другого события.

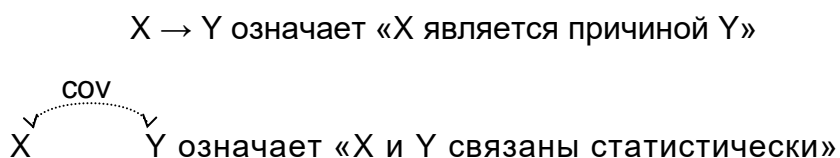


Рис. 8. Обозначение статистической связи

Статистическая или корреляционная связь устанавливает лишь сам факт взаимосвязи, но ничего не говорит о направлении причинно-следственной связи. Известно, что имеется сильная положительная корреляционная взаимосвязь между общим числом больных и количеством больниц. Такую корреляционную связь можно объяснить наличием общего фактора – численности населения, влияющего на обе переменные. Таким образом, даже из тесной зависимости случайных величин не всегда следует их причинная взаимообусловленность.

Таблица 2

			Ковариация
--	--	--	------------

Характеристики статистической связи	количественных переменных	Анализ парных связей	Коэффициент корреляции
			Корреляционное отношение
		Анализ множественных связей	Коэффициент множественной корреляции
	порядковых переменных	Анализ парных связей	Ранговый коэффициент корреляции Спирмена
			Ранговый коэффициент корреляции Кендалла
		Анализ множественных связей	Коэффициент конкордации Кендалла
	номинальных переменных	Анализ множественных связей	Таблица сопряженности

КАК ПОЯВЛЯЮТСЯ СТАТИСТИЧЕСКИЕ ВЗАИМОСВЯЗИ

Рассмотрим на простом примере влияние на прогноз возмущающих неконтролируемых воздействий. Пусть система определена линейным уравнением $Y=0,5X$. Это функциональная зависимость (рис.3а).

Что произойдет, если в систему войдут несколько более сложные причинные отношения

$$X \xrightarrow{0,5} Y \xleftarrow{0,25} E.$$

Добавление новой переменной ведет к уменьшению степени чисто линейного соответствия между X и Y. Точки на диаграмме уже не ложатся строго на прямую и больше нельзя сделать точного прогноза значений Y для нового случая (рис.3б).

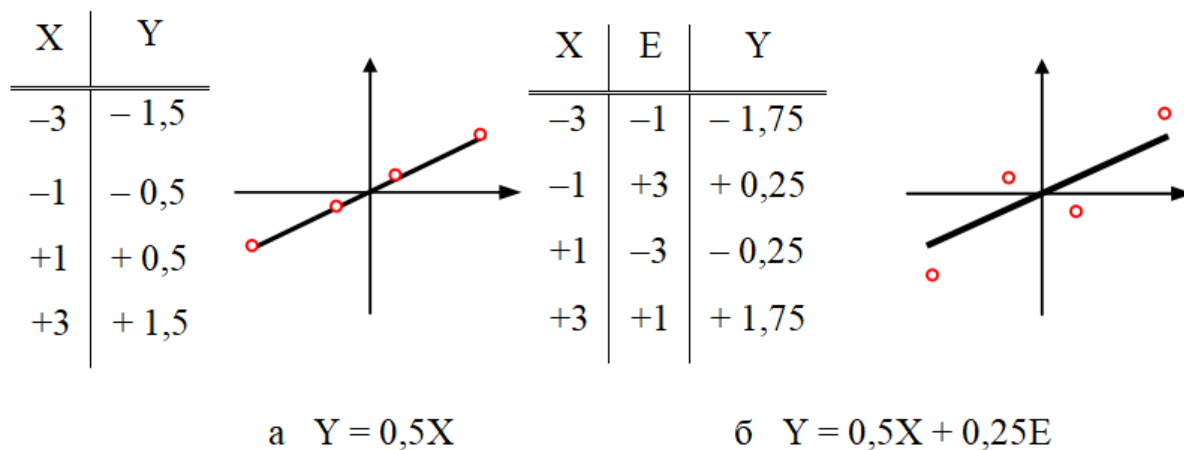


Рис. 3. Влияние возмущений на прогноз

Это показывает, каким способом возмущения искажают прогноз. В этом случае у Y имеется дополнительный разброс, производимый E, и поэтому X объясняет меньшую долю общего разброса Y. Можно сказать, что во втором случае Y имеет большую дисперсию, чем в первом.

Таким образом, наличие множества возмущающих неконтролируемых воздействий ведет к отклонению от функциональных зависимостей и появлению статистических связей.

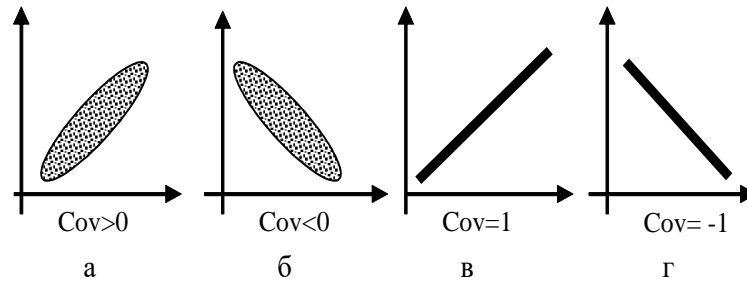
АНАЛИЗ ТЕСНОТЫ СВЯЗИ МЕЖДУ КОЛИЧЕСТВЕННЫМИ ПЕРЕМЕННЫМИ

Диаграммы рассеивания

О наличии или отсутствии статистической связи можно судить по виду диаграммы рассеивания. Представленные на диаграмме данные часто могут быть вписаны в геометрическую фигуру, имеющую форму эллипса (рис 10а и 10б). В этом случае можно предположить наличие линейной статистической связи. Ориентация эллипса отражает положительную (с ростом X растет и Y) или отрицательную связь. В общем случае, чем уже эллипс, тем выше степень тесноты связи. На рис. 11а отображена положительная корреляционная связь, характеризующаяся правым поворотом эллипса, рис.10б отображает отрицательную корреляционную связь.

Сужение или расширение эллипса ведет соответственно к двум крайним случаям:

Рис. 10. Наличие ковариационной связи



1. Линия – функциональная линейная зависимость (рис. 10в и 10г) или отсутствие зависимости (рис. 11а).

2. Круг – отсутствие зависимости (рис. 11б).

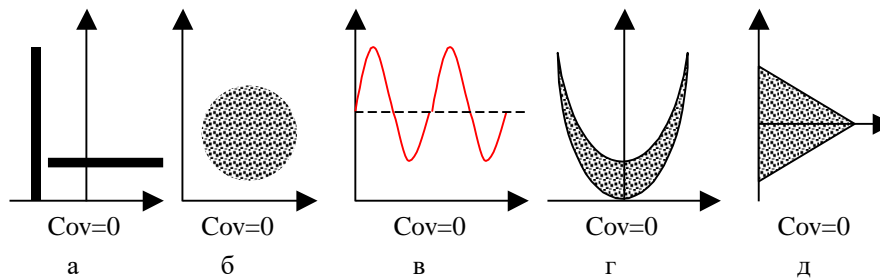


Рис. 11. Отсутствие ковариационной

На рис 11в отображен случай нелинейной функциональной зависимости с нулевой ковариацией. На рис 11г отображен случай нелинейной статистической зависимости с нулевой ковариацией.

Ковариация

Диаграммы рассеивания полезны для установления факта существования соответствия между двумя переменными и его формы. Однако для точного исследования необходимо оценить степень этого соответствия. Наиболее подходящая мера для этого – ковариация. Ковариация есть математическое ожидание произведения двух центрированных случайных величин:

$$\text{Cov}(X, Y) = K_{XY} = E(\dot{X} \cdot \dot{Y}) = E((X - EX)(Y - EY)). \quad (1)$$

Если $X=Y$, имеем:

$$\text{Cov}(X, X) = K_{XX} = E(\dot{X} \cdot \dot{X}) = E(X - EX)^2 = DX, \quad (2)$$

то есть ковариация случайной величины равна ее дисперсии.

Выборочная оценка ковариации обычно вычисляется по формуле:

$$\hat{\text{Cov}}(X, Y) = \hat{K}_{xy} = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x})(y_j - \bar{y}). \quad (3)$$

Дадим геометрическую интерпретацию ковариации. Предположим, что мы имеем два ряда наблюдений, представленных на диаграмме рассеяния (рис.12).

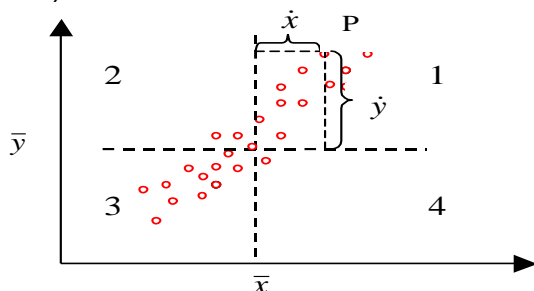


Рис. 5 Интерпретация ковариации

Рис. 12. Интерпретация ковариации

Разобьем диаграмму на четыре квадранта с помощью перпендикуляров к осям, проведенных из точки с координатами \bar{X} и \bar{Y} . Тем самым для любой точки P с координатами (x_i, y_i) будут определены отклонения $\dot{x}_i = x_i - \bar{x}_i$ и $\dot{y}_i = y_i - \bar{y}_i$.

Такая процедура носит название центрирования (рис.13). В результате этого происходит перенос всех точек диаграммы рассеяния в новую систему координат $\dot{X} \dot{Y}$ с центром в начале координат.

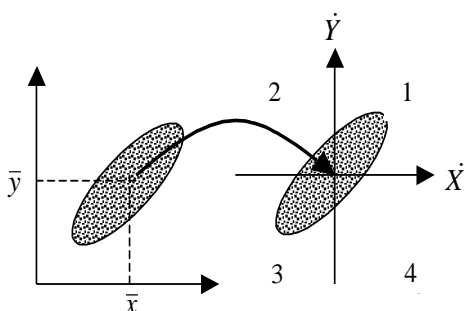


Рис. 13. Перенос координат – центрирование

Из рассмотрения диаграммы очевидным образом следует, что:

- для всех точек квадранта 1 произведение $\dot{x}_i \dot{y}_i$ положительно;
- для всех точек квадранта 2 произведение $\dot{x}_i \dot{y}_i$ отрицательно;
- для всех точек квадранта 3 произведение $\dot{x}_i \dot{y}_i$ положительно;
- для всех точек квадранта 4 произведение $\dot{x}_i \dot{y}_i$ отрицательно.

Следовательно, величина $\sum \dot{x}_i \dot{y}_i$ может служить мерой зависимости между переменными X и Y. Если зависимость положительна, так что большая часть точек лежит в 1 и 3-м квадрантах, то $\sum \dot{x}_i \dot{y}_i$ становится положительной. Если, наоборот, зависимость отрицательна, так что большая часть точек лежит во 2-м и 4-м

квадрантах, то $\sum \dot{x}_i \dot{y}_i$ становится отрицательной. Наконец, когда нет связи между X и Y , точки рассеиваются по всем четырем квадрантам и $\sum \dot{x}_i \dot{y}_i$ становится очень малой.

Теперь можно сделать выводы из анализа графиков рис. 10 и рис. 11:

1. Ковариационная связь будет положительной, если с ростом X возрастает и Y .
2. Ковариационная связь будет отрицательной, если с ростом X уменьшается Y .
3. Ковариационная связь будет отсутствовать, если экспериментальные данные ложатся на прямую, параллельную одной из координат.
4. Ковариация как мера статистической взаимосвязи пригодна для линейных связей.
5. Для функциональной линейной связи ковариация максимальна, для функциональной нелинейной связи может быть равной нулю (рис. 11в).
6. Нулевая ковариация не всегда означает отсутствие статистической связи. Для нелинейной статистической связи она может быть равной нулю (рис. 11г).
7. Ковариационная связь между случайными переменными – это частный случай их статистической связи вообще. Например, на рис. 11д ковариационная связь отсутствует, однако статистическая связь существует: с ростом X существенно уменьшается дисперсия – т. е. существует дисперсионная связь.

Коэффициент парной корреляции

Ковариация обладает двумя недостатками:

- ее численное значение может быть произвольно увеличено в результате добавления «отдаленных наблюдений»;
- оно может быть произвольно изменено путем выбора единиц измерения переменных X и Y .

Эти недостатки можно исправить, выразив отклонения в единицах стандартных отклонений, усреднение которых даст нам смешанный момент или коэффициент корреляции.

Пусть случайные векторы X и Y имеют двумерное нормальное распределение $N(\mu_1, \mu_2, \sigma_1, \sigma_2)$. Тогда простой (или смешанный) момент между X и Y есть коэффициент парной корреляции Пирсона

$$\text{Cor}(X,Y) = \rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \quad (4)$$

Для случайной выборки получим оценку коэффициента корреляции

$$r = \hat{\rho} = \frac{\text{cov}(X,Y)}{s_x s_y} \quad (5)$$

Коэффициент корреляции является мерой тесноты только для линейной статистической связи между анализируемыми признаками. Только в случае совместной нормальной распределенности исследуемых случайных величин X и Y коэффициент корреляции r имеет четкий смысл как характеристика степени связи между ними. Если же априори допускается возможность отклонения от линейного вида зависимости, то можно построить примеры, когда, несмотря на $r=0$, исследуемые переменные оказываются связанными чисто функциональным соотношением. Поэтому о величинах, для которых $r=0$, обычно говорят, что они не коррелированы, и только после дополнительного анализа, можно сказать, следует ли отсюда их независимость. И, наоборот, из высокой степени коррелированности величин при сильных отклонениях распределения случайных величин X и Y от нормального не следует их столь же тесная взаимосвязь.

Как видно из рис.14, после добавления нового объекта коэффициент корреляции значительно возрос. Дело здесь в том, что последний объект является аномальным, резко выделяющимся, так что всю совокупность наблюдений уже нельзя считать выборкой из одной и той же нормальной генеральной совокупности.

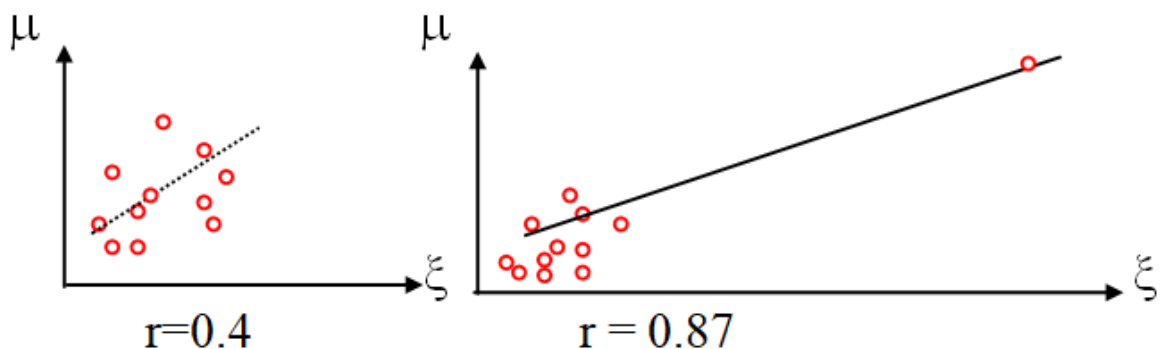
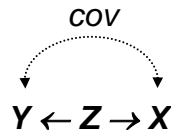


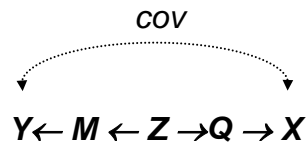
Рис. 14. Отклонение от нормального распределения

Но даже из тесной зависимости случайных величин не всегда следует их причинная взаимообусловленность. Рассмотрим примеры так называемой «ложной» корреляции. Например, переменные X и Y могут быть статистически коррелированы, если

- переменные зависят от входной переменной:



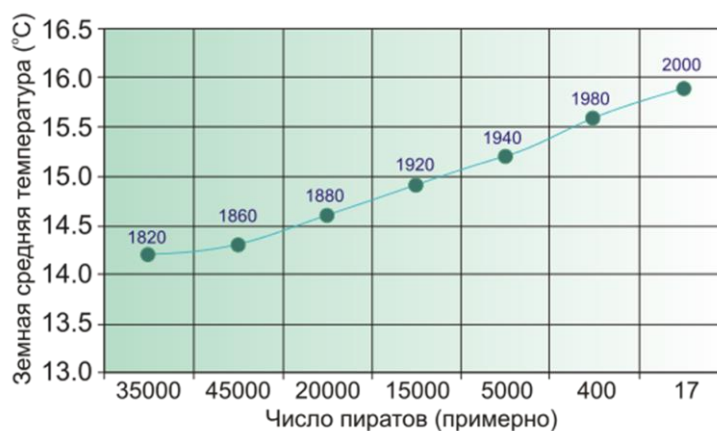
- существует координация между влияющими на них переменными



Это приводит к необходимости введения таких измерителей статистической связи, которые были бы очищены от опосредованного влияния других переменных, давали бы оценку степени тесноты связи между переменными X и Y при условии, что значения остальных переменных зафиксированы на постоянном уровне. Такой коэффициент корреляции называется частным, очищенным.

Например, можно наблюдать зависимость между ростом человека и длиной волос: чем выше человек, тем короче его волосы. Однако эта корреляция ложная, так как обе переменные «Рост» и «Длина волос» зависят от одной входной переменной «Пол»: женщины в среднем ниже мужчин и носят длинные волосы. Исключив влияние этой переменной (зафиксировав ее значение), мы обнаружим отсутствие стохастической взаимосвязи.

Влияние количества пиратов на глобальное потепление



Частные коэффициенты корреляции

Характеризуют тесноту связи между двумя случайными величинами при исключении влияния остальных случайных величин. Частный коэффициент корреляции величин X_1 и X_2 , входящих в систему $\{X_1, X_2, X_3, \dots, X_n\}$ относительно величин X_3, X_4, \dots, X_n обозначается через $r_{12.34\dots n}$. Его числовое значение находится из формулы

$$r_{12.34\dots n} = -\frac{P_{12}}{\sqrt{P_{11}P_{22}}}, \quad (6)$$

где $P_{11}, P_{12}, \dots, P_{22}$ – миноры детерминанта квадратной корреляционной матрицы (19).

Коэффициент множественной корреляции

Коэффициент корреляции Пирсона характеризует связь двух случайных величин. Коэффициент множественной корреляции используется для описания системы случайных величин $\{X^{(1)}, X^{(2)}, X^{(3)}, \dots, X^{(p)}\}$. Он служит характеристикой корреляции между величиной $X^{(1)}$ с одной стороны, и всей совокупностью величин $\{X^{(2)}, X^{(3)}, \dots, X^{(p)}\}$ с другой. Например, мы хотим выяснить, существует ли связь между ростом ($X^{(1)}$) с одной стороны и весом ($X^{(2)}$), средним ростом родителей ($X^{(3)}$) с другой. Формально он определен для любой многомерной системы. Получим его из следующих соображений.

Пусть зависимость задана моделью регрессии

$$Y = F(X) + E$$

где $F(X) = \hat{Y}$ – регрессия; E – остаток, ошибка.

Определим остаточную дисперсию как средний квадрат расстояний между наблюдаемыми и ожидаемыми значениями

$$\sigma_{\varepsilon}^2 = E(Y - \hat{Y})^2. \quad (7)$$

Принимая во внимание, что ошибки прогноза не коррелированы с независимой переменной X , можно прийти к важному выводу:

$$DY = D[F(X)] + DE \quad (8)$$

или

$$\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_{\varepsilon}^2,$$

т.е. общая дисперсия предсказываемой переменной равна дисперсии предсказываемых значений плюс дисперсия остатков. Это означает, что полная вариация исследуемой зависимой переменной Y складывается из

контролируемой нами вариации функции регрессии и из не поддающейся нашему контролю вариации остаточной, случайной компоненты.

Какая доля общей дисперсии σ_y^2 обуславливается изменчивостью функции регрессии $F(X)$? Ответ на этот вопрос приводит к понятию множественной корреляции R_{xy} :

$$R_{xy}^2 = \frac{\frac{\sigma_{\hat{y}}^2}{2}}{\frac{\sigma_y^2}{2}} = \frac{\sigma_y^2 - \sigma_\varepsilon^2}{\sigma_y^2} = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}. \quad (9)$$

Квадрат коэффициента множественной корреляции (КМК) часто называют коэффициентом детерминации. Если коэффициент детерминации определяет долю дисперсии, которая объясняется изменением $F(X)$, то оставшаяся доля дисперсии $1 - R_{xy}^2$ объясняется воздействием неконтролируемой, случайной остаточной компоненты (помехи).

Установим связь КМК с коэффициентом парной корреляции Пирсона. Для этого рассмотрим ситуацию, когда для заданного значения $X=x$ получено множество значений Y . Их распределение, называемое условным распределением Y при данном $X=x$, есть нормальное одномерное распределение с условным средним

$$\mu_{Y|X} = \mu_Y + \frac{\sigma_{XY}}{\sigma_X} (X - \mu_X) \quad (10)$$

и условной дисперсией

$$\sigma_{Y|X}^2 = \sigma_Y^2 (1 - \rho^2). \quad (11)$$

Легко понять, что дисперсия $\sigma_{Y|X}^2$, обусловленная влиянием помехи, есть не что иное, как остаточная дисперсия σ_ε^2 . Простой расчет приводит к выводу, что в нормальной двумерной схеме множественный коэффициент корреляции совпадает с парным коэффициентом корреляции. Отметим, что приведенная выше формула (11) может быть использована для расчета остаточной дисперсии и при множественной регрессии. В этом случае

$$\sigma_\varepsilon^2 = \sigma_Y^2 (1 - R_{XY}^2). \quad (12)$$

Пример. Что означает фраза «коэффициент корреляции между ростом (X) и весом (Y), оцененный по выборке из 100 человек, равен 0,8»? Спроектируем данные на плоскость признаков РОСТ и ВЕС. Поскольку это совершенно

разные физические единицы, оба признака предварительно пронормируем. У нормированных признаков среднее станет нулевым, а стандартное отклонение – единичным. Если бы взаимосвязь признаков была точно линейной, то все объекты лежали бы точно на прямой $Y = aX$. Однако на самом деле каждому значению роста соответствует не только одно значение веса, а некоторое распределение весов людей, которое имеет ненулевую дисперсию $\sigma_{Y|X}^2 = \sigma_\varepsilon^2$. Коэффициент корреляции дает возможность узнать, какую долю составляет стандартное отклонение этого распределения от стандартного отклонения признака «вес»:

$$\sigma_{Y|X}^2 = \sigma_Y^2(1 - \rho^2) = 1 \cdot (1 - 0.8^2) = 0.36; \quad \sigma_{Y|X} = 0.6.$$

Другая интерпретация значения коэффициента корреляции в данном примере – объяснение различий в значениях признака «вес». Из-за наличия не функциональной, а статистической зависимости дисперсия веса объясняется дисперсией роста только на $r^2 \times 100\% = 64\%$.

Лекция 3. Регрессионный анализ

РЕГРЕССИОННЫЙ АНАЛИЗ

Поставим задачу регрессионного анализа в общем виде: по результатам наблюдений количественных переменных

$$(x_i^{(1)}, \dots, x_i^{(p)}, y_j^{(j)}); \quad i=1, \dots, N; \quad j=1, \dots, m$$

в условиях действия случайных факторов \mathbf{E} найти и оценить статистическую связь между \mathbf{X} и \mathbf{Y} . Решение этой задачи и составляет содержание регрессионного анализа (РА).

Связь между \mathbf{X} и \mathbf{Y} представляется с помощью математической модели, которая называется уравнением регрессии. Как уже отмечалось выше, эффект, вызываемый многими из переменных, настолько слаб, что даже при наличии достоверных данных его статистическая оценка трудна и ненадежна. Поэтому влияние этого эффекта учитывается переменной ε :

$$y = F(\mathbf{X}) + \varepsilon = \hat{Y} + \varepsilon, \quad (1)$$

где y – скалярная результирующая переменная ($m=1$),

$F(\mathbf{X}) = \hat{Y}$ – функция регрессии.

Значок $\hat{}$ будем применять для обозначения предсказанного наиболее вероятного значения случайной величины.

Регрессией в прямом смысле называется функция вида

$$y = b_0 + \sum_{i=1} b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i=1} b_{ij} x_i^2 + \dots + \varepsilon, \quad (2)$$

где b_0, b_i, b_{ij} - коэффициенты регрессии;

ε - невязка (ошибка, отклонение), обусловленная недостаточной пригодностью модели и ошибкой эксперимента. Эти причины обычно являются смешанными.

Регрессия (2) называется множественной. Наибольший практический интерес представляет множественная *линейная* регрессия, для которой $b_{ij} = 0$. В том случае, если исследуется влияние одной переменной на результат эксперимента, то выражение (2) упрощается к виду

$$y = b_0 + b_1 x + \varepsilon. \quad (3)$$

Выражение (3) представляет собой линейную одномерную регрессию. Кроме того, часто используется нелинейное представление регрессии (логарифмическое, экспоненциальное, гиперболическое и другое задаваемое исследователем). В этом случае переменная X подвергается соответствующему нелинейному преобразованию.

В понятие регрессии может быть вложено два смысла.

1. Регрессия в первом смысле: среднее по столбцам (устаревший взгляд).
2. Подбор аналитического выражения для линии регрессии.

Рассмотрим простой пример, иллюстрирующий смысл регрессии. Пусть изучается связь между весом $X^{(1)}$ и ростом $X^{(2)}$ мужчин одного возраста в данном регионе. Возможны три постановки задачи этого исследования:

1. Оценить силу связи, т.е. вычислить $r(x^{(1)}, x^{(2)})$.
2. Узнать, каков наиболее вероятный рост людей данного веса, т.е. найти регрессию $X^{(2)}$ от $X^{(1)}$.
3. Узнать, каков наиболее вероятный вес людей данного роста, т.е. найти регрессию $X^{(1)}$ от $X^{(2)}$.

Получение регрессии включает три этапа:

1. Задается вид уравнения регрессии, включающего, кроме X и Y , несколько неизвестных параметров, которые находятся по опытным данным.
2. Определяются неизвестные коэффициенты, входящие в уравнение регрессии.
3. Проводится статистический анализ регрессии.

Рассмотрим выполнение этих этапов на примере простой линейной регрессии.

ВЫБОР МОДЕЛИ РЕГРЕССИИ

Первый и одновременно один из главных вопросов, который должен решать пользователь, приступая к описанию результирующей переменной Y методом регрессионного анализа – выбор модели (вида функции) регрессии, т.к. собственно регрессионный анализ оценивает лишь параметры этой модели. Основная

информация для выбора модели должна быть получена на этапе качественного решения, и заключаться в отборе независимых переменных, наиболее сильно влияющих на зависимую результирующую переменную, а также в проверке линейности взаимосвязи зависимой и независимой переменных.

Обычно уравнение регрессии задается (постулируется) в виде полинома. Наиболее распространенными и удобными регрессионными моделями можно считать следующие:

- Модель первого порядка по X , линейная относительно коэффициентов:

$$\hat{Y} = \beta_0 + \beta_1 X. \quad (4)$$

- Модель второго порядка по X , линейная относительно коэффициентов:

$$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 X^2. \quad (5)$$

На практике могут использоваться любые другие модели, например:

$$\hat{Y} = \alpha e^{\beta X}; \quad \hat{Y} = \alpha X^\beta; \quad \hat{Y} = \alpha + \beta^* 1/X, \quad (6)$$

в том числе и нелинейные относительно коэффициентов, например:

$$\hat{Y} = \beta_0(x) + \beta_1(x)X. \quad (7)$$

Аппарат линейного регрессионного анализа наиболее разработан, линейные зависимости наиболее просты и понятны. Поэтому стремятся использовать преобразования как X , так и Y (а возможно и обеих этих величин), приводящих изучаемую закономерность именно к линейному виду. Современные статистические программы позволяют получить корреляции и диаграммы рассеяния для разнообразных комбинаций преобразований X и Y , например $(X, \log Y)$, $(\log X, Y)$, $(\log Y, \log X)$, $(\sqrt{X}, \log Y)$, и т.д. Преобразование, для которого получается наибольшее по абсолютной величине значение r_{xy} , будет тем преобразованием, которому соответствует наиболее сильная линейная зависимость.

Проще всего такие преобразования выполняются, если полученная линейная зависимость содержит не более двух постоянных коэффициентов.

Отметим, что неудачно выбранная модель может вызвать лишние затраты времени, хотя и не приводит к принципиальным ошибкам в результате, поскольку адекватность модели проверяется на последнем этапе и плохая модель не используется. Поэтому на этапе выбора вида регрессии надо использовать всю имеющуюся информацию. Как правило, выбор выполняется на основе опыта, интуиции, исходя из содержания задачи, т.е. выбор лежит вне статистических соображений. Например, в регрессионной зависимости *Вес коровы* = $f(\text{параметры коровы})$ была использована формула расчета объема цилиндра, что удачно «легло» на экспериментальные данные.

Предположим, что имеется выборка парных наблюдений $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Существует два способа получения такой выборки:

1. При каждом фиксированном значении аргумента $X=x_i$ ($i=1, \dots, k$) производится несколько наблюдений результирующей переменной $Y: y_{i1}, y_{i2}, \dots, y_{ini}$. При таком подходе только Y является случайной величиной.
2. Из генеральной совокупности случайным образом выбираются N объектов и у каждого из них измеряются переменные X и Y . Здесь случайными являются обе величины X и Y . Преимущество этого метода получения выборки заключается в том, что мы можем сделать статистические выводы относительно коэффициента корреляции между X и Y , в то время как при первом методе этого сделать нельзя.

Независимо от способа получения выборки для определения существования и степени линейной зависимости необходимо выполнить два предварительных шага:

- Анализ диаграммы рассеяния. Допустимо ли предположение о линейной зависимости? Есть ли необходимость в преобразованиях зависимости к линейному виду?
- Вычисление выборочного коэффициента корреляции.

Если предполагается линейная зависимость между X и Y , то теоретическая модель задается уравнением

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i=1, \dots, n \quad (8)$$

где y_i, x_i – значения Y и X в i -м опыте;

β_0, β_1 – неизвестные параметры, “истинные” коэффициенты регрессии, соответствующие модели

$$Y_{уст} = \beta_0 + \beta_1 X \quad (9)$$

ε_i – ошибка, объясняющая отличие наблюдаемого значения от “истинного” и обусловленная воздействием на Y случайных факторов.

Для линейной модели принят ряд гипотез относительно свойств случайного возмущения ε . Эти гипотезы относятся к его среднему, дисперсии и ковариации:

$$E[\varepsilon_i] = 0 \quad i=1, \dots, n \quad \text{– среднее равно нулю;}$$

$$D[\varepsilon_i] = \sigma_\varepsilon^2 \quad 1, \dots, n \quad \text{– дисперсия постоянна и однородна;}$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j) \quad \text{– ошибки не коррелированы.}$$

Если гипотеза о линейной зависимости справедлива и предположения об ошибке ε удовлетворяются, то графическая модель выглядит так:

Для каждого значения $X=x_i$ имеется распределение Y (не обязательно нормальное) со средним значением $\beta_0 + \beta_1 x_i$ и дисперсией $\sigma = 2\varepsilon$, $i=1, \dots, n$.

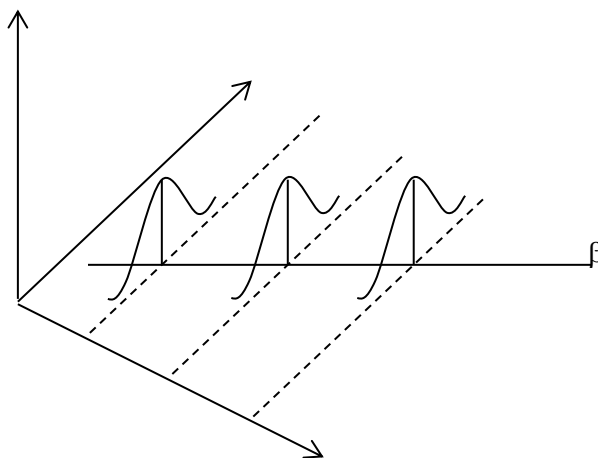


Рис 1. Модель линейной регрессии

ОЦЕНКА ПАРАМЕТРОВ МОДЕЛИ

После подбора типа зависимости по форме кривой оценивают число « k » коэффициентов, которые нужно вычислить. Для линейной модели $k=2$. Для оценки параметров модели можно использовать два метода:

1. **Метод избранных точек.** На кривой зависимости $y=f(x)$ выбирают k точек по числу определяемых коэффициентов. Для сложных кривых эти точки желательно выбрать как на гладких частях кривых, так и в областях экстремума, на перегибах и т. д. Таким образом получают систему k уравнений с k неизвестными, решая которую вычисляют значения коэффициентов.

2. **Метод наименьших квадратов (МНК).** Решение той же задачи по МНК приводит к наиболее точным результатам:

- число пар точек (x_i, y_i) может быть выбрано гораздо больше, чем число коэффициентов
- МНК позволяет вычислить не только значения коэффициентов, но и оценить их значимость, а также адекватность уравнения исходной кривой.

Найдем оценку неизвестных значений β_0 и β_1 методом наименьших квадратов.

Очевидно, что нам никогда не удастся найти β_0 и β_1 , так как для этого потребуются провести бесконечное число опытов. В наших силах лишь оценить эти коэффициенты. Наилучшие оценки $\beta_0=b_0$ и $\beta_1=b_1$ получаются минимизацией соответственно по β_0 и β_1 суммы квадратов отклонений

$$R_{SS} = \sum_{i=1}^N (y_i - y_{iucm})^2 \rightarrow \min \quad (10)$$

R_{SS} есть мера ошибки, возникающей при аппроксимации выборки прямой. Из (10) с учетом (8) и (9) следует другая запись условия МНК

$$R_{SS} = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^N \varepsilon_i^2 \rightarrow \min, \quad (11)$$

позволяющая найти β_0 и β_1 .

Продифференцировав (11) по β_0 и β_1 и приравняв полученные частные производные к нулю, имеем:

$$\begin{aligned} \frac{\partial R_{SS}}{\partial \beta_0} &= 2 \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)(-1) = 0 \\ \frac{\partial R_{SS}}{\partial \beta_1} &= 2 \sum_{i=1}^N (-x_i)(y_i - \beta_0 - \beta_1 x_i) = 0 \end{aligned} \quad (12)$$

Уравнения (12) получили название нормальных, их число в системе всегда равно числу неизвестных параметров модели. Решив систему уравнений и принимая во внимание, что N конечно, найдем оценки коэффициентов регрессионной модели:

$$b_0 = \bar{y} - b_1 \bar{x} \quad (13)$$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (14)$$

Эти оценки не смещены и имеют минимальную дисперсию среди всех несмещенных оценок β_0 и β_1 , линейно зависящих от наблюдений y_1, y_2, \dots, y_N . Оценкой уравнения регрессии (или прямой наименьших квадратов) будет

$$\hat{y} = b_0 + b_1 x \quad (15)$$

Так что оценка значения Y при $X=x_i$ есть $\hat{y}_i = b_0 + b_1 x_i$. Разница между наблюдаемым Y и оцененным значением \hat{Y} при $X=x_i$ называется отклонением или остатком

$$d_i = y_i - \hat{y}_i \quad (16)$$

доставляет минимум сумм квадратов отклонений

$$\hat{R}_{SS} = \sum_{i=1}^N d_i^2 \quad (17)$$

Соотношение между теоретической регрессионной прямой ($N=\infty$) и прямой наименьших квадратов можно увидеть на рис. 2 и 3.

Следует различать такие понятия как

- истинная функция регрессии, отражающая действительную, истинную взаимосвязь;
- теоретическая аппроксимирующая функция регрессии, характеризующая результат, к которому мы бы неограниченно приближались при $N=\infty$;

- выборочная аппроксимирующая функция регрессии (прямая наименьших квадратов).

К сожалению, для практики статистических исследований достаточно типична ситуация, когда исследователь «не угадывает» класс допустимых решений, т.е. истинная функция регрессии не принадлежит к выбранному классу. Другими словами, как бы мы не увеличивали объем выборки, мы не сможем добиться сходимости выборочной аппроксимирующей функции регрессии к истинной функции регрессии.

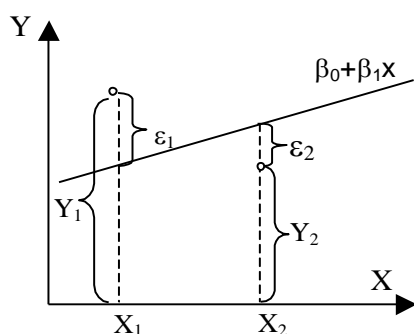


Рис.2. Теоретическая регрессия

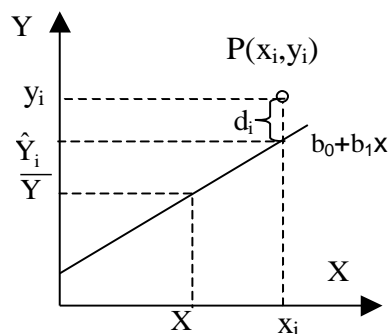


Рис.3. Прямая наименьших квадратов

АНАЛИЗ РЕГРЕССИИ

На этом этапе мы должны ответить на следующие три вопроса:

1. Все ли коэффициенты $\{b_i\}$ полученной регрессии значимо отличны от нуля, т.е. не является ли их отличие от нуля фиктивным, обусловленным естественным разбросом случайной величины Y , по которой они вычислены. Например, надо ли учитывать b_0 на рис. 4а или следует считать, что $b_0 = 0$.
2. Удачно ли выбрана исходная модель, то есть адекватно ли отражает полученные регрессии экспериментальный материал, положенный в ее основу. Например, не имеем ли мы дело с ситуацией, изображенной на рис. 4б, когда отличие \hat{Y} от Y отчетливо заметно на фоне естественного разброса Y .
3. Какова точность полученной регрессии, то есть каков доверительный интервал у каждого из предсказанного ею значения \hat{Y} (заштрихованная область на рис.4а).

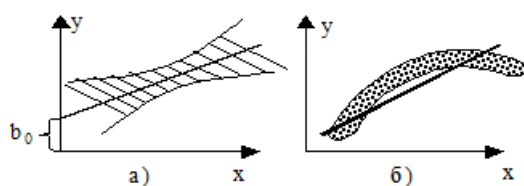


Рис.4. Анализ регрессии

Проверка значимости коэффициентов модели

На этом этапе проверяются гипотезы о значимом отличии от нуля каждого из коэффициентов регрессионной модели.

Для линейной модели

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (18)$$

по очереди проверяется серия гипотез

1) $H_0: \beta_1=0; H_1: \beta_1 \neq 0;$

2) $H_0: \beta_2=0; H_1: \beta_2 \neq 0;$

.....

n) $H_0: \beta_p=0; H_1: \beta_p \neq 0.$

Гипотеза $H_0: \beta_k=0$ для $k=1, \dots, p$ может рассматриваться как гипотеза о том, что переменная X_k не улучшает предсказание Y по сравнению с предсказанием, получаемым с помощью регрессии Y по $(p-1)$ остальным переменным.

Проверку значимости можно осуществлять двумя равноценными способами:

- проверкой по t-критерию Стьюдента;
- построением доверительного интервала.

Прежде всего необходимо найти дисперсию коэффициента регрессии Db_i . Например, для простой линейной модели после несложных преобразований выражений (13,14) можно получить:

$$Db_1 = s_{b_1}^2 = \frac{DY}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad (19)$$

$$Db_0 = s_{b_0}^2 = s_{b_1}^2 \frac{\sum_{i=1}^N x_i^2}{N}, \quad (20)$$

Воспользуемся критерием Стьюдента для проверки значимости

$$t = \frac{b - 0}{s_b}$$

и сравним с критическим значением T-распределения на уровне значимости α при числе степеней свободы $df = N-2$, с которым определяется дисперсия s_Y^2 , а следовательно и s_b^2 . В программах регрессионного анализа величины s_{b_0} и s_{b_1} называются *стандартными ошибками коэффициентов* регрессии. Для каждого β_i стандартная ошибка коэффициента s_{b_i} есть оценка стандартного отклонения b_i от β_i . Если

- $|t| < t_{1-\alpha/2, (df)}$, то отличие от нуля случайно и H_0 не отвергается;

- $|t| < t_{1-\alpha/2, (df)}$, то отличие от нуля следует признать значимым и отвергнуть гипотезу H_0 .

К аналогичным результатам можно прийти, вычислив доверительный интервал для истинных коэффициентов β

$$b - t_{1-\alpha/2, (df)} S_b < \beta < b + t_{1-\alpha/2, (df)} S_b \quad (21)$$

и определив, входит ли в них нулевое значение.

Адекватность модели

Под адекватностью модели простой линейной регрессии подразумевается, что никакая другая модель не дает значимого улучшения в предсказании Y . Например, мы хотим проверить, значимо ли улучшится предсказание Y с помощью полиномиальной регрессии $y_i = \beta_0 + \beta_1 x + \dots + \beta_m x_m + \varepsilon_i$ для $m \geq 2$. Нулевой гипотезой в этом случае будет:

$$H_0: \beta_2 = \dots = \beta_m = 0. \quad (22)$$

Ниже (рис.5, а,б) приведены два рисунка с одинаковым расположением экспериментальных точек, с одинаковым разбросом относительно линии регрессии, но с различным средним разбросом в точках, иначе говорят с различной дисперсией воспроизводимости.

Разброс в точках показан отрезками прямых, составляющими доверительный интервал $\pm 2S$. Модель считается адекватной только в первом случае.

В данном случае разброс в точках такого же порядка, что и разброс относительно линии регрессии. Поэтому можно предположить, что модель пригодна. Во втором случае опыты «слишком» точны, т.е. указывают на более сложную нелинейную модель, в которой точность ее предсказания была бы сравнима с точностью эксперимента.

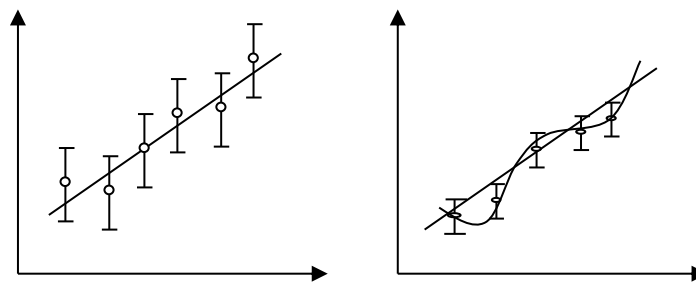


Рис.5. Проверка адекватности

Разброс опытных значений Y относительно линии регрессии Y можно оценить по остаточной сумме квадратов R_{SS} . Неудобство использования R_{SS} состоит в том, что R_{SS} зависит от числа коэффициентов в уравнении. Введите столько коэффициентов, сколько вы провели независимых опытов, и получите остаточную сумму, равную нулю. Поэтому предпочитают относить на один «свободный» опыт. Остаток суммы квадратов, деленный на число степеней свободы, называется остаточной дисперсией, или дисперсией адекватности.

$$S_{\text{АД}}^2 = \frac{R_{\text{SS}}}{df_{\text{АД}}}, \quad (23)$$

где $df=N-p$ – число степеней свободы;

$p=2$ – количество коэффициентов линейной модели.

Предположим, что имеется k различных значений для X , например x_1, \dots, x_k . Для каждого из этих x_i имеется n_i наблюдений $y_{i1}, y_{i2}, \dots, y_{ini}$ переменной Y , $i=1, \dots, k$. Пусть $n_i > 1$ и $\sum n_i = N$. Тогда модель линейной регрессии может быть записана в следующем виде:

$$Y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \quad j=1, \dots, n_i, \quad i=1, \dots, k, \quad (24)$$

а разброс в i -ой точке находим по известной формуле для несмещенной дисперсии:

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2. \quad (25)$$

Полученную группу выборочных дисперсий S_i^2 можно считать оценками для одной и той же генеральной дисперсии σ^2 .

Наилучшая оценка этой дисперсии имеет вид

$$S_{\text{ВОС}}^2 = \frac{1}{k} \sum_{i=1}^k S_i^2 = \frac{1}{k} \sum_{i=1}^k \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (26)$$

и называется дисперсией воспроизводимости.

Если число опытов в каждой точке одинаково, т.е. $n_1=n_2=n_k=m$, то

$$S_{\text{ВОС}}^2 = \frac{1}{k(m-1)} \sum_{i=1}^k \sum_{j=1}^m (y_{ij} - \bar{y}_i)^2 \quad (27)$$

с числом степеней свободы $df_{\text{вос}}=k(m-1)$.

Для проверки гипотезы об однородности разброса в точках и разброса относительно линии регрессии

$$H_0: S_{\text{АД}}^2 = S_{\text{ВОС}}^2; \quad H_1: S_{\text{АД}}^2 > S_{\text{ВОС}}^2 \quad (28)$$

используется F – критерий:

$$F = \frac{S_{\text{АД}}^2}{S_{\text{ВОС}}^2} \quad (29)$$

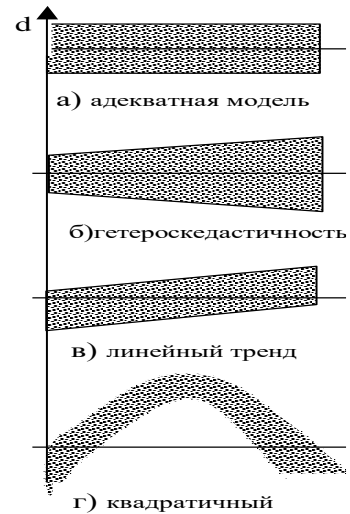


Рис.6. Примеры графиков остатков

Если $F < F_{\alpha, df_{ад}, df_{вос}}$, то модель можно считать адекватной. Если гипотеза об адекватности отвергается, необходимо переходить к более сложной форме математического описания, либо, если это касается планирования активного эксперимента, проводить эксперимент с меньшим интервалом варьирования.

Исследование регрессионных остатков

Интересную статистическую информацию может дать анализ остатков

$$d_i = y_i - \hat{y}_i, i=1, \dots, N$$

Для проверки адекватности модели можно использовать график d_i в зависимости от x_i или \hat{y}_i . Рассмотрим типичные графики остатков:

1. Если остатки попадают в горизонтальную полосу с центром на оси абсцисс, модель можно рассматривать как адекватную (рис. 6а).
2. Если полоса расширяется, когда X или Y возрастают, это указывает на гетероскедастичность, т.е. на отсутствие постоянства дисперсии. В частности, стандартное отклонение может быть функцией $\beta_0 + \beta_1 x$, что делает необходимым преобразование Y (рис. 6б).
3. Если остатки представляются линейным трендом, в модель необходимо ввести независимую дополнительную переменную (рис. 6в).
4. График вида, представленного на рис 6г, указывает, что в модель должен быть добавлен квадратичный член.

Оценка точности регрессии

Точность регрессии задается доверительным интервалом, внутри которого предсказанное среднее значение $\hat{Y}_{ист}$ располагается с вероятностью $1 - \alpha$. Определим дисперсию среднего значения $D\hat{Y}$.

$$D\hat{Y} = \left[\frac{1}{N} + \frac{(x - \bar{x})^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \right] DY \quad (30)$$

Располагая $D\hat{Y}$ или ее оценкой $s^2\hat{Y}$, определенной в соответствии с (30), найдем доверительный интервал

$$\hat{Y} - t_{1-\alpha/2, (df)} S\hat{Y} < \hat{Y}_{ист} < \hat{Y} + t_{1-\alpha/2, (df)} S\hat{Y} \quad (31)$$

Из (31) следует, что наибольшая точность предсказаний соответствует центральной части диапазона изменения X ; с удалением от среднего значения X точность предсказаний падает.

Лекция 4. Метод главных компонент

ВВЕДЕНИЕ

С данными бывают две проблемы: либо их слишком мало, либо их слишком много. Сегодня мы поговорим о второй проблеме.

Есть такая область исследования — определение авторства текстов. Допустим, у нас есть массив текстов, автор которых неизвестен. Может быть, эти тексты принадлежат одному и тому же человеку, может быть, разным. Может быть, мы догадываемся о том, кто является автором, а может быть и нет. Если бы эти тексты были написаны ручкой на бумаге, мы могли бы ответить на какие-то вопросы об авторстве, сравнивая почерки. Но сейчас перед нами чаще «голый» текст, не содержащий таких естественных индивидуальных признаков, как почерк.

Тем не менее, разумно предположить, что эти признаки всё-таки есть. Одни люди пишут длинными предложениями, другие короткими. Одни используют много разных слов, другие ограничиваются небольшим запасом. Одни используют много глаголов, другие мало. У каждого автора есть свой стиль и этот стиль можно извлечь из текста. Как это сделать?

Для каждого текста можно вычислить огромное количество параметров. Среднюю длину предложения. Разброс длин предложений. Распределение слов. Процент существительных, прилагательных, глаголов. И ещё кучу всего. В общем, можно превратить каждый текст в длинный-длинный вектор. Взяв много текстов, можно записать их параметры в широкую табличку.

Но дальше возникает проблема: что, собственно говоря, делать с этим массивом данных? Как его обрабатывать? Как находить связи между разными параметрами? **Как их визуализировать?**

Если у нас есть один параметр, можно нарисовать для него гистограмму. Если параметра два, можно нарисовать диаграмму рассеяния (scatter plot), показывающую, как распределён каждый из них и как они связаны между собой. С большим количеством параметров так сделать нельзя: пространство, в котором можно было бы нарисовать соответствующие картинки, имеет высокую размерность и не помещается на двумерной бумаге или экране.

Что же делать? Есть разные подходы. Например, можно рассмотреть всевозможные пары параметров и для каждой нарисовать свою диаграмму рассеяния. Таким образом, однако, удастся визуализировать только попарные связи между параметрами, а этого зачастую недостаточно. К тому же этот метод приводит к появлению очень большого количества картинок, если число параметров велико.

Другой подход строится в предположении, что наш набор параметров избыточен и для описания наиболее важных свойств текстов достаточно всего нескольких чисел. Это кажется разумным: если среди наших многочисленных параметров есть такие, которые находятся в сильной связи друг с другом (а это вполне естественное предположение), то часть из них можно отбросить. Например, если один из параметров выражается через другие, то его можно просто выкинуть без потери информации.

Но как выяснить, какой именно набор параметров хорошо описывает наш набор данных, но при этом имеет небольшую избыточность? Иными словами, ***как уменьшить размерность пространства, в котором живут данные, потеряв при этом минимум информации?***

Способы решения этой задачи, называются **методами уменьшения размерности** (dimensionality reduction).

Сущность проблемы

Общее число признаков $X^{(1)}, X^{(2)}, \dots, X^{(p)}$, регистрируемых на каждом из множества обследуемых объектов очень велико – порядка ста и более. Тем не менее имеющиеся многомерные наблюдения следует подвергнуть статистической обработке, осмыслить либо внести в БД для того чтобы иметь возможность их использовать в нужный момент. Суть проблемы состоит в том, чтобы представить каждое из наблюдений в виде вектора Z некоторых вспомогательных показателей $Z^{(1)}, Z^{(2)}, \dots, Z^{(u)}$ с существенно меньшим чем числом компонент ($u < p$)

От $X^{(1)}, X^{(2)}, \dots, X^{(p)}$ к $Z^{(1)}, Z^{(2)}, \dots, Z^{(u)}$

Зачем?

- Необходимость наглядного представления исходных данных (визуализация), что достигается их проецированием на специально подобранное трехмерное пространство ($u=3$), плоскость ($u=2$), или числовую прямую;
- Стремление к лаконизму исследуемых моделей, обусловленному необходимостью упрощения счета и интерпретации полученных статистических выводов;
- Необходимость существенного сжатия объемов хранимой статистической информации (без видимых потерь в ее информативности)

Предпосылки перехода

Существуют следующие основания говорить о возможности перехода от большого числа p исходных показателей анализируемой системы к существенно меньшему числу u наиболее информативных переменных

- Дублирование информации от сильно взаимосвязанных признаков
- Неинформативность признаков, мало меняющихся при переходе от одного объекта к другому

- Возможность агрегирования, то есть простого или взвешенного суммирования, по некоторым признакам

Методы снижения размерности

К основным методам снижения размерности относятся метод главных компонент и факторный анализ.

Другие способы уменьшения размерности данных — это

- метод независимых компонент,
- многомерное шкалирование, а также многочисленные нелинейные обобщения:
 - метод главных кривых и многообразий,
 - поиск наилучшей проекции (англ. Projection Pursuit),
 - нейросетевые методы «узкого горла»,
 - самоорганизующиеся карты Кохонена и др.

МЕТОД ГЛАВНЫХ КОМПОНЕНТ

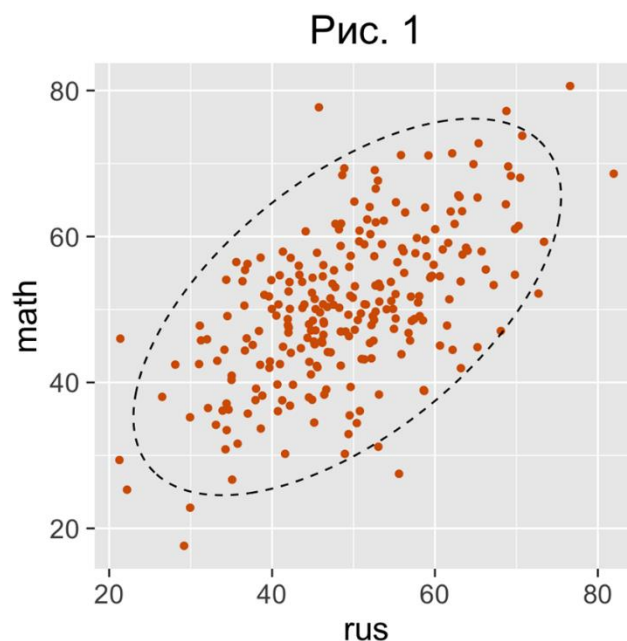
Метод главных компонент (principal component analysis, PCA) — один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Изобретён Карлом Пирсоном в 1901 году. Применяется во многих областях, в том числе, в эконометрике, биоинформатике, обработке изображений, для сжатия данных, в общественных науках.

Вычисление главных компонент может быть сведено к вычислению сингулярного разложения матрицы данных или к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных. Иногда метод главных компонент называют преобразованием Кархунена — Лозва или преобразованием Хотеллинга.

Пример «Школьные оценки»

У нас есть табличка с результатами теста для школьников по двум предметам — например, по русскому языку и математике. Можно нарисовать вот такую картинку.

	rus	math
1	38.62	33.67
2	46.22	54.53
3	46.40	38.32
4	53.17	51.07
5	62.86	65.64



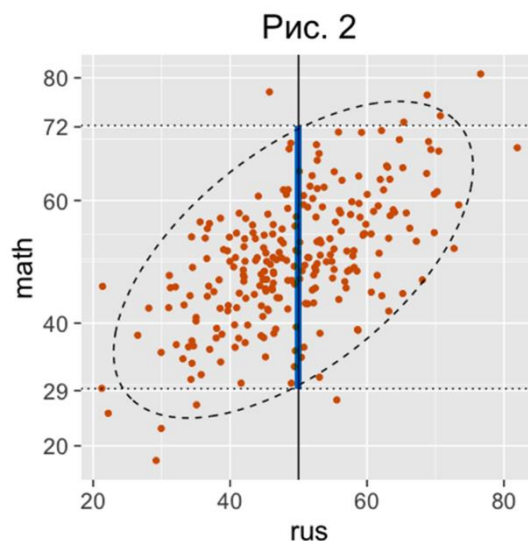
Мы видим, что оценки по этим двум предметам скоррелированы — среди школьников, получающих высокие баллы по математике, много тех, кто также получает высокие баллы по русскому языку, и наоборот (это согласуется с нашей интуицией). Есть и исключения, лежащие вне эллипса — *выбросы*. Для простоты изложения мы сейчас будем пренебрегать выбросами, а также считать, что если бы мы взяли побольше школьников, то соответствующие им точки практически заполнили бы весь эллипс.

Это наши исходные данные. Теперь предположим, что нам надо уменьшить размерность — вместо двух чисел на каждого школьника хранить только одно число. Например, мы выбираем, что записать в аттестат, и по закону это должно быть только одно число, а не два. И мы хотим в этом единственном числе закодировать как можно больше информации о школьнике. Как в этом случае поступить?

Можно просто отбросить одну из переменных и оставить другую. Например, можно записывать в аттестат только оценку по русскому языку, а оценку по математике игнорировать.

С одной стороны, это не такая уж и плохая стратегия. Как мы обсуждали выше, оценки скоррелированы, и человек, сдавший русский на высокий балл, скорее всего не имеет проблем и с математикой, и наоборот.

С другой стороны, какая-то информация всё-таки теряется. Представьте себе, что мы знаем о школьнике только тот факт, что он сдал русский на 50 баллов. Что это говорит о математических способностях школьника? Если не знать настоящих данных, а смотреть только на эллипс на картинке, то мы увидим, что соответствующая точка может находиться в любом месте синего отрезка — его оценка по математике в этом случае колеблется примерно между 29 и 72 — разброс очень большой.



Конечно, понятно, что, заменяя два числа одним, мы потеряем какую-то информацию, но хотелось бы всё-таки эти потери минимизировать.

Можем ли мы сделать что-то лучшее, чем сообщать только одну оценку из двух? Оказывается, можем.

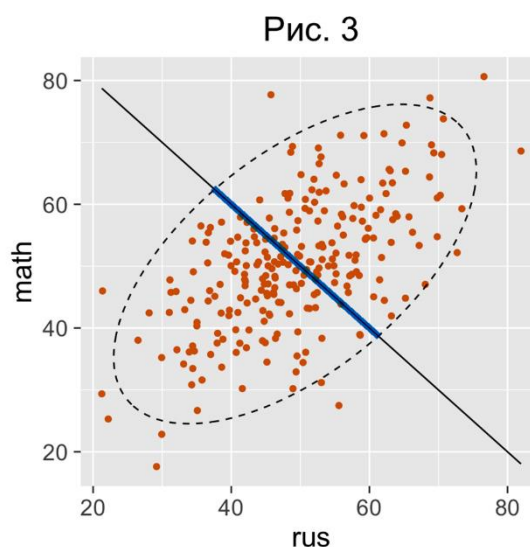
Мы можем сконструировать новое число из двух имеющихся!

Давайте рассмотрим самый простой вариант: впишем в аттестат сумму оценок за русский язык и математику. Иными словами, мы введём новую переменную — обозначим её через **PC1** (почему так — будет ясно позднее), которая связана с нашими старыми переменными таким образом:

$$\mathbf{PC1 = rus + math}$$

Посмотрим, насколько этот метод лучше. Пусть мы знаем, что для некоторого школьника $PC1=100$. Что тогда можно сказать про его оценки?

Проведём на графике прямую, соответствующую условию $PC1=100$, то есть $rus+math=100$. Она пройдёт из левого верхнего угла в правый нижний и пересечёт наш эллипс по некоторому отрезку.

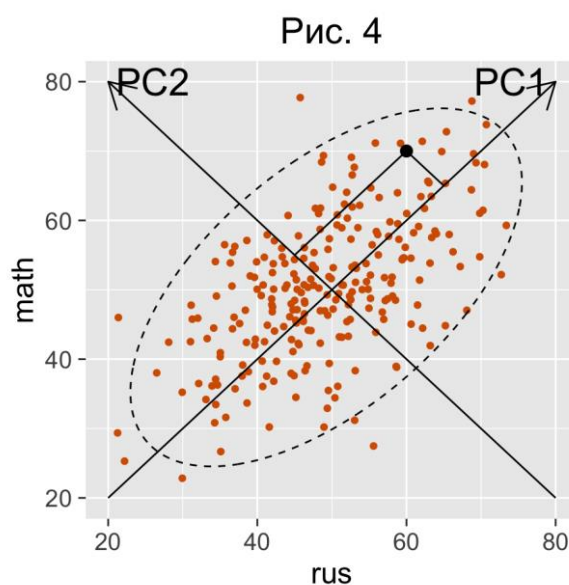


Как и в прошлый раз, наш школьник мог бы оказаться в любой точке этого отрезка: на этот раз мы не знаем наверняка ни оценку по математике (она колеблется где-то между 39 и 63), ни оценку по русскому языку (она между 37 и 61).

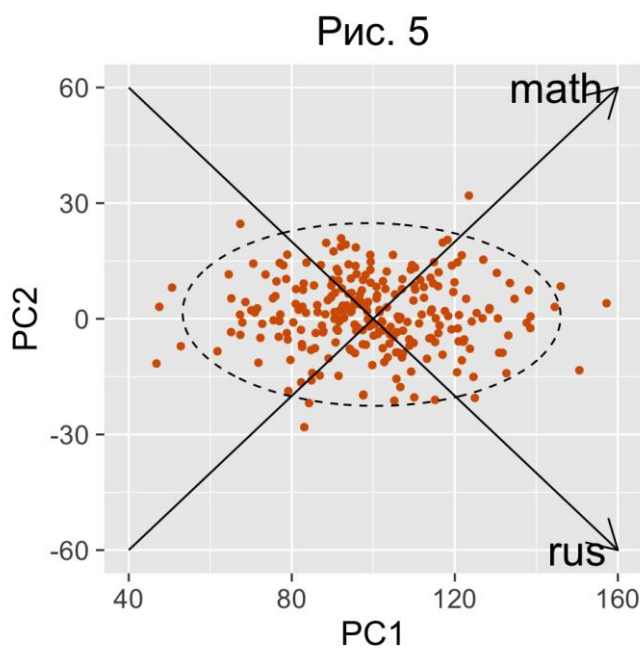
Тем не менее, если измерять степень нашего незнания длиной того самого отрезка, на котором может оказаться школьник, то мы видим, что она уменьшилась: новый отрезок короче старого, потому что сейчас мы пересекаем эллипс «поперек», а раньше пересекали «наискосок». Поэтому сообщать наше число PC1 лучше, чем сообщать только одну из оценок (если, конечно, мы не знаем заранее, что получателью этой информации какая-то из двух оценок важнее другой).

Метод главных компонент — это история про введение новой, более экономной системы координат, в которой описывать наши данные проще. Вот как эта система координат будет устроена в нашем примере с оценками.

В качестве первой координаты точки мы возьмём PC1, то есть сумму её старых координат, а в качестве второй координаты (обозначим её через PC2) возьмём разность её старых координат: $PC2 = \text{math} - \text{rus}$



Новая система координат



Аппроксимация данных линейными многообразиями

Метод главных компонент начинался с задачи наилучшей аппроксимации конечного множества точек прямыми и плоскостями (Пирсон, 1901). Дано конечное множество векторов $x_1, x_2, \dots, x_m \in \mathbb{R}^n$, для каждого $k = 0, 1, \dots, n-1$ среди всех k -мерных линейных многообразий в \mathbb{R}^n найти такое множество линейных комбинаций $L_k \subset \mathbb{R}^n$, что сумма квадратов уклонений x_i от L_k минимальна:

$$\sum_{i=1}^m \text{dist}^2(x_i, L_k) \rightarrow \min$$

где $\text{dist}(x_i, L_k)$ — евклидово расстояние от точки до линейного многообразия.

Три вида проекций

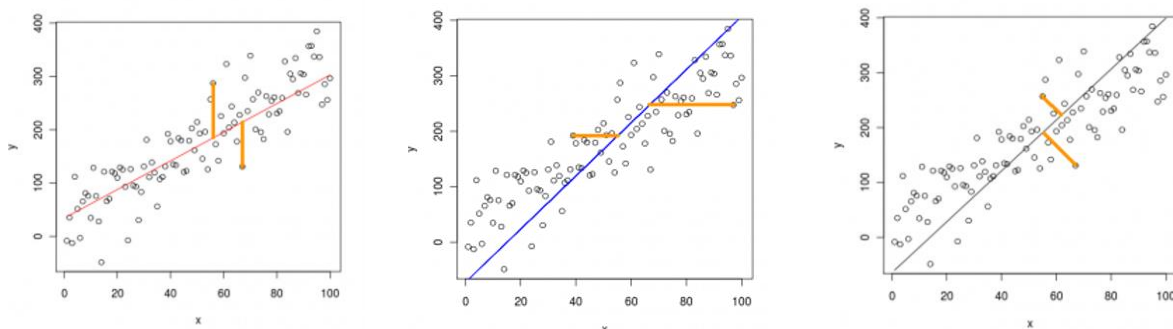
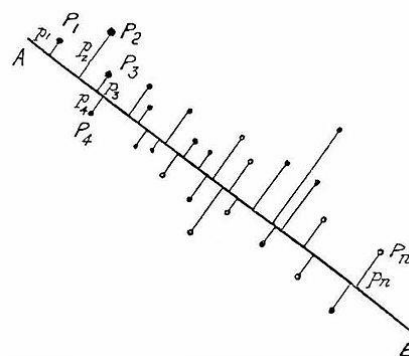


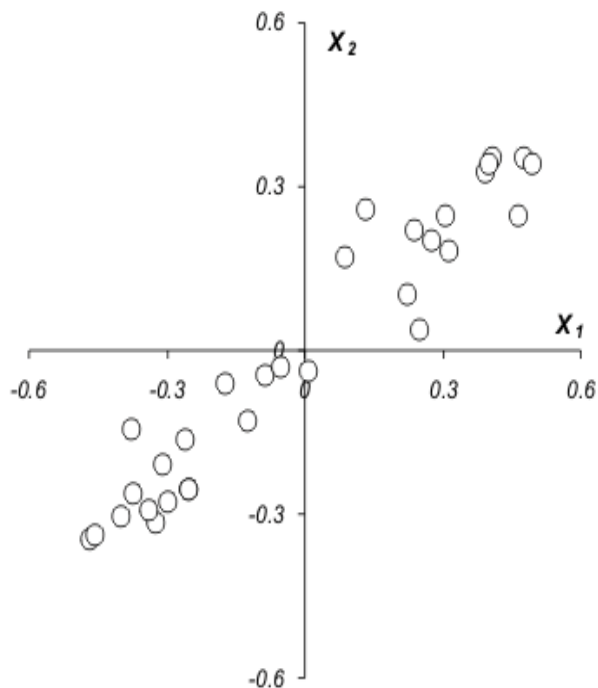
Иллюстрация к знаменитой работе Пирсона (1901): даны точки P_i на плоскости, p_i — расстояние от P_i до прямой AB . Ищется прямая AB , минимизирующая сумму $\sum_i p_i^2$



Интуитивный подход

Постараемся передать суть метода главных компонент, используя интуитивно-понятную геометрическую интерпретацию. Начнем с простейшего случая, когда имеются только две переменные x_1 и x_2 . Такие данные легко изобразить на плоскости.

	X_1	X_2
1	0.407	0.353
2	0.475	0.355
3	-0.088	-0.045
4	0.394	0.325
5	0.274	0.202
6	0.131	0.258
7	-0.053	-0.031
8	-0.124	-0.128
9	-0.469	-0.344
10	0.088	0.171
11	-0.261	-0.162
12	0.401	0.341
13	-0.376	-0.143
14	-0.251	-0.255
15	-0.325	-0.316
16	0.464	0.248
17	-0.310	-0.207
18	0.307	0.247
19	-0.399	-0.303
20	-0.253	-0.253
21	-0.341	-0.291



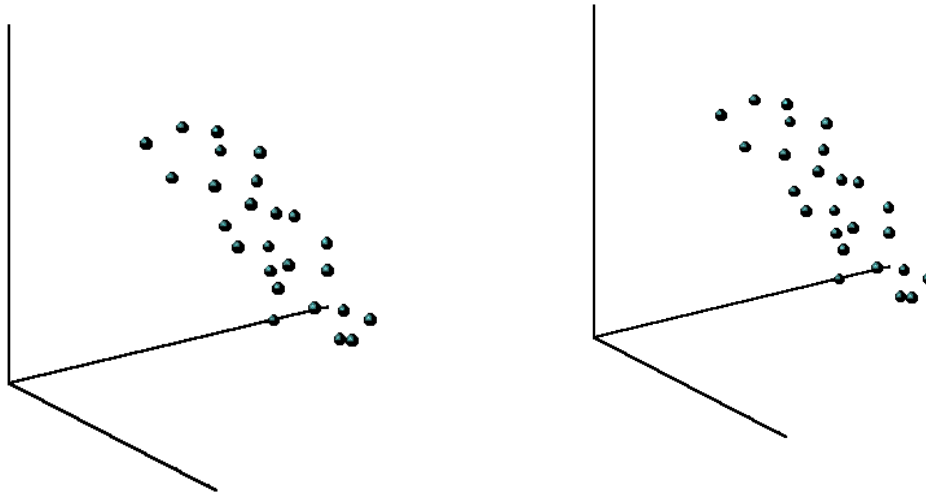
Каждой строке исходной таблицы (т.е. образцу) соответствует точка на плоскости с соответствующими координатами. Они обозначены **пустыми** кружками. Проведем через них прямую так, чтобы вдоль нее происходило максимальное изменение данных. На рисунке эта прямая выделена синим цветом; она называется **первой главной компонентой – PC1**. Затем спроецируем все исходные точки на эту ось. Получившиеся точки закрасены **красным** цветом. Теперь мы можем предположить, что на самом деле все наши экспериментальные точки и должны были лежать на этой новой оси. Просто какие-то неведомые силы отклонили их от правильного, идеального положения, а мы вернули их на место. Тогда все отклонения от новой оси можно считать шумом, т.е. ненужной нам информацией.

Правда, мы должны быть в этом уверены. Проверить шум ли это, или все еще важная часть данных, можно поступив с этими остатками так же, как мы поступили с исходными данными – найти в них ось максимальных изменений. Она называется **второй главной компонентой (PC2)**. И так надо действовать, до тех пор, пока шум уже не станет действительно шумом, т.е. случайным хаотическим набором величин.

В общем, многомерном случае, процесс выделения главных компонент происходит так:

- Ищется центр облака данных, и туда переносится новое начало координат – это нулевая главная компонента (PC0)
- Выбирается направление максимального изменения данных – это первая главная компонента (PC1)

- Если данные описаны не полностью (шум велик), то выбирается еще одно направление (PC2) – перпендикулярное к первому, так чтобы описать оставшееся изменение в данных и т.д.



метод главных компонент – как проекция на подпространство

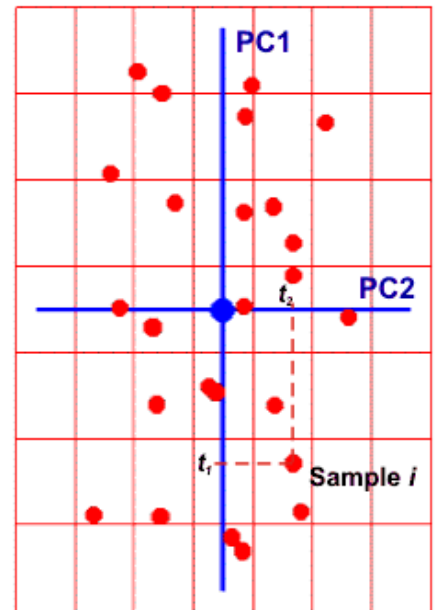
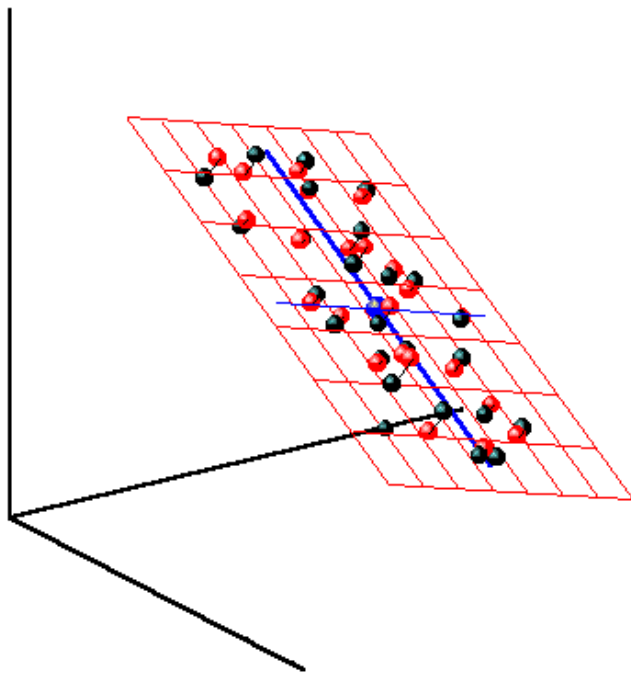
В результате мы переходим от большого количества переменных к новому представлению, размерность которого значительно меньше. Часто удается упростить данные на порядки: от 1000 переменных перейти всего к двум. При этом ничего не выбрасывается – все переменные учитываются. В то же время несущественная для сути дела часть данных отделяется, превращается в шум.

Найденные главные компоненты и дают нам искомые скрытые переменные, управляющие устройством данных.

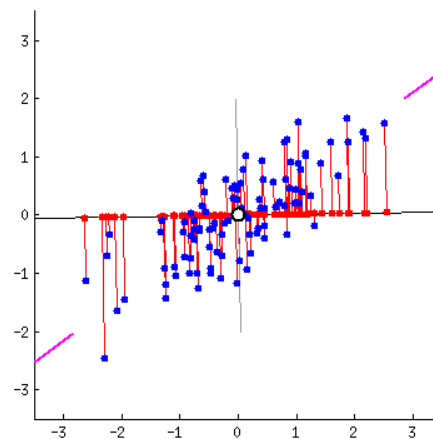
PCA найдет «лучшую» линию в соответствии с двумя разными критериями того, что является «лучшим».

Во-первых, изменение значений вдоль этой линии должно быть максимальным. Обратите внимание на то, как изменяется «разброс» (мы называем это «дисперсией») красных точек, когда линия вращается; видите ли вы, когда он достигает максимума?

Во-вторых, если мы восстановим исходные две характеристики (положение синей точки) от новой (положение красной точки), ошибка восстановления будет определяться длиной соединительной красной линии. Наблюдайте, как изменяется длина этих красных линий во время вращения линии; видите ли вы, когда общая длина достигает минимума?



Вот как выглядят эти проекции для разных линий (красные точки - это проекции синих точек):



Второй главный компонент рассчитывается таким же образом, при условии, что он не коррелирован (то есть перпендикулярен) первому главному компоненту и учитывает следующую по величине дисперсию. Это продолжается до тех пор, пока не будет вычислено p главных компонент, равное исходному количеству переменных.

С геометрической точки зрения, главные компоненты представляют собой Векторы данных, которые объясняют максимальное количество отклонений. Главные

компоненты – новые оси, которые обеспечивают лучший угол для оценки данных, чтобы различия между наблюдениями были лучше видны.

Поскольку существует столько главных компонент, сколько переменных в наборе, главные компоненты строятся таким образом, что первый из них учитывает наибольшую возможную дисперсию в наборе данных.

Постановка задачи

Линейное преобразование как переход к новой системе координат

Таблицу экспериментальных данных (ТЭД) геометрически можно интерпретировать как облако экспериментальных данных в p -размерном пространстве, где p – количество признаков, свойств объектов. Каждая точка такого пространства представляет собой один объект исследования.

- Из векторной алгебры хорошо известно, что точка (вектор) в заданной системе координат может быть представлена и в другой системе координат (в другом базисе), причем количество этих базисов бесконечно.
- Новая система координат Y_1, Y_2, \dots, Y_p может быть выражена через старые координаты X_1, X_2, \dots, X_p , как их линейная комбинация.

Геометрия PCA

Каждый исходный образец x_i (строка в матрице X) можно представить как вектор в p -мерном пространстве с координатами

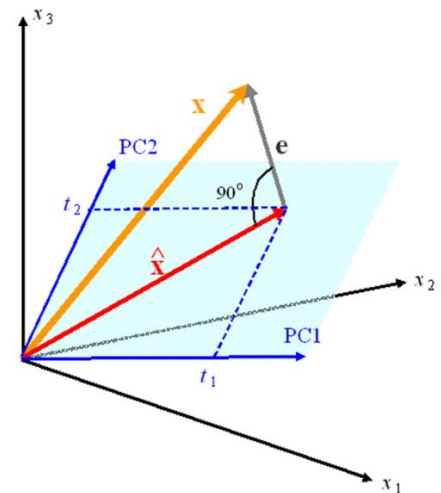
$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$

PCA проецирует его в вектор, лежащий в пространстве главных компонент,

$$\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{iA})$$

размерностью A . В исходном пространстве этот же вектор \mathbf{t}_i имеет координаты

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$$



Постановка задачи МГК

Дано:

1. Случайные переменные x_1, \dots, x_p
2. Вектор средних $m = (m_1, \dots, m_p)'$
3. Ковариационная матрица $\Sigma^{p \times p} = (\sigma_{ij})$

$$X = (\bar{X}_1 \bar{X}_2 \dots \bar{X}_p) = \begin{array}{l|cccc} & X_1 & X_2 & \dots & X_p \\ \text{объект } 1 & X_{11} & X_{12} & \dots & X_{1p} \\ \text{объект } 2 & X_{21} & X_{22} & \dots & X_{2p} \\ \dots & \cdot & \cdot & \cdot & \cdot \\ \text{объект } n & X_{n1} & X_{n2} & \dots & X_{np} \end{array}$$

Задание

Найти такие линейные комбинации исходных переменных:

$$\begin{aligned} Y_1 &= \sum_{j=1}^p \alpha_{1j} X_j \\ &\dots \\ Y_p &= \sum_{j=1}^p \alpha_{pj} X_j \end{aligned} \tag{1}$$

что

$$1. \text{cov}(Y_i Y_j) = 0 \quad i, j = 1, 2, \dots, p, \quad i \neq j \tag{2}$$

$$2. D(Y_1) > D(Y_2) > \dots > D(Y_p) \tag{3}$$

$$3. \sum_{i=1}^p D(Y_i) = \sum_{i=1}^p \sigma_{ii} = D_0 \tag{4}$$

Из формул видно, что

1. Переменные Y_1, \dots, Y_p не коррелированы (2) и упорядочены по возрастанию дисперсии (3).
2. Общая дисперсия D_0 после преобразования остается без изменений (4).

Ковариационная матрица представляет собой симметричную матрицу размера $p \times p$ (где p – количество измерений), где в качестве ячеек пребывают коэффициенты ковариации, связанные со всеми возможными парами исходных переменных.

Поскольку ковариация переменной с самой собой – это ее дисперсия, на главной диагонали (от верхней левой ячейки к нижней правой), у нас фактически есть дисперсии каждой исходной переменной. А поскольку ковариация коммутативна (в ячейке XY значение равно YX), элементы матрицы симметричны относительно главной диагонали.

Матричная запись

Уравнение (1) можно записать в матричной форме:

$$Y = \begin{pmatrix} Y_1 \\ \cdot \\ \cdot \\ Y_p \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \dots & \alpha_{1p} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \alpha_{p1} & \dots & \alpha_{pp} \end{pmatrix} \begin{pmatrix} X_1 \\ \cdot \\ \cdot \\ X_p \end{pmatrix} = AX, \text{ где } A - \text{ матрица перехода}$$

Вычисление ГК

Первой главной компонентой Y_1 исследуемой системы показателей $X = (X_1, \dots, X_p)$ называется такая нормировано-центрированная линейная комбинация этих показателей, которая среди всех других обладает наибольшей дисперсией. Из уравнения имеем

$$Y_1 = \alpha_{11}\bar{X}_1 + \alpha_{12}\bar{X}_2 + \dots + \alpha_{1p}\bar{X}_p = \alpha_{11} \begin{pmatrix} X_{11} \\ \cdot \\ \cdot \\ X_{n1} \end{pmatrix} + \dots + \alpha_{1p} \begin{pmatrix} X_{1p} \\ \cdot \\ \cdot \\ X_{np} \end{pmatrix} = \begin{pmatrix} \alpha_{11}X_{11} + \dots + \alpha_{1p}X_{1p} \\ \cdot \\ \cdot \\ \alpha_{11}X_{n1} + \dots + \alpha_{1p}X_{np} \end{pmatrix} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}$$

Тогда имеем

$$Y_1 = \alpha_{11}X_1 + \dots + \alpha_{1p}X_p$$

Требуется найти такие $\alpha_{11}, \dots, \alpha_{1p}$, чтобы величина

$$DY_1 = \sum_{i=1}^p \sum_{j=1}^p \alpha_{1i} \alpha_{1j} \sigma_{ij}$$

была максимальной при

$$\sum_{j=1}^p \alpha_{1j} = 1$$

Это условие обеспечивает единственность решения.

Решение

$$a_1 = (\alpha_{11}, \dots, \alpha_{1p})'$$

называется собственным вектором и соответствует максимальному собственному значению матрицы. Это собственное значение равно дисперсии DY_1 . Линейная комбинация (5) называется главной компонентой переменных x_1, \dots, x_p . Она объясняет

$$\frac{DY_1}{D_0} \% \text{ общей дисперсии.}$$

Аналогично можно показать, что

$$Y_k = \alpha_k X$$

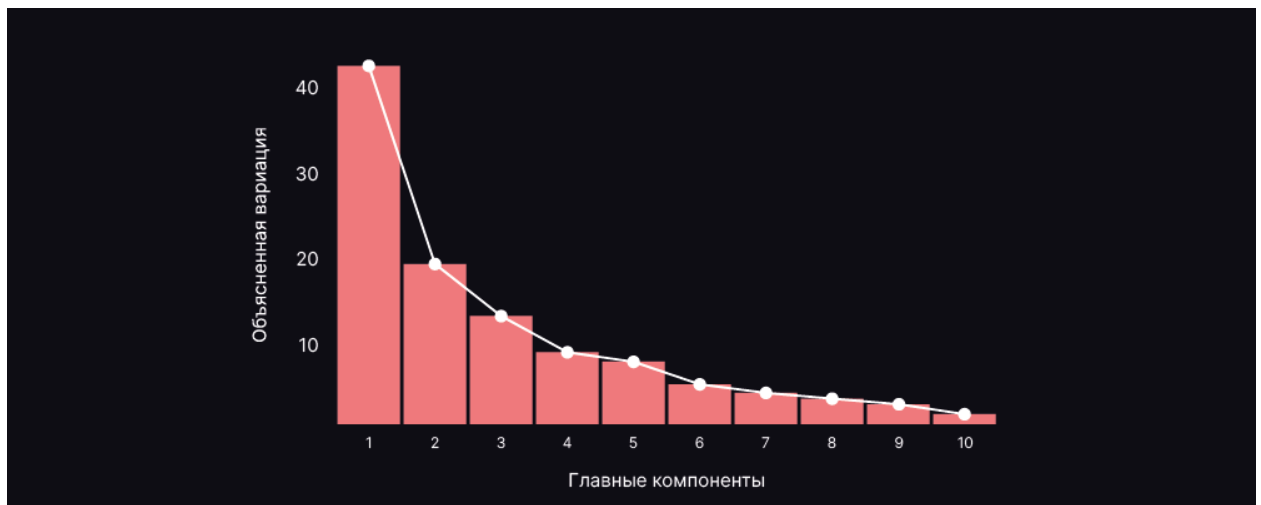
где α_k – собственный вектор матрицы Σ , соответствующей k-тому по величине собственному значению λ_k этой матрицы.

Переменные Y_1, \dots, Y_k будут объяснять

$$100 \cdot (DY_1 + \dots + DY_k) / D_0 \%$$

общей дисперсии.

Главная компонента – это новая переменная, смесь исходных. Эти комбинации выполняются таким образом, что новые переменные (то есть главные компоненты) не коррелированы, и большая часть информации в исходных переменных помещается в первых компонентах. Итак, идея состоит в том, что 10-мерный датасет дает нам 10 главных компонент, но PCA пытается поместить максимум возможной информации в первый, затем максимум оставшейся информации во второй и так далее, пока не появится что-то вроде того, что показано на графике ниже:



Такая организация информации в главных компонентах позволит нам уменьшить размерность без потери большого количества информации за счет отбрасывания компонент с низкой информативностью.

Здесь важно понимать, что главные компоненты менее интерпретируемы и не имеют никакого реального значения, поскольку они построены как линейные комбинации исходных переменных.

Матрицу ковариации для нашей выборки X можно представить в виде произведения $X^T X$. Из отношения Релея вытекает, что максимальная вариация нашего набора данных будет достигаться вдоль собственного вектора этой матрицы, соответствующего максимальному собственному значению. Таким образом главные

компоненты, на которые мы бы хотели спроецировать наши данные, являются просто собственными векторами соответствующих топ-к штук собственных значений этой матрицы.

Дальнейшие шаги просты — надо просто умножить нашу матрицу данных на эти компоненты и мы получим проекцию наших данных в ортогональном базисе этих компонент. Теперь если мы транспонируем нашу матрицу данных и матрицу векторов главных компонент, мы восстановим исходную выборку в том пространстве, из которого мы делали проекцию на компоненты. Если количество компонент было меньше размерности исходного пространства, мы потеряем часть информации при таком преобразовании.

Таким образом, соотношение для определения всех “р” ГК вектора X может быть представлено в виде

$$Y = AX,$$

где $Y = (Y_1, \dots, Y_p)^T$, $X = (X_1, \dots, X_p)^T$, а матрица A состоит из строк $\alpha_j = (\alpha_{j1}, \dots, \alpha_{jp})$, $j = 1, 2, \dots, p$, являющихся собственными векторами матрицы Σ , соответствующим собственным числам λ_j . При этом сама матрица A по построению является ортогональной, т.е.

$$AA^T = A^T A = I$$

Коэффициент α_{ij} показывает, какой вклад вносит старый признак $X^{(i)}$ в i-тую главную компоненту. Поэтому матрица $A = \{\alpha_{ij}\}$ называется матрицей нагрузок на ГК.

Для визуального анализа данных часто используют проекции исходных векторов на плоскость первых двух главных компонент. Обычно хорошо видна структура данных, выделяются компактные кластеры объектов и отдельно выделяющиеся вектора.

Когда матрица ковариации неизвестна, она оценивается вторичной ковариационной матрицей S. Для получения оценок ГК следует применить описанную выше процедуру к матрице S. В результате получаются оценки a_{ij} коэффициентов α_{ij} , $i, j = 1, \dots, p$.

Для получения ГК можно использовать вместо ковариационной матрицы корреляционную. Обычно, когда “р” переменных измеряются в различных единицах, не имеющих между собой ничего общего, линейные комбинации переменных бывает трудно интерпретировать. В этом случае переходят к z-оценкам.

$$z_i = (x_i - \mu) / \sigma_i \text{ или } z_i = (x_i - \bar{x}) / S_i \quad i = 1, \dots, p$$

При этом общая дисперсия D_0 равняется числу переменных.

Корреляция между переменной X_i и главной компонентой Y_i задается величиной

$$\text{corr}(X_i, Y_i) = \alpha_{ij} \frac{S_j}{S_i}$$

где S_i, S_j оценки стандартных отклонений X_i и Y_j . Следовательно, для сравнения вкладов переменных x_1, \dots, x_p в Y_j следует сравнить величины $\alpha_{ij}/s_i, i = 1, \dots, p$.

Когда известна корреляционная матрица, достаточно сравнить коэффициенты α_{ij} . В этом случае самый большой коэффициент показывает, какая переменная внесла наибольший вклад в j -тую ГК.

Корреляционная матрица

Исходная таблица данных. Ищем корреляцию предметов. В таблице хранятся набранные учениками баллы по дисциплинам.

	Родной язык	Обществоведение	Естествознание	Иностран. язык	Математика
Ученик 1	92	83	77	156	38
Ученик 2	97	82	68	114	33
Ученик 3	100	100	93	176	44
Ученик 4	89	77	100	158	46
Ученик 5	95	79	75	140	37
Ученик 6	99	96	84	174	42
Ученик 7	97	87	98	190	49
Ученик 8	93	77	73	132	35
Ученик 9	89	75	72	132	35
Ученик 10	98	93	70	186	37

	Р	О	Е	И	М
Р	1	0.86	0.03	0.46	0.17
О	0.86	1	0.26	0.71	0.38
Е	0.03	0.26	1	0.59	0.97
И	0.46	0.71	0.59	1	0.75
М	0.17	0.38	0.97	0.75	1

Рассчитаем корреляционную матрицу

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & 1 & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & 1 & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & 1 & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.86 & 0.03 & 0.46 & 0.17 \\ 0.86 & 1 & 0.26 & 0.71 & 0.38 \\ 0.03 & 0.26 & 1 & 0.59 & 0.97 \\ 0.46 & 0.71 & 0.59 & 1 & 0.75 \\ 0.17 & 0.38 & 0.97 & 0.75 & 1 \end{pmatrix}$$

r_{ij} и r_{ji} равны

Обратите внимание:

1. Матрица симметрична. Почему?
2. По диагонали размещены единицы. Почему?

Собственные значения и собственные векторы матриц

У каждой матрицы есть собственный вектор и собственное число. Например, собственное значение для матрицы

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \text{ то есть } \begin{cases} a_1 + 2a_2 = \lambda a_1 \\ 3a_1 + 4a_2 = \lambda a_2 \end{cases}$$

это такое значение λ , которое удовлетворяет вот этому уравнению:

$$R \cdot V = \lambda V$$

где R – матрица, для которой ищется решение;

V – искомый собственный вектор.

λ – собственное число.

Решение основано на простой форме в виде детерминанта матрицы

$$\text{Det}(R - \lambda I) = 0$$

Это дает для квадратной матрицы уравнение

$$\text{Det} \begin{vmatrix} 1-\lambda & r_{12} \\ r_{12} & 1-\lambda \end{vmatrix} = 0$$

которое может быть представлено в виде:

$$(1-\lambda)^2 - r_{12}^2 = 0$$

Раскрывая скобки и группируя, получаем

$$\lambda^2 - 2\lambda + (1 - r_{12}^2) = 0$$

Собственные числа могут быть получены при решении квадратного уравнения. Для двумерной корреляционной матрицы собственные числа имеют вид:

$$\lambda_1 = 1 + r_{12}$$

$$\lambda_2 = 1 - r_{12}$$

Если между двумя переменными имеется функциональная линейная зависимость, то одно собственное число будет 2, другое – 0. Для некоррелированных переменных оба собственных числа будут равны 1. Заметим, что

- сумма собственных чисел $\lambda_1 + \lambda_2 = (1 + r_{12}) + (1 - r_{12}) = 2$ равна числу переменных;
- а произведение $\lambda_1 \lambda_2 = (1 - r_{12}^2)$ равна детерминанту корреляционной матрицы;

Эти свойства сохраняются для корреляционных матриц любой размерности, причем первое большее собственное число представляет величину дисперсии, соответствующую первой главной оси, а второе собственное число – величину дисперсии, соответствующую второй главной оси и т. д.

Так как при использовании корреляционной матрицы сумма собственных чисел равна m (число переменных), можем получить доли дисперсии, соответствующую данному направлению или компоненте:

$$\frac{\lambda_k}{m}$$

Набор собственных значений $\lambda_1, \dots, \lambda_n$ матрицы A называется спектром A .

Пример 1

$$\begin{pmatrix} -10 & 6 \\ -18 & 11 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -10 \times 1 + 6 \times 2 \\ -18 \times 1 + 11 \times 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\begin{pmatrix} -10 & 6 \\ -18 & 11 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -10 \times 2 + 6 \times 3 \\ -18 \times 2 + 11 \times 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \end{pmatrix} = - \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Следовательно,

- 2 и -1 являются собственными значениями данной матрицы;
- Собственным вектором для 2 является вектор $(1,2)^T$, а для -1 вектор $(2,3)^T$.

Пример 2

Матрица, имеющая P строк и P столбцов, как правило, имеет P пар собственных значений и собственных векторов.

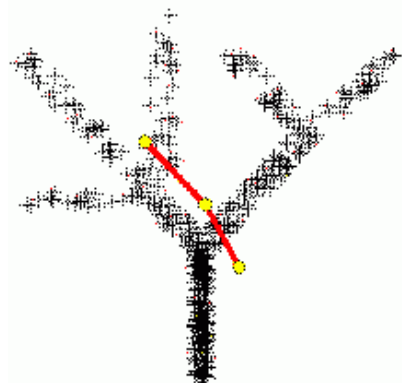
$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \times 1 + 0 \times 0 + 0 \times 0 \\ 0 \times 1 + 4 \times 0 + 0 \times 0 \\ 0 \times 1 + 0 \times 0 + 6 \times 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \times 0 + 0 \times 1 + 0 \times 0 \\ 0 \times 0 + 4 \times 1 + 0 \times 0 \\ 0 \times 0 + 0 \times 1 + 6 \times 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \times 0 + 0 \times 0 + 0 \times 1 \\ 0 \times 0 + 4 \times 0 + 0 \times 1 \\ 0 \times 0 + 0 \times 0 + 6 \times 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} = 6 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Пределы применимости и ограничения эффективности метода

Метод не всегда эффективно снижает размерность при заданных ограничениях на точность. **Прямые и плоскости не всегда обеспечивают хорошую аппроксимацию.** Например, данные могут с хорошей точностью следовать какой-нибудь кривой, а эта кривая может быть сложно расположена в пространстве данных. В этом случае метод главных компонент для приемлемой точности потребует нескольких компонент (вместо одной), или вообще не даст



снижения размерности при приемлемой точности.

Для работы с такими «кривыми» главными компонентами изобретен метод главных многообразий и различные версии нелинейного метода главных компонент. Больше неприятностей могут доставить данные сложной топологии. Для их аппроксимации также изобретены различные методы, например, самоорганизующиеся карты Кохонена, нейронный газ или топологические грамматики.

Применение МГК

1. **Визуализация данных**
2. **Сжатие данных.** Экономия памяти, храня только значения компонентов вместо значений всех пикселей для каждой фотографии. (как подобрать платье с помощью метода главных компонент)
3. **Обработка изображений.**
4. Индексация видео
5. Психодиагностика
6. Общественные науки

Лекция 5. Факторный анализ

При исследовании сложных объектов и систем часто невозможно измерить показатели, определяющие свойства этих объектов. Такие показатели называют факторами. Иногда нам неизвестны даже число и содержательный смысл этих факторов. Для измерений могут быть доступны иные величины, тем или иным способом зависящие от этих факторов. При этом, когда влияние неизвестного фактора проявляется в нескольких измеряемых признаках, эти признаки могут обнаруживать тесную связь между собой, поэтому общее число факторов может быть гораздо меньше, чем число измеряемых переменных-признаков, которое обычно выбирается исследователем в той или иной мере произвольно. Для обнаружения влияющих на измеряемые признаки факторов используются методы факторного анализа (ФА).

При анализе различных явлений и процессов часто необходимо учитывать большое число показателей, многие из которых связаны и дублируют друг друга. Нередко эти признаки лишь в косвенной мере отражают наиболее существенные, но не поддающиеся наблюдению и измерению внутренние явления.

Идея факторного анализа состоит в том, что из матрицы корреляции извлекается информация о взаимной зависимости признаков друг с другом. Если соответствующий коэффициент корреляции достаточно высок, то можно сказать, что эти признаки лишние и их можно заменить одним фактором. Всё это можно хорошо проиллюстрировать на графике линейной регрессии. **Если некоторый фактор**

сопоставить соответствующей линии регрессии, то можно говорить о снижении размерности. При этом новый фактор является линейной комбинацией прежних признаков.

Вообще, факторный анализ преследует две цели:

1. сокращение числа переменных (редукция данных) ;
2. классификацию переменных - определение структуры взаимосвязей между переменными.

Поэтому факторный анализ используется либо как метод сокращения данных, либо как метод классификации.

Существуют две модели факторного анализа:

1. метод главных компонент
2. анализ главных факторов.

Основное отличие этих методов в том, что в **методе главных компонент мы допускаем, что должна учитываться вся дисперсия, в то время как в анализе главных факторов мы используем только общую дисперсию.** В большинстве случаев эти два метода обычно дают очень близкие результаты. Тем не менее, метод главных компонент часто используется как метод снижения размерности, в то время как метод главных факторов часто предпочтителен, когда цель анализа — обнаружить структуру.

ФА можно применять только к количественным признакам! Использование порядковых, номинальных признаков для проведения ФА недопустимо. Это требование обусловлено тем, что входной информацией для ФА являются элементы ковариационной матрицы. Кроме того, представление переменных в виде линейной комбинации скрытых факторов и использование оценок факторов через линейные комбинации наблюдаемых переменных для порядковых переменных невозможно.

Сущность метода факторного анализа

Сущность метода факторного анализа – переход от описания некоторого множества объектов, заданного большим набором косвенных признаков, к описанию меньшим числом максимально информативных глубинных переменных. Такие переменные называются факторными и являются факторами исходных признаков. **В ФА исходят из того, что признаки, входящие в исследуемый набор, не являются независимыми, напротив, они коррелированы.**

Наличие корреляции между двумя или более признаками может трактоваться следующим образом: либо один из признаков определяет остальные, либо существует некий не включенный в набор скрытый параметр, оказывающий влияние на другие признаки. Такого рода скрытые параметры – общие факторы, а методы ФА предназначены для их выявления. Термин «общий» подчеркивает, что каждый фактор F_i имеет существенное значение **для анализа всех переменных.** В то же время

предполагается, что существуют и специфические или характерные факторы, каждый из которых влияет **только на одну конкретную переменную**.

Пример. Модель ФА можно интерпретировать в терминах интеллектуальных тестов.

<table style="display: inline-table; border: none;"> <tr> <td style="border: none;"></td> <td style="border: none; text-align: center;">p</td> </tr> <tr> <td style="border: none; text-align: center;">n</td> <td style="border: none;"></td> </tr> </table>		p	n		Тесты				
		p							
n									
	1	2	3	...	p				
Экзаменуемый 1	оценка								
...									
Экзаменуемый n									

В качестве ненаблюдаемых общих факторов, от которых будут зависеть оценки по всем тестам, выступают такие факторы как характеристика общей одаренности, характеристики его математических, технических или гуманитарных способностей.

Постановка задачи

Пусть

X – p-мерный случайный вектор;

F – m-мерный вектор, компонентами которого являются непосредственно не наблюдаемые переменные (факторы), то есть

$$\mathbf{F} = (F_1, \dots, F_m);$$

U – вектор сумм ненаблюдаемых ошибок и специфических факторов. Согласно основному предположению факторного анализа, каждое конкретное измерение x_i может рассматриваться как сумма воздействий некоторого небольшого числа общих факторов $\{F_i\}$, взятых с определенными весами λ_{ij} , специфического фактора s_i , воздействующего только на данную переменную, и ошибки измерения e_i . Поскольку s_i и e_i в факторном анализе неразличимы, их обычно рассматривают как сумму $u_i = s_i + e_i$.

$\Lambda^{p \times m} = \{\lambda_{ij}\}$ – матрица факторных нагрузок ($i=1..p, j=1..m$). Определим матрицу Λ порядка $p \times m$ как матрицу, элементами которой являются факторные веса, определяющие нагрузку i-ой переменной на j-ый фактор.

Примем следующие допущения:

1. Общие факторы не коррелированы и имеют единичные дисперсии.
2. Характерные факторы не коррелированы и $Du_i = \tau_i^2, i = 1..p$, где τ_i^2 – специфическая дисперсия или специфичность i-ой исходной переменной.
3. Переменные f_i и u_i не коррелированы.
4. Зависимость признаков от общих факторов линейная.

Таким образом, каждый из признаков x_i , входящий в исследуемый набор, может быть представлен как функция небольшого числа общих факторов и специфического фактора:

$$x_i = f(F_1, F_2, \dots, F_m, U),$$

или с учетом линейности модели:

$$x_1 = \sum_{j=1}^m \lambda_{1j} F_j + U_1, \quad \dots \quad x_p = \sum_{j=1}^m \lambda_{pj} F_j + U_p.$$

Принимая во внимание допущения 1, 2, 3 получим:

$$Dx_i = \sum_{j=1}^m \lambda_{ij}^2 + \tau_i^2 + 2 \sum_{j < q=1}^m \lambda_{ij} \lambda_{iq} r_{F_i F_q} + 2\tau_i \sum_{j=1}^m \lambda_{ij} r_{F_j U_i} = \sum \lambda_{ij}^2 + \tau_i^2 = 1.$$

Первая сумма указывает долю дисперсии переменной x_i , приходящуюся на все общие факторы. Можно также записать:

$$Dx_i = h_i^2 + \tau_i^2$$

где h_i^2 – общность переменной x_i , τ_i^2 – характеристика переменной x_i .

Техника факторного анализа направлена на оценку фактических нагрузок λ_{ij} , а также общих факторов с помощью значений исходных переменных. **После того, как факторные нагрузки будут найдены, остается задача наилучшей интерпретации общих факторов.** Для этого используется *метод вращения фактора*, который из-за субъективности является наиболее спорной частью факторного анализа.

Надо понимать, что на практике невозможно получить точную структуру факторной модели, можно только пытаться найти оценки параметров факторной структуры с использованием определенных статистических или практических критериев.

Необходимо отметить связь метода главных компонент и факторного анализа. В модели главных компонент вся дисперсия приписывается общим факторам, тогда как в факторном анализе дисперсия состоит из двух частей: дисперсии, обусловленной наличием общих факторов, и дисперсии, обусловленной специфичностью.

Порядок выполнения факторного анализа

На первом шаге процедуры факторного анализа происходит стандартизация заданных значений переменных (z-преобразование); затем при помощи стандартизованных значений рассчитывают корреляционные коэффициенты Пирсона между рассматриваемыми переменными.

Исходным элементом для дальнейших расчётов является корреляционная матрица. Для понимания отдельных шагов этих расчётов потребуются хорошие знания, прежде всего, в области операций над матрицами. Для построенной корреляционной матрицы определяются, так называемые, **собственные значения и соответствующие им собственные векторы**, для определения которых используются оценочные значения диагональных элементов матрицы (так называемые относительные дисперсии простых факторов).

Собственные значения сортируются в порядке убывания. После для чего обычно отбирается столько факторов, сколько имеется собственных значений, **превосходящих по величине единицу**. Собственные векторы, соответствующие этим собственным значениям, образуют факторы; элементы собственных векторов получили название **факторной нагрузки**. Их можно понимать как коэффициенты корреляции между соответствующими переменными и факторами. Для решения такой задачи определения факторов были разработаны многочисленные методы, наиболее часто употребляемым из которых является метод определения главных факторов (компонентов).

Описанные выше шаги расчёта ещё не дают однозначного решения задачи определения факторов. Основываясь на геометрическом представлении рассматриваемой задачи, поиск однозначного решения называют задачей вращения факторов. И здесь имеется большое количество методов, наиболее часто употребляемым из которых является ортогональное вращение по так называемому методу варимакса. **Факторные нагрузки повёрнутой матрицы могут рассматриваться как результат выполнения процедуры факторного анализа**. Кроме того, на основании значений этих нагрузок необходимо попытаться дать толкование отдельным факторам.

Если факторы найдены и истолкованы, то на последнем шаге факторного анализа, отдельным наблюдениям можно присвоить значения этих факторов, так называемые факторные значения. Таким образом, для каждого наблюдения значения большого количества переменных можно перевести в значения небольшого количества факторов.

Этапы факторного анализа

При решении задач факторного анализа выделяют этапы:

1. Формулировка проблемы
2. Подготовка исходной матрицы данных (ТЭД). Расчет соответствующей матрицы взаимосвязей признаков.
3. Выделение первоначальных ортогональных факторов (факторизация).
4. Вращение – преобразование факторов, облегчающее их интерпретацию.
5. Интерпретация данных.
6. Оценка значений факторов. Подсчет факторных значений по каждому фактору для каждого наблюдения.

В ходе выполнения ФА некоторые этапы можно опустить. Например, в качестве исходной матрицы данных можно использовать корреляционную матрицу или эквивалентную ей любую другую матрицу связей, подсчитанную в ходе какой-то другой вычислительной процедуры. Тогда работа сразу начинается с третьего этапа.

Кроме того, так как нет наблюдений, невозможно и произвести оценку значений факторов. Именно этот случай рассмотрен в одном из примеров ниже.

1. Формулировка проблемы

Формулировка проблемы включает несколько задач. Во-первых, четкое определение целей факторного анализа. Переменные, подвергаемые факторному анализу, задаются исходя из прошлых исследований, теоретических выкладок и по усмотрению исследователя. Важно, чтобы переменные измерялись в интервальной или относительной шкале. Выборка должна быть подходящего размера. Опыт подсказывает, что рекомендуется брать выборку, по крайней мере, в четыре или пять раз больше, чем число переменных. Часто при исследованиях размер выборки мал, и это отношение значительно меньше. В таких случаях следует осторожно интерпретировать результаты.

2. Вычисление матрицы взаимосвязей признаков

В основе анализа чаще всего лежит матрица корреляций между переменными. Ее анализ дает исследователям ценную информацию. Целесообразность выполнения факторного анализа определяется наличием корреляций между переменными. На практике так обычно и бывает. Если же корреляции между всеми переменными небольшие, то факторный анализ бесполезен. Следует также ожидать, что переменные, тесно взаимосвязанные между собой, должны также тесно коррелировать с одним и тем же фактором или факторами.

Для проверки целесообразности использования факторной модели анализа зависимости переменных существует несколько статистик. С помощью критерия сферичности Бартлетта проверяется нулевая гипотеза об отсутствии корреляций между переменными в генеральной совокупности: другими словами, рассматривается утверждение о том, что корреляционная матрица совокупности — это единичная матрица, в которой все диагональные элементы равны 1, а все остальные равны 0. Проверка с помощью критерия сферичности основана на преобразовании детерминанта корреляционной матрицы в статистику хи-квадрат. При большом значении статистики нулевую гипотезу отклоняют. Если же нулевую гипотезу не отклоняют, то целесообразность выполнения факторного анализа вызывает сомнения. Другая полезная статистика — критерий адекватности выборки Кайзера—Мейера—Олкина (КМО). Данный коэффициент сравнивает значения наблюдаемых коэффициентов корреляции со значениями частных коэффициентов корреляции. Небольшие значения КМО-статистики указывают на то, что корреляции между парами переменных нельзя объяснить другими переменными и что использование факторного анализа нецелесообразно.

Пример. В таблице ниже приведена корреляционная матрица между 8-ю признаками (размерами) человеческого тела. Как видно из анализа таблицы многие признаки имеют сильные корреляционные связи (красным цветом отмечены связи с $r > 0.7$).

Признаки		Коэффициенты корреляции между признаками							
		1	2	3	4	5	6	7	8
1	Рост	1	0,85	0,81	0,86	0,47	0,4	0,3	0,38
2	Размах рук		1	0,88	0,83	0,28	0,33	0,28	0,42
3	Длина предплечья			1	0,8	0,38	0,32	0,24	0,35
4	Длина ноги				1	0,44	0,33	0,33	0,37
5	Вес					1	0,76	0,73	0,33
6	Окружность бедер						1	0,58	0,58
7	Окружность груди							1	0,54
8	Ширина груди								1

3. Определение главных факторов (факторизация)

Первой задачей факторного анализа является определение по ковариационной (или корреляционной) матрице $S = \{s_{ij}\}$ (или $R = \{r_{ij}\}$) оценок $\hat{\lambda}_{ij}$ факторных нагрузок и оценок специфичности $\hat{\tau}_i^2$. Для определения искомых оценок существует ряд методов:

- метод главных факторов,
- центроидный метод,
- метод наибольшего правдоподобия.

Все эти методы строятся как методы решения задач на поиск экстремума некоторого критерия качества матрицы факторных нагрузок Λ при определенных ограничениях. Различие между методами определяется видом используемых критериев.

Чаще всего используется *метод главных факторов*. В качестве исходной информации задают:

1. Число общих факторов.
2. Вид матрицы, к которой следует применить факторный анализ.

Число общих факторов определяется целым числом m или постоянной константой c . В последнем случае m полагается равным числу собственных значений, превосходящих c . При определении числа m можно руководствоваться критериями:

1. **Критерий Кайзера или критерий собственных чисел.** Этот критерий предложен Кайзером, и является, вероятно, наиболее широко используемым. Отбираются только факторы с собственными значениями равными или большими 1. Это означает, что если фактор не выделяет дисперсию, эквивалентную, по крайней мере, дисперсии одной переменной, то он опускается.

2. **Критерий каменистой осыпи или критерий отсеивания.** Он является графическим методом, впервые предложенным психологом Кэттелом. Собственные значения возможно изобразить в виде простого графика. Кэттел предложил найти такое место на графике, где убывание собственных значений слева направо максимально замедляется. Предполагается, что справа от этой точки находится только «факториальная осыпь» — «осыпь» является геологическим термином, обозначающим обломки горных пород, скапливающиеся в нижней части скалистого склона. Однако этот критерий отличается высокой субъективностью и, в отличие от предыдущего критерия, статистически необоснован.

Недостатки обоих критериев заключаются в том, что первый иногда сохраняет слишком много факторов, в то время как второй, напротив, может сохранить слишком мало факторов; однако оба критерия вполне хороши при нормальных условиях, когда имеется относительно небольшое число факторов и много переменных. На практике возникает важный вопрос: когда полученное решение может быть содержательно интерпретировано. В этой связи предлагается использовать ещё несколько критериев.

3. **Критерий значимости.** Он особенно эффективен, когда модель генеральной совокупности известна и отсутствуют второстепенные факторы. Но критерий непригоден для поиска изменений в модели и реализуем только в факторном анализе по методу наименьших квадратов или максимального правдоподобия.
4. **Критерий доли воспроизводимой дисперсии.** Факторы ранжируются по доле детерминируемой дисперсии, когда процент дисперсии оказывается незначительным, выделение следует остановить. Желательно, чтобы выделенные факторы объясняли более 80 % разброса. Недостатки критерия: во-первых, субъективность выделения, во-вторых, специфика данных может быть такова, что все главные факторы не смогут совокупно объяснить желательного процента разброса. Поэтому главные факторы должны вместе объяснять не меньше 50,1 % дисперсии.
5. **Критерий интерпретируемости и инвариантности.** Данный критерий сочетает статистическую точность с субъективными интересами. Согласно ему, главные факторы можно выделять до тех пор, пока будет возможна их ясная интерпретация. Она, в свою очередь, зависит от величины факторных нагрузок, то есть если в факторе есть хотя бы одна сильная нагрузка, он может быть интерпретирован. Возможен и обратный вариант — если сильные нагрузки имеются, однако интерпретация затруднительна, от этой компоненты предпочтительно отказаться.

Напомним, что

- Сумма собственных чисел равна числу переменных, то есть p .
- Общая дисперсия может быть вычислена как след ковариационной матрицы (то есть равна сумме ее диагональных элементов).

Тогда можно прийти к выводу, что для корреляционной матрицы общая дисперсия будет равна сумме собственных чисел, то есть p . Тогда, разделив собственное

число на число переменных, можем получить долю дисперсии, соответствующую данному фактору или компоненте:

$$\text{Доля, соответствующая данной компоненте} = \frac{\text{Соответствующее собственное число}}{p}$$

При проведении факторного анализа все расчеты носят последовательный характер. Процедура выполнения вычислительных операций схематично представлена на рис. 1.



При этом приводится подробное описание важнейших матриц. Вертикальные стрелки соответствуют основным проблемам, возникающим при проведении факторного анализа в тех местах схемы, куда указывают эти стрелки.

Любой метод факторного анализа начинается с Y матрицы исходных данных. По ней вычисляется корреляционная матрица R . По главной диагонали корреляционной матрицы затем проставляют оценки общностей и получают $R_h = (r_{ik}^h)$. Это составляет проблему общности, которая состоит в установлении оценок общностей. Это самая первая проблема, которая возникает в ходе факторного анализа. Для ее решения в качестве общностей выбираем наибольшие значения элементов в каждом столбце матрицы R . Стрелка между R_h и A указывает на проблему факторов. Из R_h с помощью определенных способов извлекают факторы, получая в результате матрицу A . Столбцы матрицы A ортогональны и занимают

произвольную позицию в отношении переменных, определяемую методом выделения факторов. Матрица A воспроизводит R_h по равенству $R_h = A \cdot A'$

Выводы

1. На данном этапе определяется минимальное число факторов, адекватно воспроизводящих наблюдаемые корреляции, а также значения общностей каждой переменной.
2. Применение указанных методов приводит к набору ортогональных факторов, упорядоченных в порядке убывания их значимости.

Эти ограничения принимаются, чтобы обеспечить единственность решения. В результате этих ограничений

- Факторная сложность переменных, скорее всего будет больше единицы, независимо от вида истинной факторной структуры, то есть переменные будут иметь нагрузки более чем на один фактор (другими словами зависеть не от одного, а нескольких факторов);
- Все факторы, за исключением первого, являются биполярными, другими словами, некоторые переменные должны иметь положительную нагрузку на этот фактор, а некоторые – отрицательную.

Пример. Результаты факторного анализа

Собственные значения				
<i>Выделение: Главные компоненты</i>				
	Соб. зн.	% общей	Кумулятивн. собст. знач.	Кумулятивн. %
1	4,597349	57,46686	4,597349	57,46686
2	1,723141	21,53926	6,320489	79,00612

Фактор.нагрузки (без вращ.)		
<i>Выделение: Главные компоненты (Отмечены нагрузки >0,7)</i>		
	Фактор 1	Фактор 2
Рост	-0,873232	0,343122
Размах рук	-0,842998	0,444261
Длина предплечья	-0,828256	0,429134
Длина ноги	-0,855072	0,360417
Вес	-0,696270	-0,530817
Окружность бедер	-0,673535	-0,569179
Окружность груди	-0,615703	-0,616589
Ширина груди	-0,624085	-0,328739
Общ.дис.	4,597349	1,723141
Доля общ	0,574669	0,215393

Факторные нагрузки могут интерпретироваться как корреляции между соответствующими переменными и факторами – чем выше нагрузка по модулю, тем больше близость фактора к исходной переменной; таким образом, они представляют наиболее важную информацию для интерпретации полученных факторов. В сгенерированной таблице для облегчения трактовки выделены факторные нагрузки по абсолютной величине больше 0,7.

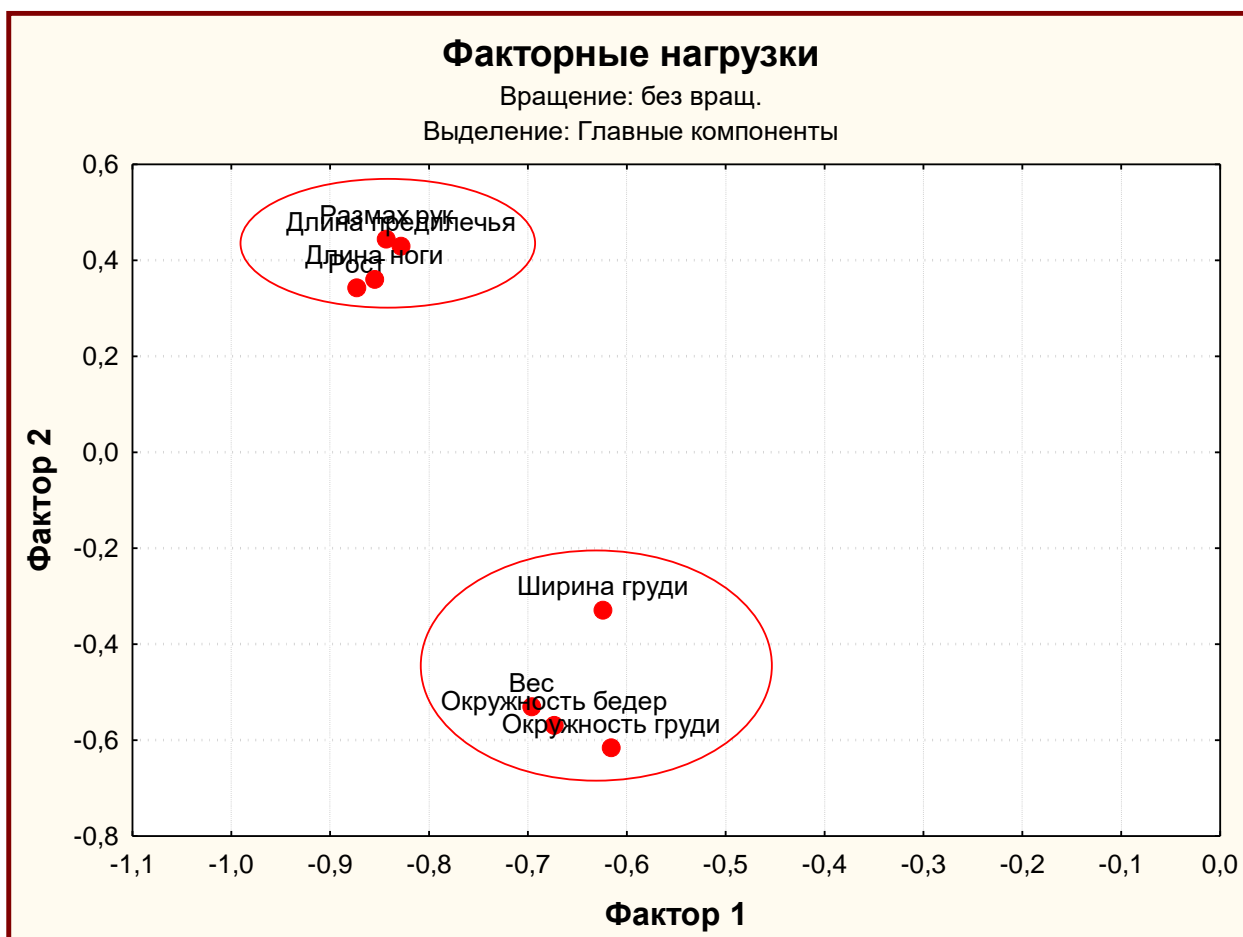


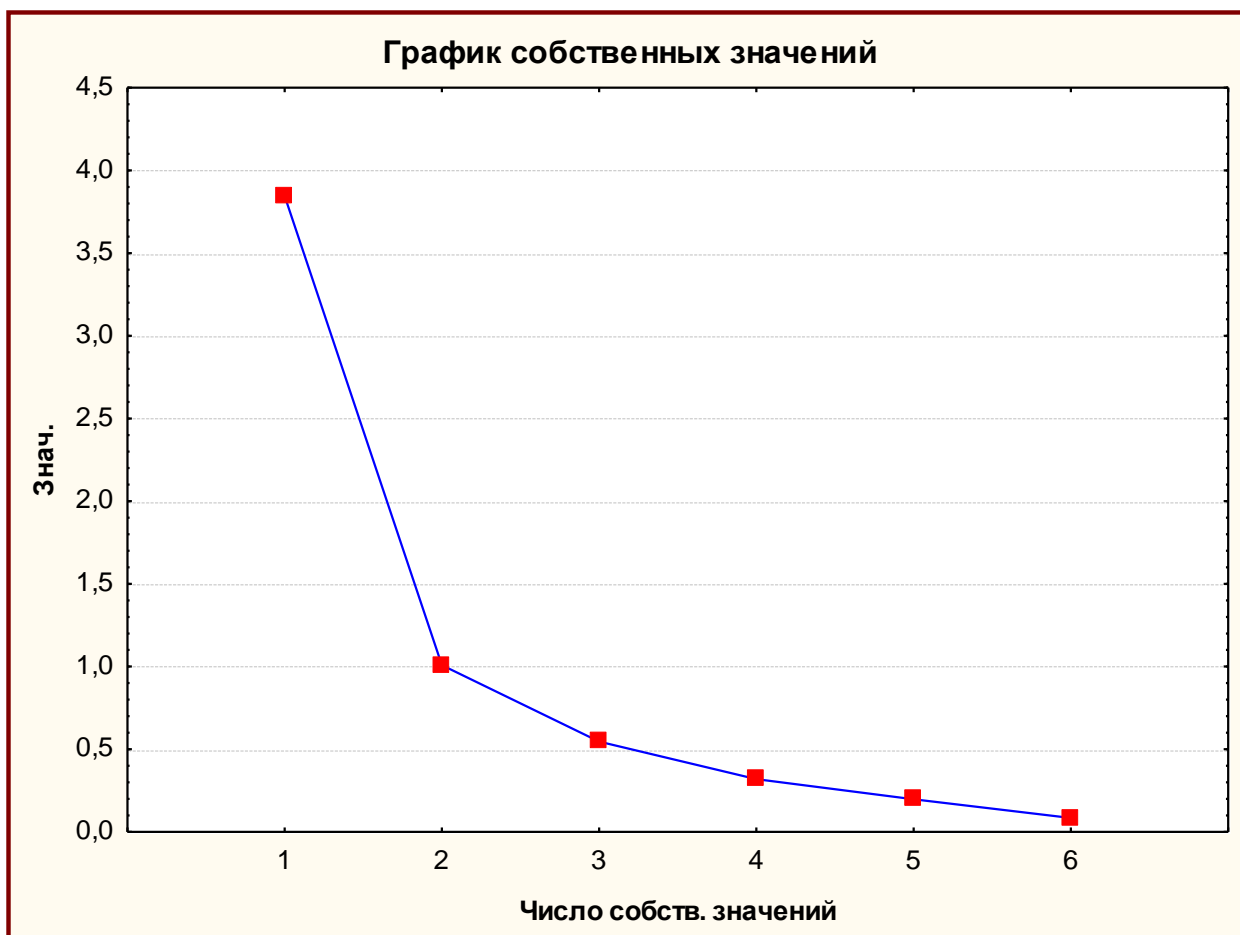
Рис.13.1.6 Агрегирование некоторых признаков человеческого тела в результате факторного анализа

Результаты факторного анализа приведены на рисунке, из которого видно, что анализируемые в примере признаки человеческого тела образуют два хорошо выраженных агрегата: *рост, размах рук, длина предплечий и длина ног* образуют один агрегат, а *вес, окружность бедер, окружность груди и ширина груди* – второй

Общности			
<i>Выделение: Главные компоненты Вращение: без вращ.</i>			
	Из 1	Из 2	Множест.
Рост	0,762535	0,880267	0,840368
Размах рук	0,710645	0,908013	0,901323
Длина предплечья	0,686008	0,870164	0,833255

Длина ноги	0,731148	0,861048	0,800860
Вес	0,484792	0,766559	0,880394
Окружность бедер	0,453649	0,777614	0,775985
Окружность груди	0,379090	0,759272	0,748005
Ширина груди	0,389482	0,497552	0,618233

График каменистой осыпи



4. Вращение факторов

Следующим шагом факторного анализа является интерпретация каждого фактора. Для этого можно воспользоваться неоднозначностью определения факторов. Полученные факторы F_1, \dots, F_m можно заменить их линейными комбинациями, которые взаимно не коррелированы и имеют единичные дисперсии. Таким образом, имеется бесконечное множество наборов факторов, удовлетворяющих данной модели.

Аналогично тому, как мы отображаем объект точкой в пространстве признаков, можно отобразить каждый признак точкой в пространстве факторов. При этом

проекция точки на факторы – есть величины факторных нагрузок. Представление всех переменных в пространстве общих факторов называется **конфигурацией**.

Ниже на рисунках представлены соответственно случайная, простая ортогональная и косоугольная структуры расположения наблюдений в пространстве общих факторов F_1 и F_2 .

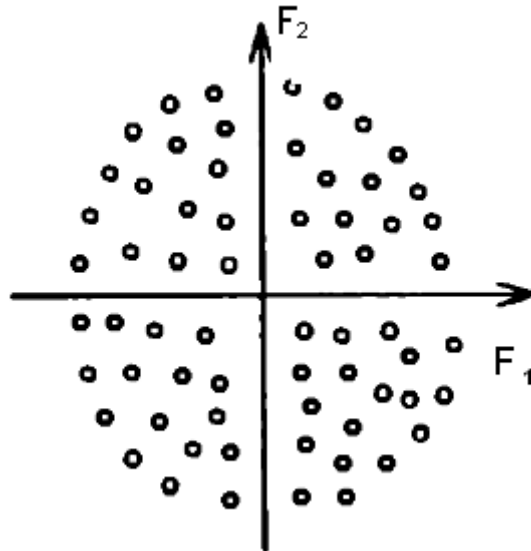
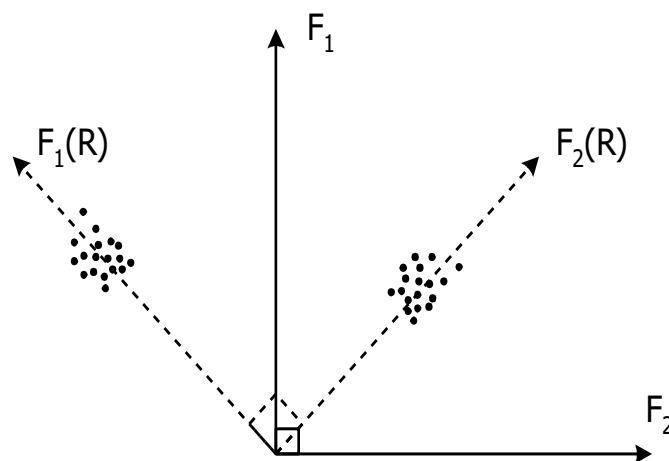


Рис. Случайная конфигурация векторов

Рис. Ортогональная простая структура



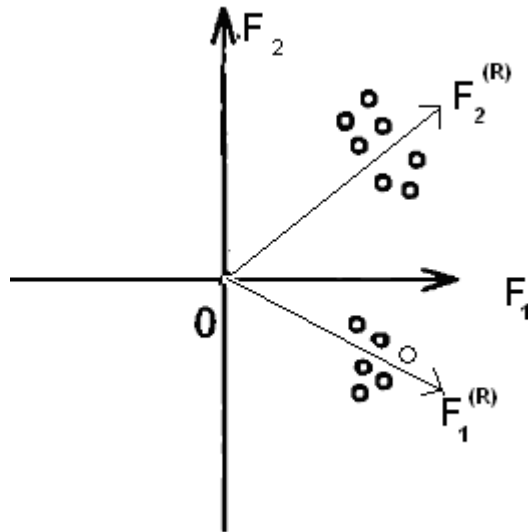


Рис. Косоугольная простая структура

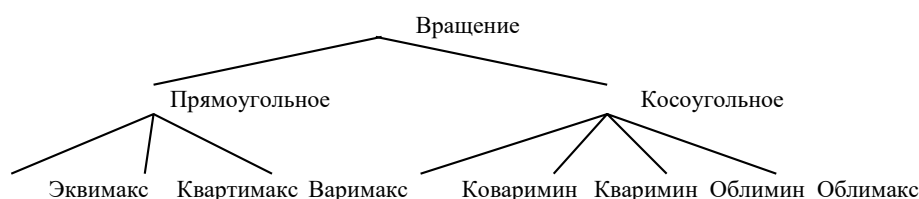
Интерпретация факторов проводится с помощью процедуры вращения. При этом число факторов и значения общностей фиксируются. Будем вращать систему координат вокруг ее начала. При таком вращении остаются неизменными расстояния от точек до начала координат (длины векторов) и углы между векторами.

Целью всех вращений является получение наиболее простой факторной структуры. Хотя трудно определить минимальные требования к простой структуре, но если взять число факторов r и число переменных n , то всегда можно сказать, какая структура наиболее проста. Факторная структура является простейшей, если когда все переменные имеют факторную сложность, равную 1, то есть когда каждая переменная имеет ненулевую нагрузку только на один общий фактор. При случайной конфигурации наблюдений невозможно получить такую структуру.

Существуют три различных подхода к проблеме вращения:

- ❶ Первый подход – графический. В этом случае исходные переменные рассматриваются как точки в факторном пространстве, координаты которых равны нагрузкам на факторы, а размерность определяется числом факторов. Вращение заключается в проведении новых осей, которые соответствуют некоторому критерию простой, легко интерпретируемой структуры. Если в пространстве факторов есть явные скопления (кластеры) точек (переменных), легко отделяемые друг от друга, простая структура получается в том случае, когда оси проведены через эти скопления. Но если такое разделение не очевидно или число факторов велико, графический метод неприменим.
- ❷ Второй подход связан с аналитическими методами. В этом случае выбирается некоторый объективный критерий, которым надо руководствоваться при выполнении вращения. В рамках этого подхода различают два вида вращения – ортогональное и косоугольное.

Существуют различные варианты ортогонального и косоугольного вращений:



После прямоугольного вращения модель может быть записана:

$$x_i = \sum_{j=1}^m c_{ij} F_j^{(R)} + U_i,$$

где $i = 1, \dots, p$, c_{ij} – нагрузки постоянных факторов.

В результате ортогонального вращения факторов общность каждой переменной x_i остается без изменений:

$$c_{ij} = \sum_{k=1}^m \hat{\lambda}_{ij} q_{kj}$$

Для облегчения интерпретации факторов постоянные q_{kj} выбираются так, чтобы результирующие нагрузки имели простую структуру. Структура фактических нагрузок считается простой, если большинство из C_{ij} не слишком сильно отличаются от 0.

При косоугольном вращении исходят из предположения, что важнее получить простую структуру факторных нагрузок, чем сохранить ортогональность факторов. Поэтому условие некоррелированности факторов ослабляется, и ищутся коррелированные факторы $F_1^{(R)}, \dots, F_m^{(R)}$ с единичными дисперсиями, являющимися линейными комбинациями факторов F_1, \dots, F_m . Так как полученные факторы могут быть коррелированными, имеется более широкая область изменения $q_{k.j}$.

❸ Третий поход заключается в задании априорной целевой матрицы. Цель вращения – нахождение факторного отображения, наиболее близкого к некоторой заданной матрице.

Пример.

Факторные нагрузки (Варимакс нормал.) <i>Выделение: Главные компоненты (Отмечены нагрузки >,7)</i>		
	Фактор 1	Фактор 2
Рост	0,898597	0,269798
Размах рук	0,937320	0,171595

Длина предплечья	0,916382	0,174380
Длина ноги	0,895005	0,244978
Вес	0,219446	0,847586
Окружность бедер	0,177866	0,863700
Окружность груди	0,103101	0,865241
Ширина груди	0,287556	0,644099
Общ.дис.	3,499945	2,820544
Доля общ	0,437493	0,352568

Общности			
<i>Выделение: Главные компоненты Вращение: Варимакс нормал.</i>			
	Из 1	Из 2	Множест.
Рост	0,807476	0,880267	0,840368
Размах рук	0,878569	0,908014	0,901323
Длина предплечья	0,839755	0,870164	0,833255
Длина ноги	0,801034	0,861048	0,800860
Вес	0,048156	0,766559	0,880394
Окружность бедер	0,031636	0,777614	0,775985
Окружность груди	0,010630	0,759272	0,748005
Ширина груди	0,082689	0,497552	0,618233

Воспроизведенные корреляции								
<i>Выделение: Главные компоненты</i>								
	Рост	Размах рук	Длина предплечья	Длина ноги	Вес	Окружность бедер	Окружность груди	Ширина груди
Рост	0,88	0,89	0,87	0,87	0,43	0,39	0,33	0,43
Размах рук	0,89	0,91	0,89	0,88	0,35	0,31	0,25	0,38
Длина предплечья	0,87	0,89	0,87	0,86	0,35	0,31	0,25	0,38
Длина ноги	0,87	0,88	0,86	0,86	0,40	0,37	0,30	0,42
Вес	0,43	0,35	0,35	0,40	0,77	0,77	0,76	0,61
Окружность бедер	0,39	0,31	0,31	0,37	0,77	0,78	0,77	0,61
Окружность груди	0,33	0,25	0,25	0,30	0,76	0,77	0,76	0,59
Ширина груди	0,43	0,38	0,38	0,42	0,61	0,61	0,59	0,50

Остаточные корреляции								
Выделение: Главные компоненты (Отмеченные остатки > 0,10)								
	Рост	Размах рук	Длина пред-плечья	Длина ноги	Вес	Окружность бедер	Окружность груди	Ширина груди
Рост	0,12	-0,04	-0,06	-0,01	0,04	0,01	-0,03	-0,05
Размах рук	-0,04	0,09	-0,01	-0,05	-0,07	0,02	0,03	0,04
Длина пред-плечья	-0,06	-0,01	0,13	-0,06	0,03	0,01	-0,01	-0,03
Длина ноги	-0,01	-0,05	-0,06	0,14	0,04	-0,04	0,03	-0,05
Вес	0,04	-0,07	0,03	0,04	0,23	-0,01	-0,03	-0,28
Окружность бедер	0,01	0,02	0,01	-0,04	-0,01	0,22	-0,19	-0,03
Окружность груди	-0,03	0,03	-0,01	0,03	-0,03	-0,19	0,24	-0,05
Ширина груди	-0,05	0,04	-0,03	-0,05	-0,28	-0,03	-0,05	0,50

5. Интерпретация факторов

Целью процедуры вращения является представление каждой исходной переменной x_i одним, или небольшим числом факторов. Нагрузки остальных факторов равны 0. Тогда задача интерпретации значительно облегчается, так как каждая нагрузка при использовании матрицы корреляции R, равна корреляции между исходной переменной и соответствующим фактором.

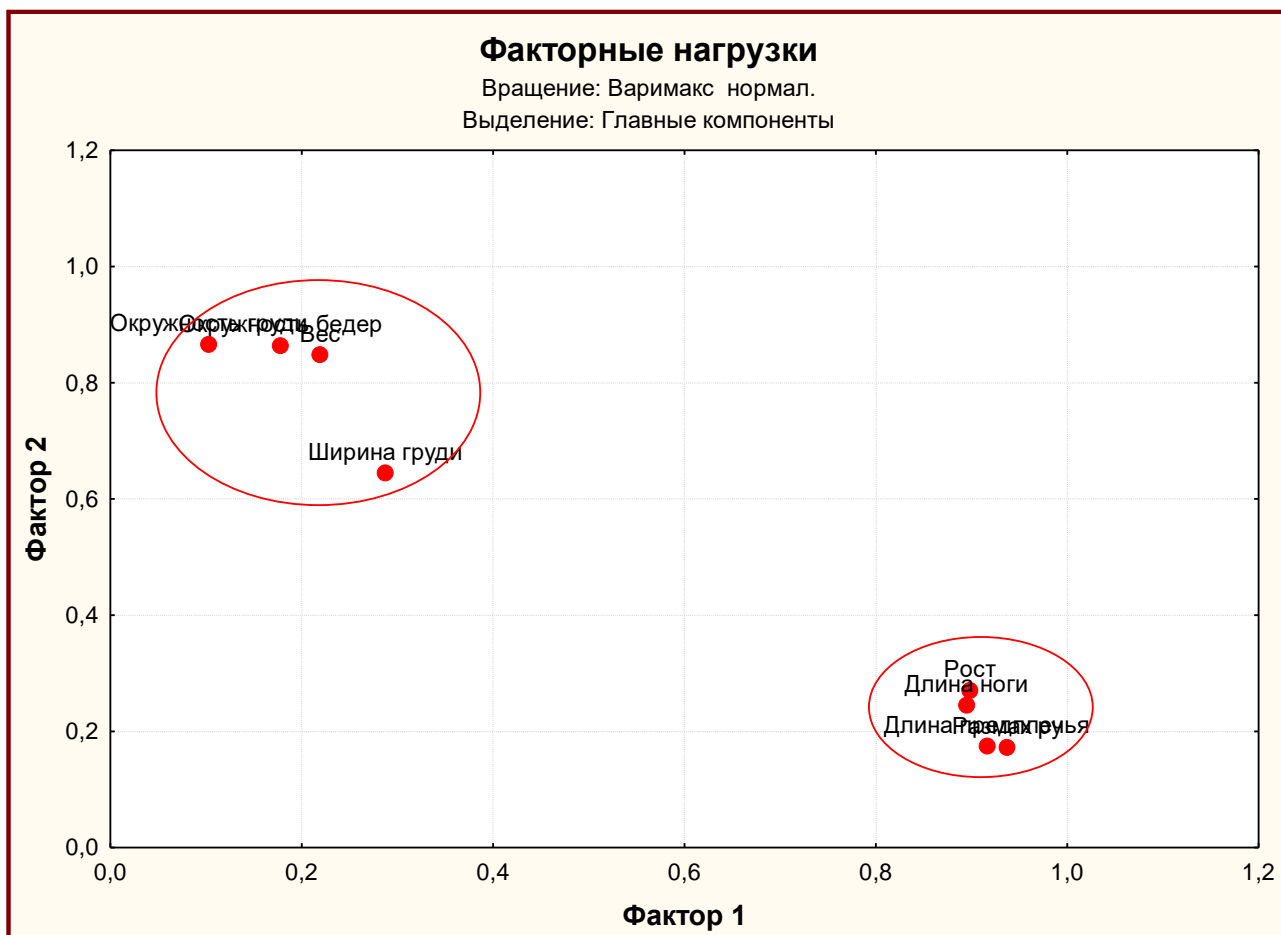
Как же интерпретировать факторы после вращения?

Для интерпретации факторов необходимо определить переменные, которые имеют высокие значения нагрузок по одному и тому же фактору. А затем этот фактор следует проанализировать с учетом этих переменных. Другое полезное средство интерпретации — графическое изображение переменных, координатами которых служат величины факторных нагрузок. Так, в конце оси расположены переменные, которые имеют большие нагрузки только в связи с этим фактором и, следовательно, характеризуют его. Переменные в начале координат имеют небольшие нагрузки в связи с обоими факторами. Переменные, расположенные вдали от осей, связаны с обоими факторами. Если фактор нельзя четко определить с точки зрения связи с исходными переменными, то его следует пометить как неопределяемый или генеральный (общий для всех переменных).

Если нет возможности провести вербальное объяснение факторов, то факторный анализ можно считать неудавшимся.

Пример. Исходя из анализа факторных нагрузок можно полагать, что значения перечисленных признаков человеческого тела в сущности являются проявлениями

двух других признаков (факторов) - **стройности** (F1) и **полноты** (F2), причем, если на содержательном уровне понятия стройности и полноты ясны любому из нас, то формализованных описаний этих свойств человеческого тела, тем более поддавшихся бы непосредственному инструментальному измерению до сих пор не существует.



6. Оценка значений общих факторов – факторное шкалирование

После интерпретации факторов необходимо вычислить их значения. Мало установить лишь сам факт существования небольшого числа скрыто существующих факторов, объясняющих природу взаимной коррелированности исходных признаков, и основную часть их дисперсии. Во многих случаях требуется определить значения факторов для данного вектора $X=(x^{(1)}, \dots, x^{(p)})$.

Под факторным шкалированием понимается процедура, позволяющая присвоить каждому объекту некоторые числовые оценки значений выделенных факторов, используя значения наблюдаемых переменных для этого объекта.

Для проведения подобной процедуры есть следующие основания:

1. после определения скрытой факторной структуры измеряемых данных для объектов исследователю может понадобиться представить каждый из этих объектов в терминах значений факторов, а не измеряемых переменных;
2. может появиться необходимость использования одного или более факторов в качестве переменных для дальнейшего анализа. ФА часто применяется в качестве средства для создания новых факторных переменных (шкал) для других исследований, чем для изучения самой скрытой структуры.

Главной целью факторного шкалирования является определение значений фактора (F) через наблюдаемые переменные (X_1, X_2, \dots). Общий фактор невозможно точно выразить посредством наблюдаемых переменных, поскольку каждая из них содержит также и характерную переменную, которую нельзя отделить от всей переменной. Можно получить лишь оценку значений общих факторов (\hat{F}) через наблюдаемые переменные. Поэтому шкалирование факторов всегда связано с некоторой неопределенностью.

Пример. Рассмотрим однофакторную модель с тремя переменными. Допустим, что все факторные нагрузки одинаковы (или что все коэффициенты корреляции равны). Этот пример показан на рисунке слева.

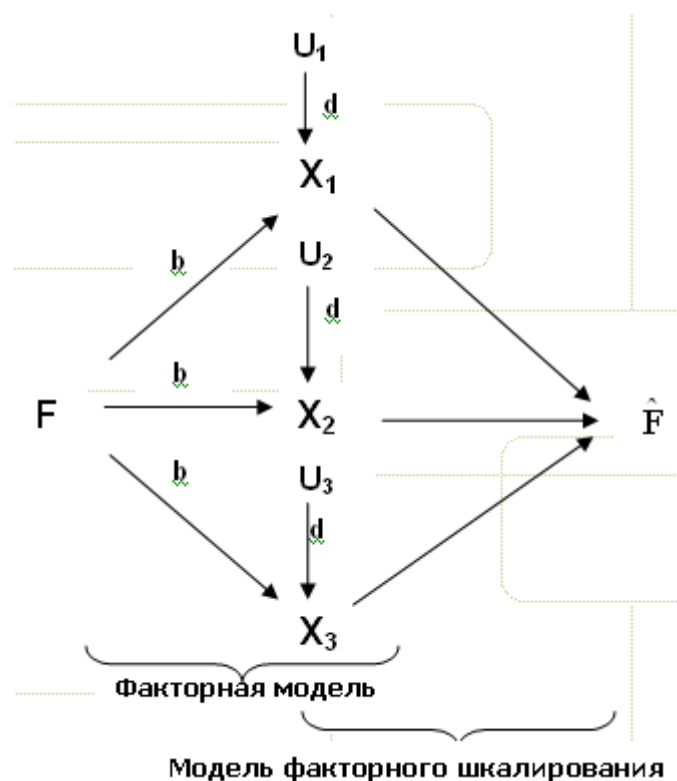


Рис. Графическая модель, иллюстрирующая зависимость между фактором и его оценкой

Для такой модели вычислить наблюдаемые коэффициенты корреляции между переменными можно с помощью перемножения факторных нагрузок. Так как все факторные нагрузки одинаковы, коэффициент корреляции будет равен квадрату факторной нагрузки:

$$r_{ij} = b_i b_j = b^2_i = b^2_j = h^2$$

В качестве оценки значения фактора берется линейная комбинация признаков X_1 , X_2 , X_3 . Так как каждая из этих переменных имеет одинаковую нагрузку от общих факторов, то естественно сложить их, беря соответствующие значения с одинаковым весом. Окончательное выражение будет иметь вид

$$\hat{F} = X_1 + X_2 + X_3,$$

а соответствующая диаграмма представлена в правой части рисунка.

Стандартного метода оценки значений факторов не существует. Как правило, для этой цели используется техника регрессионного анализа. Фактор представляет собой линейную комбинацию исходных переменных. Если рассматривать факторы как зависимые переменные, а исходные переменные x_i , $i=1\dots p$, считать независимыми, то можно записать следующие уравнения:

$$\hat{F}_j = \sum_{i=1}^p b_{ij} z_i \quad j=1, \dots, m$$

где \hat{F}_j - оценка значения j -го параметра,

z_i – стандартизованная оценка i -ой переменной;

b_{ij} - оценки коэффициентов регрессии, иногда они называются коэффициентами значений факторов.

Веса или коэффициенты значения фактора, используемые для объединения нормированных переменных, получают из матрицы коэффициентов значения фактора. Большинство компьютерных программ позволяет вычислить значения факторов. Только в анализе главных компонент можно вычислить точные значения факторов. Более того, в анализе главных компонент эти значения не взаимосвязаны. В анализе общих факторов оценки значений факторов получают, но нет гарантии, что факторы не будут коррелировать между собой. Значения факторов можно использовать вместо исходных переменных в последующем многомерном анализе. Например, используя матрицу коэффициентов значения фактора в табл. 19.3, можно вычислить два значения фактора для каждого респондента. Если, нормированные значения переменной умножить на соответствующий коэффициент значения фактора, то получится значение данного фактора.

Определение подгонки модели

Последняя стадия факторного анализа заключается в определении соответствия модели факторного анализа исходным данным, т.е. степени ее подгонки. Основное допущение, лежащее в основе факторного анализа, состоит в том, что наблюдаемая корреляция между переменными может быть свойственна общим факторам. Следовательно, корреляции между переменными можно вывести или воспроизвести из определенных корреляций между переменными и факторами. Изучив разности между наблюдаемыми корреляциями (данными в исходной корреляционной матрице) и вычисленными корреляциями (определенными из матрицы факторных нагрузок), можно определить соответствие модели исходным данным. Эти разности называют остатками (residuals). Если много остатков с большими значениями, то факторная модель не обеспечивает хорошее соответствие данным и требует пересмотра.

Лекция 6. Классификация

Задача классификации — задача, в которой имеется множество *объектов* (ситуаций), разделённых, некоторым образом, на *классы*. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется *выборкой*. Классовая принадлежность остальных объектов неизвестна. Требуется построить алгоритм, способный *классифицировать* произвольный объект из исходного множества.

Классифицировать объект — значит, указать номер (или наименование) класса, к которому относится данный объект.

Классификация объекта — номер или наименование класса, выдаваемый алгоритмом классификации в результате его применения к данному конкретному объекту.

В математической статистике задачи классификации называются также задачами дискриминантного анализа. В машинном обучении задача классификации решается, в частности, с помощью методов искусственных нейронных сетей при постановке эксперимента в виде обучения с учителем.

Существуют также другие способы постановки эксперимента — обучение без учителя, но они используются для решения другой задачи — кластеризации или таксономии. В этих задачах разделение объектов обучающей выборки на классы не задаётся, и требуется классифицировать объекты только на основе их сходства друг с другом. В некоторых прикладных областях, и даже в самой математической статистике, из-за близости задач часто не различают задачи кластеризации от задач классификации.

Некоторые алгоритмы для решения задач классификации комбинируют обучение с учителем с обучением без учителя, например, одна из версий нейронных сетей Кохонена — сети векторного квантования, обучаемые с учителем.

Типы входных данных

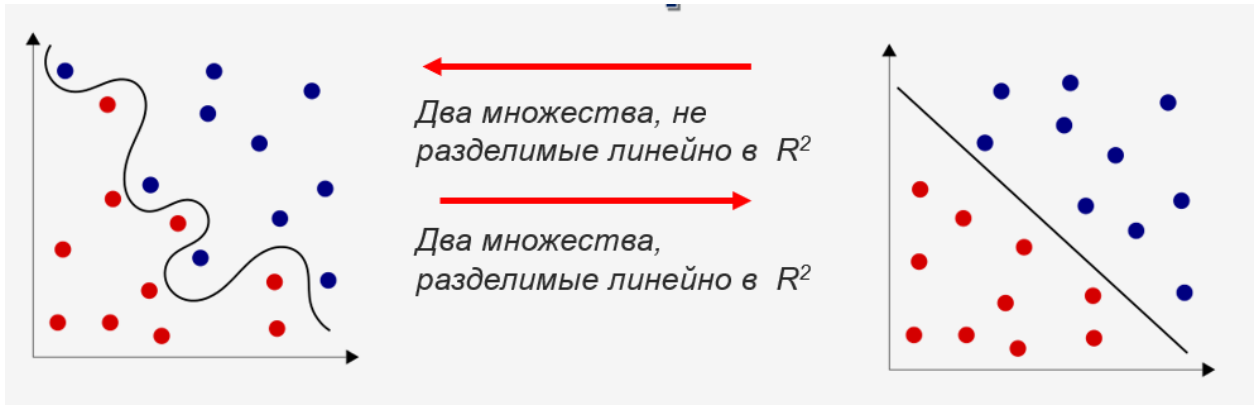
- **Признаковое описание** — наиболее распространённый случай. Каждый объект описывается набором своих характеристик, называемых *признаками*. Признаки могут быть числовыми или нечисловыми.
- **Матрица расстояний между объектами**. Каждый объект описывается расстояниями до всех остальных объектов обучающей выборки. С этим типом входных данных работают немногие методы, в частности, метод ближайших соседей, метод парзеновского окна, метод потенциальных функций.
- **Временной ряд** или **сигнал** представляет собой последовательность измерений во времени. Каждое измерение может представляться числом, вектором, а в общем случае — признаковым описанием исследуемого объекта в данный момент времени.
- **Изображение** или **видеоряд**.
- Встречаются и более сложные случаи, когда входные данные представляются в виде **графов**, текстов, результатов запросов к базе данных, и т. д. Как правило, они приводятся к первому или второму случаю путём предварительной обработки данных и извлечения признаков.
- Классификацию сигналов и изображений называют также распознаванием образов.

Типы классов

- **Двухклассовая классификация**. Наиболее простой в техническом отношении случай, который служит основой для решения более сложных задач.
- **Многоклассовая классификация**. Когда число классов достигает многих тысяч (например, при распознавании иероглифов или слитной речи), задача классификации становится существенно более трудной.
- **Непересекающиеся классы**.
- **Пересекающиеся классы**. Объект может относиться одновременно к нескольким классам.
- **Нечёткие классы**. Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Линейная сепарабельность

Два множества точек в двумерном пространстве называются линейно сепарабельными (линейно делимыми), если они могут быть полностью отделены единственной прямой. Для n -мерного пространства два набора точек линейно делимы, если они могут быть отделены $(n-1)$ -мерной гиперплоскостью.



В математических терминах: пусть X_0 и X_1 — два множества точек в n -мерном пространстве. Тогда X_0 и X_1 линейно разделимы, если существует $n+1$ действительных чисел w_1, w_2, \dots, w_{n+1} , таких, что каждая точка $x \in X_0$ удовлетворяет

$$\sum_{i=1}^n w_i x_i \geq w_{n+1}$$

и каждая точка $x \in X_1$ удовлетворяет

$$\sum_{i=1}^n w_i x_i < w_{n+1}$$

где x_i — i -й компонент x .

Линейный классификатор

Линейный классификатор — способ решения задач классификации, когда решение принимается на основании линейного оператора над входными данными. Класс задач, которые можно решать с помощью линейных классификаторов, обладают, соответственно, свойством линейной сепарабельности.

Пусть вектор \mathbf{x} из действительных чисел представляет собой входные данные, а на выходе классификатора вычисляется показатель y по формуле:

$$y = f(\vec{w} \cdot \vec{x}) = f\left(\sum_j w_j x_j\right)$$

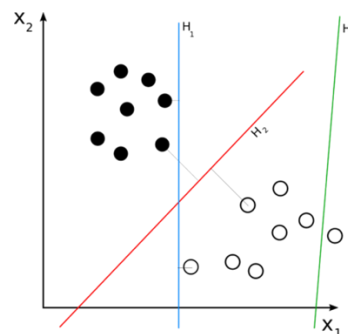
здесь ω - действительный вектор весов, а f - функция преобразования скалярного произведения.

Иными словами, вектор весов ω - ковариантный вектор или линейная форма отображения \mathbf{x} в R . Значения весов вектора ω определяются в ходе машинного обучения на подготовленных образцах. Функция f обычно простая пороговая функция,

отделяющая один класс объектов от другого. В более сложных случаях функция f имеет смысл вероятности того или иного решения.

Операцию линейной классификации для двух классов можно себе представить как отображение объектов в многомерном пространстве на гиперплоскость, в которой те объекты, которые попали по одну сторону разделяющей линии, относятся к первому классу ("да"), а объекты по другую сторону - ко второму классу ("нет").

На картинке множества чёрных и белых шаров разделяются синей и красной линией. При этом красная линия проводит более точную классификацию, потому что **она максимально отстоит от обоих множеств**. Зелёная линия не является линейным классификатором, она не разделяет два множества.



Генеративная и дискриминативная модели

Существует два подхода к определению параметров ω для линейного классификатора - *генеративные* или *дискриминативные* модели.

Генеративная модель использует условное распределение $P(x|\text{class})$ Например:

- **Дискриминантный анализ (LDA)** — предполагает нормальное распределение по Гауссу.
- **Наивный байесовский классификатор** с Бернуллиевской моделью событий.

Дискриминативные модели стремятся улучшить качество выходных данных на наборе образцов для обучения. Например:

- **Логистическая регрессия** — стремление достичь максимального сходства через вектор ω из предположения, что наблюдаемый набор образцов генерировался в виде биномиальной модели от выходных данных.
- **Простой Перцептрон** — алгоритм коррекции всех ошибок на входном наборе образцов.
- **Метод опорных векторов** — алгоритм расширения разделительной зоны в гиперплоскости решений между образцами входных данных.

Дискриминантный анализ

Дискриминативные модели более точны, однако при неполной информации в данных легче использовать условное распределение.

Дискриминантный анализ — раздел вычислительной математики, представляющий набор методов статистического анализа для решения задач распознавания образов, который используется для принятия решения о том, какие переменные разделяют (то есть «дискриминируют») возникающие наборы данных (так называемые «группы»). В отличие от кластерного анализа в дискриминантном анализе группы известны априори.

С вычислительной точки зрения дискриминантный анализ очень похож на дисперсионный анализ (см. раздел Дисперсионный анализ). Рассмотрим следующий простой пример. Предположим, что вы измеряете рост в случайной выборке из 50 мужчин и 50 женщин. Женщины в среднем не так высоки, как мужчины, и эта разница должна найти отражение для каждой группы средних (для переменной Рост). Поэтому переменная Рост позволяет вам провести дискриминацию между мужчинами и женщинами лучше, чем, например, вероятность, выраженная следующими словами: "Если человек большой, то это, скорее всего, мужчина, а если маленький, то это вероятно женщина".

Основная идея дискриминантного анализа заключается в том, чтобы определить, отличаются ли совокупности по среднему какой-либо переменной (или линейной комбинации переменных), и затем использовать эту переменную, чтобы предсказать для новых членов их принадлежность к той или иной группе.

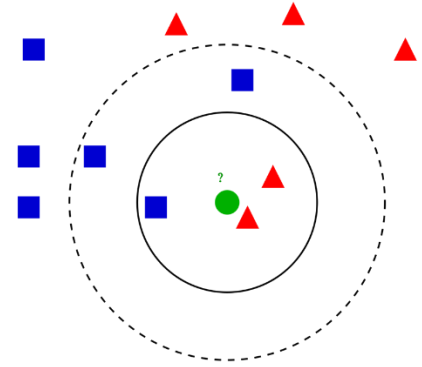
Метод k-ближайших соседей

Метод k-ближайших соседей (*k-nearest neighbors algorithm*, k-NN) — метрический алгоритм для автоматической классификации объектов или регрессии.

В случае использования метода для классификации объект присваивается тому классу, который является наиболее распространённым среди k соседей данного элемента, классы которых уже известны. В случае использования метода для регрессии, объекту присваивается среднее значение по k ближайшим к нему объектам, значения которых уже известны.

Алгоритм может быть применим к выборкам с большим количеством атрибутов (многомерным). Для этого перед применением нужно определить функцию расстояния; классический вариант такой функции — евклидова метрика.

Пример классификации k-ближайших соседей. Тестовый образец (зелёный круг) должен быть классифицирован как синий квадрат (класс 1) или как красный треугольник (класс 2). Если $k = 3$, то он классифицируется как 2-й класс, потому что внутри меньшего круга 2 треугольника и только 1 квадрат. Если $k = 5$, то он будет классифицирован как 1-й класс (3 квадрата против 2 треугольников внутри большего круга)



Выделение значимых атрибутов

Некоторые значимые атрибуты могут быть важнее остальных, поэтому для каждого атрибута может быть задан в соответствии определённый вес (например, вычисленный с помощью тестовой выборки и оптимизации ошибки отклонения). Таким образом, каждому атрибуту k будет задан в соответствии вес z_k , так что значение атрибута будет попадать в диапазон $[0; z_{k\max}(k)]$ (для нормализованных значений по минимакс-методу). Например, если атрибуту присвоен вес 2,7, то его нормализованно-взвешенное значение будет лежать в диапазоне $[0; 2,7]$

Взвешенный способ

При взвешенном способе во внимание принимается не только количество попавших в область определённых классов, но и их удалённость от нового значения.

Для каждого класса j определяется оценка близости:

$$Q_j = \sum_{i=1}^n \frac{1}{d(x, a_i)^2}$$

где $d(x, a_i)$ — расстояние от нового значения x до объекта a_i .

У какого класса выше значение близости, тот класс и присваивается новому объекту.

С помощью метода можно вычислять значение одного из атрибутов классифицируемого объекта на основании дистанций от попавших в область объектов и соответствующих значений этого же атрибута у объектов:

$$x_k = \frac{\sum_{i=1}^n k_i d(x, a_i)^2}{\sum_{i=1}^n d(x, a_i)^2}$$

где

a_i — i -ый объект, попавший в область,

k_i — значение атрибута k у заданного объекта a_i ,

x — новый объект,

x_k — k -ый атрибут нового объекта.

Наивный байесовский классификатор

Наивный байесовский классификатор — простой вероятностный классификатор, основанный на применении теоремы Байеса со строгими (наивными) предположениями о независимости.

В зависимости от точной природы вероятностной модели, наивные байесовские классификаторы могут обучаться очень эффективно. Во многих практических приложениях для оценки параметров для наивных байесовских моделей используют метод максимального правдоподобия; другими словами, можно работать с наивной байесовской моделью, не веря в байесовскую вероятность и не используя байесовские методы.

Несмотря на наивный вид и, несомненно, очень упрощенные условия, наивные байесовские классификаторы часто работают намного лучше нейронных сетей во многих сложных жизненных ситуациях.

Достоинством наивного байесовского классификатора является малое количество данных, необходимых для обучения, оценки параметров и классификации.

Как работает наивный байесовский алгоритм?

Давайте рассмотрим пример. Ниже представлен обучающий набор данных, содержащий один признак «Погодные условия» (weather) и целевую переменную «Игра» (play), которая обозначает возможность проведения матча. На основе погодных условий мы должны определить, состоится ли матч. Чтобы сделать это, необходимо выполнить следующие шаги.

- Шаг 1. Преобразуем набор данных в частотную таблицу (frequency table).
- Шаг 2. Создадим таблицу правдоподобия (likelihood table), рассчитав соответствующие вероятности. Например, вероятность облачной погоды (overcast) составляет 0,29, а вероятность того, что матч состоится (yes) — 0,64.
- Шаг 3. С помощью теоремы Байеса рассчитаем апостериорную вероятность для каждого класса при данных погодных условиях. Класс с наибольшей апостериорной вероятностью будет результатом прогноза.

Задача. *Состоится ли матч при солнечной погоде (sunny)?*

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table			
Weather	No	Yes	
Overcast		4	=4/14 0.29
Rainy	3	2	=5/14 0.36
Sunny	2	3	=5/14 0.36
All	5	9	
	=5/14	=9/14	
	0.36	0.64	

Sunny – Солнечная погода. *Rainy* – Дождливая погода. *Overcast* – Облачная погода

Мы можем решить эту задачу с помощью описанного выше подхода.

$$P(\text{Yes} \mid \text{Sunny}) = P(\text{Sunny} \mid \text{Yes}) * P(\text{Yes}) / P(\text{Sunny})$$

Здесь мы имеем следующие значения:

- $P(\text{Sunny} \mid \text{Yes}) = 3 / 9 = 0,33$
- $P(\text{Sunny}) = 5 / 14 = 0,36$
- $P(\text{Yes}) = 9 / 14 = 0,64$

Теперь рассчитаем $P(\text{Yes} \mid \text{Sunny})$:

- $P(\text{Yes} \mid \text{Sunny}) = 0,33 * 0,64 / 0,36 = 0,60$

Значит, при солнечной погоде более вероятно, что матч состоится.

Аналогичным образом с помощью НБА можно прогнозировать несколько различных классов на основе множества признаков. Этот алгоритм в основном используется в области классификации текстов и при решении задач многоклассовой классификации.

Распознавания образов

Распознавание образов — это отнесение исходных данных к определённому классу с помощью выделения существенных признаков, характеризующих эти данные, из общей массы данных.

При постановке задач распознавания стараются пользоваться математическим языком, стремясь — в отличие от теории искусственных нейронных сетей, где основой является получение результата путём эксперимента, — заменить эксперимент логическими рассуждениями и математическими доказательствами.

Классическая постановка задачи распознавания образов:

- Дано множество объектов. Относительно них необходимо провести классификацию. Множество представлено подмножествами, которые называются классами.
- Заданы: информация о классах, описание всего множества и описание информации об объекте, принадлежность которого к определённому классу неизвестна.
- Требуется по имеющейся информации о классах и описании объекта установить — к какому классу относится этот объект.

Наиболее часто в задачах распознавания образов рассматриваются монохромные изображения, что дает возможность рассматривать изображение как функцию на плоскости. Если рассмотреть точечное множество на плоскости T , где функция $f(x,y)$ выражает в каждой точке изображения его характеристику — яркость, прозрачность, оптическую плотность, то такая функция есть формальная запись изображения.

Множество же всех возможных функций $f(x,y)$ на плоскости T есть модель множества всех изображений X . Вводя понятие сходства между образами можно поставить задачу распознавания. Конкретный вид такой постановки сильно зависит от последующих этапов при распознавании в соответствии с тем или иным подходом.

Методы распознавания графических образов

- Для оптического распознавания образов можно применить метод перебора вида объекта под различными углами, масштабами, смещениями и т. д. Для букв нужно перебирать шрифт, свойства шрифта и т. д.
- Второй подход — найти контур объекта и исследовать его свойства (связность, наличие углов и т. д.)
- Ещё один подход — использовать искусственные нейронные сети. Этот метод требует либо большого количества примеров задачи распознавания (с правильными ответами), либо специальной структуры нейронной сети, учитывающей специфику данной задачи.

Направления в распознавании образов

- Изучение способностей к распознаванию, которыми обладают живые существа, объяснение и моделирование их;
- Развитие теории и методов построения устройств, предназначенных для решения отдельных задач в прикладных целях.

Примеры задач распознавания образов

- Оптическое распознавание символов
- Распознавание штрих-кодов
- Распознавание автомобильных номеров
- Распознавание лиц
- Распознавание речи
- Распознавание изображений
- Распознавание локальных участков земной коры, в которых находятся месторождения полезных ископаемых
- Классификация документов

Деревья решений

Один из самых популярных методов классификации и регрессии – деревья решений. Деревья решений используются в повседневной жизни в самых разных областях человеческой деятельности, порой и очень далеких от машинного обучения. Деревом решений можно назвать наглядную инструкцию, что делать в какой ситуации. Приведем пример из области консультирования научных сотрудников института. Высшая Школа Экономики выпускает инфо-схемы, облегчающие жизнь своим сотрудникам. Вот фрагмент инструкции по публикации научной статьи на портале института.



В терминах машинного обучения можно сказать, что это элементарный классификатор, который определяет форму публикации на портале (книга, статья, глава книги, препринт, публикация в "НИУ ВШЭ и СМИ") по нескольким признакам: типу публикации (монография, брошюра, статья и т.д.), типу издания, где опубликована статья (научный журнал, сборник трудов и т.д.) и остальным.

Зачастую дерево решений служит обобщением опыта экспертов, средством передачи знаний будущим сотрудникам или моделью бизнес-процесса компании.

Например, до внедрения масштабируемых алгоритмов машинного обучения в банковской сфере задача кредитного скоринга решалась экспертами. Решение о выдаче кредита заемщику принималось на основе некоторых интуитивно (или по опыту) выведенных правил, которые можно представить в виде дерева решений.

В этом случае можно сказать, что решается задача бинарной классификации (целевой класс имеет два значения: "Выдать кредит" и "Отказать") по признакам "Возраст", "Наличие дома", "Доход" и "Образование".



Дерево решений как алгоритм машинного обучения – по сути то же самое: объединение логических правил вида "**Значение признака a меньше x И Значение признака b меньше y... => Класс 1**" в структуру данных "Дерево". Огромное преимущество деревьев решений в том, что они легко интерпретируемы, понятны человеку. Например, по схеме на рисунке выше можно объяснить заемщику, почему ему было отказано в кредите. Скажем, потому что у него нет дома и доход меньше 5000. Как мы увидим дальше, многие другие, хоть и более точные, модели не обладают этим свойством и могут рассматриваться скорее как "черный ящик", в который загрузили данные и получили ответ.

В связи с этой "понятностью" деревьев решений и их сходством с моделью принятия решений человеком (можно легко объяснять боссу свою модель), деревья решений получили огромную популярность, а один из представителей этой группы методов классификации, C4.5, рассматривается первым в списке 10 лучших алгоритмов интеллектуального анализа данных.

Деревья принятия решений являются удобным инструментом в тех случаях, когда требуется не просто классифицировать данные, но ещё и объяснить почему тот или иной объект отнесён к какому-либо классу.

Лекция 7. Кластеризация

Древняя китайская классификация животных

Животные подразделяются на: а) принадлежащих императору; б) набальзамированных; в) дрессированных; г) молочных поросят; д) сирен; е) сказочных; ж) бродячих собак; з) включенных в данную классификацию; и) дрожащих, как сумасшедшие к) неисчислимым; л) нарисованных самой лучшей китайской кисточкой;

м) других; н) тех, которые только что разбили цветочную вазу и о) тех, которые издавна напоминают мух

(Хорхе Луис Борхес, Другие исследования: 1937-1952).

Задача кластеризации сходна с задачей классификации, является ее логическим продолжением, но ее отличие в том, что классы изучаемого набора данных заранее не предопределены.

Синонимами термина "кластеризация" являются "автоматическая классификация", "обучение без учителя" и "таксономия".

Кластеризация предназначена для разбиения совокупности объектов на однородные группы (кластеры или классы). Если данные выборки представить как точки в признаковом пространстве, то задача кластеризации сводится к определению "сгущений точек".

Кластерный анализ объединяет различные процедуры, используемые для проведения кластеризации. В результате применения данных процедур исходная совокупность объектов разделяется на кластеры или группы (классы) сходных между собой объектов.

Список прикладных областей, где она применяется, широк: сегментация изображений, маркетинг, борьба с мошенничеством, прогнозирование, анализ текстов и многие другие. На современном этапе кластеризация часто выступает первым шагом при анализе данных. После выделения схожих групп применяются другие методы, для каждой группы строится отдельная модель. Аналитику часто легче выделить группы схожих объектов, изучить их особенности и построить для каждой группы отдельную модель, чем создавать одну общую модель на всех данных. Таким приемом постоянно пользуются в маркетинге, выделяя группы клиентов, покупателей, товаров и разрабатывая для каждой из них отдельную стратегию.

Следует отметить, что в результате применения различных методов кластерного анализа могут быть получены кластеры различной формы. Например, возможны кластеры "цепочного" типа, когда кластеры представлены длинными "цепочками", кластеры удлиненной формы и т.д., а некоторые методы могут создавать кластеры произвольной формы.

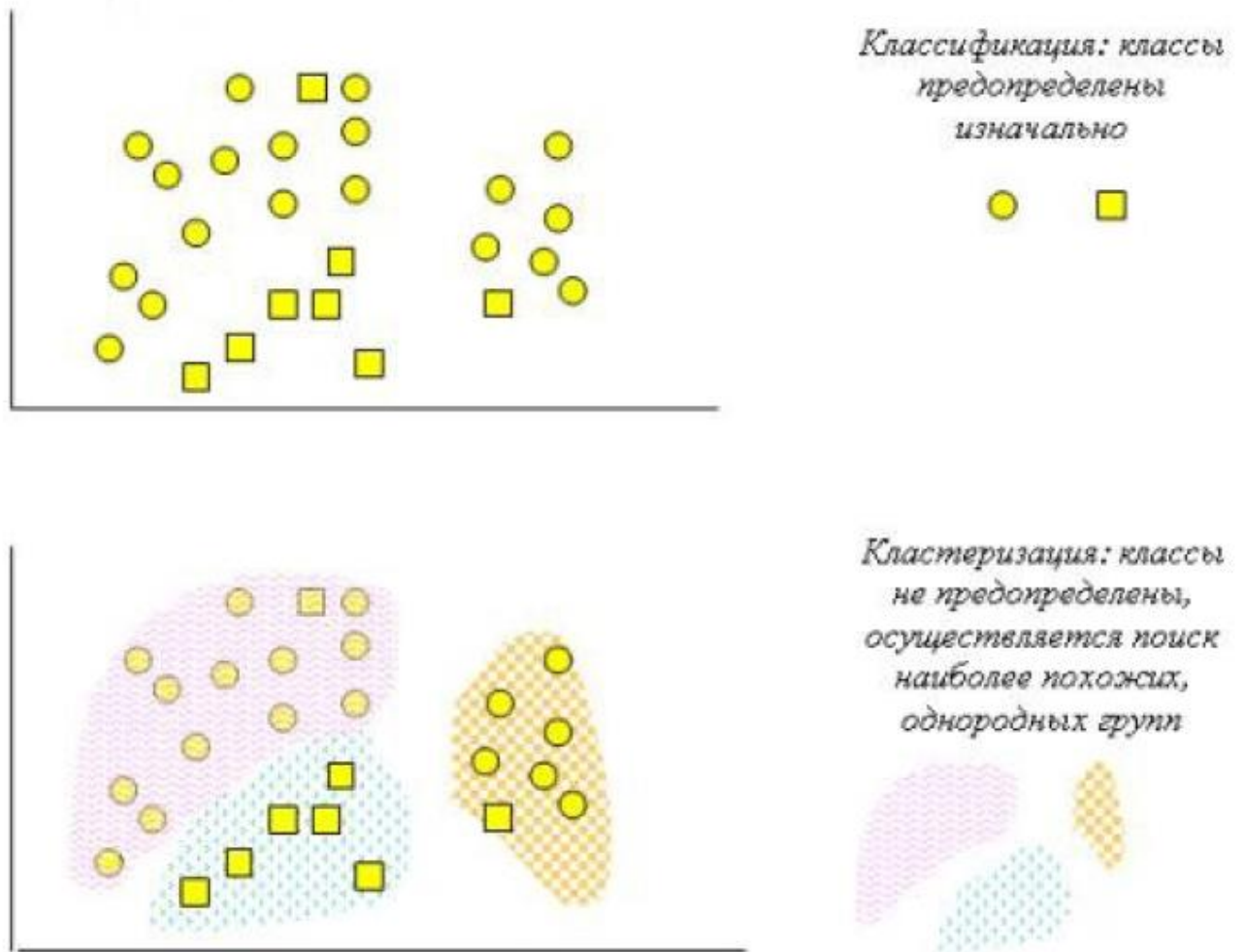


Рис. 7.1 Сравнение задач классификации и кластеризации

Кластеры могут быть непересекающимися, или эксклюзивными (non-overlapping, exclusive), и пересекающимися (overlapping). Схематическое изображение непересекающихся и пересекающихся кластеров дано на рис. 7.2.

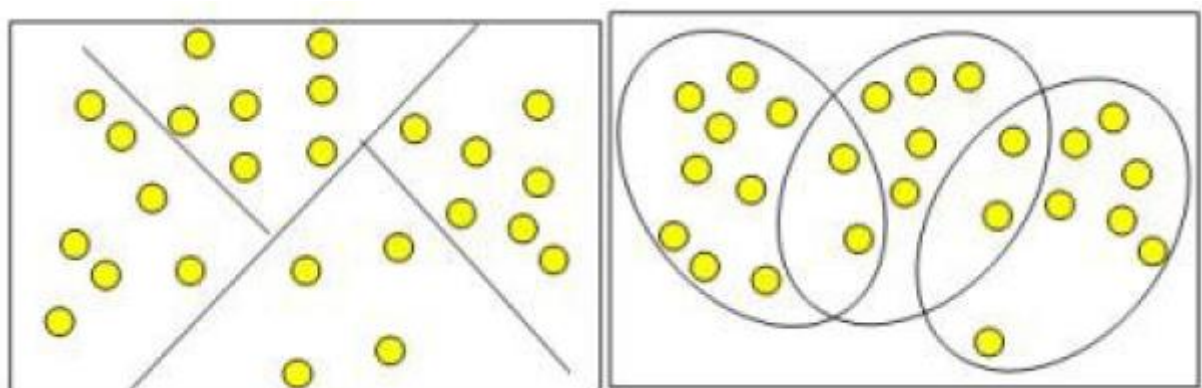


Рис. 7.2 Непересекающиеся и пересекающиеся кластеры

Говоря о кластерной совокупности объектов подразумеваем, что каждый из них задан соответствующей строкой матрицы X (ТЭД), либо задана матрица попарных расстояний объектов. Тогда проблема кластеризации заключается в том, чтобы всю анализируемую совокупность объектов $O=\{O_1, O_2, \dots, O_n\}$, статистически представленную в виде матриц X или P , разбить на сравнительно небольшое число (заранее известное или нет) однородных в определенном смысле, групп или классов.

Полученные в результате разбиения классы часто называют кластерами (таксонами, образами), а методы их нахождения – соответственно кластер-анализом, численной таксономией, распознаванием образов с самообучением.

Перед началом кластеризации каждый исследователь должен решить для себя, какую из двух задач он решает:

1. Обычная задача разбиения на гиперобласти группирования.
2. Определение естественного расслоения исходных наблюдений на четко выраженные кластеры. Каждый кластер не разбивается на столь же удаленные части.

Если первая задача всегда имеет решение, то во втором случае результат может быть и отрицательным. Например, множество исходных наблюдений образует один общий кластер.

Можно выделить следующие этапы кластерного анализа:

1. Получение выборки для кластеризации.
2. Определение множества признаков, по которым будут определяться объекты.
3. Вычисление значений меры сходства между объектами.
4. Применение методов кластерного анализа для создания групп.
5. Проверка достоверности результатов.

Меры сходства

То, что некоторые вещи обнаруживают между собой сходство или различие, является весьма важным моментом для кластеризации.

Выделяют четыре группы коэффициентов сходства:

- Коэффициенты корреляции
- Меры расстояния
- Коэффициенты ассоциативности
- Вероятные коэффициенты сходства.

Наиболее широкое распространение получили меры расстояния, которые мы и рассмотрим.

На практике меры расстояния точнее было бы назвать мерами несходства: если для большинства используемых коэффициентов большие значения соответствуют большему сходству, то для мер расстояния все наоборот.

Количественное оценивание сходства отталкивается от понятия метрики. При таком подходе объекты представляются точками пространства, расстояние между которыми и определяет их сходство или различие.

Выделим такие понятия как «расстояние между отдельными объектами» (мера близости объектов) и «расстояние между классами» (мера близости классов) и рассмотрим их отдельно.

Расстояние между отдельными объектами.

Выбор метрики (или меры близости) является узловым моментом исследования, от которого во многом зависит окончательный вариант разбиения объектов на классы.

В каждой конкретной задаче этот выбор должен производиться по-своему и опираться на физическую, статистическую природу вектора наблюдений X .

Рассмотрим некоторые примеры расстояний и мер близости.

- Расстояние Махаланобиса.

В общем случае зависимых компонент $x^{(1)} \dots x^{(p)}$ вектора наблюдений X и их различной значимости в решении вопроса об отнесении объекта к тому или иному классу обычно пользуются взвешенным расстоянием Махаланобиса.

$$d_0(x_i, x_j) = \sqrt{(x_i - x_j)' \Sigma^{-1} (x_i - x_j)}$$

Здесь

Σ - ковариационная матрица генеральной совокупности, из которой извлекается матрица наблюдений X ;

Σ^{-1} – матрица, обратная ковариационной.

\wedge - симметричная матрица весовых коэффициентов X_{mq} набора, чаще всего выбирается диагональной.

Когда корреляция между переменными равна нулю, расстояние Махаланобиса эквивалентно евклидову расстоянию.

Такое расстояние между объектами удобно тем, что оно инвариантно по отношению к линейным преобразованиям признакового пространства, т.е. любое линейное преобразование (растяжение, сдвиг, поворот и т. п.) оставляет расстояние Махаланобиса между объектами неизменным. Это расстояние применяют в том случае, если признаки $x(1), x(2), \dots, x(p)$ значительно коррелированы между собой, а также, если признаки имеют различную значимость для классификации.

- Обычное евклидово расстояние

$$d_E(x_i, x_j) = \sqrt{\sum_{k=1}^p \left(x_i^{(k)} - x_j^{(k)} \right)^2}$$

- Взвешенное евклидово расстояние. Формула для вычисления евклидова расстояния между двумя объектами в пространстве p переменных с учетом коэффициентов важности этих переменных преобразуется к виду:

$$d_{bE}(x_i, x_j) = \sqrt{\sum_{k=1}^p w_k \left(x_i^{(k)} - x_j^{(k)} \right)^2}$$

Обычно применяется в ситуациях, в которых так или иначе удается приписать каждой из компонент $X^{(k)}$ некоторый неотрицательный «вес» w_k , пропорциональный степени его важности.

a) Хеммингово расстояние

$$d_H(x_i, x_j) = \sum_{k=1}^p \left| x_i^{(k)} - x_j^{(k)} \right|$$

Это расстояние используется в основном как мера различия объектов, признаки которых измерены в шкалах наименований и порядка. В случае дихотомических признаков расстояние Хэмминга показывает число несовпадающих у объектов признаков.

Первоначально данная мера была введена для бинарных кодов, но она вполне применима и для сравнений любых упорядоченных наборов, которые состоят из элементов принимающих дискретные значения.

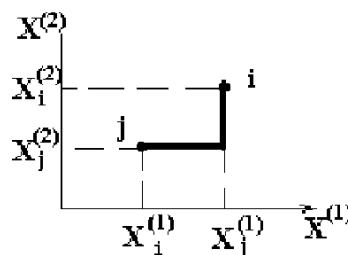
b) Расстояние Минковского

$$d_m(x_i, x_j) = \left(\sum_{k=1}^p \left| x_i^{(k)} - x_j^{(k)} \right|^r \right)^{1/r}$$

при $r=1$ имеем манхэттенское расстояние, при $r=2$ – евклидово.

c) Манхэттенское расстояние или расстояние городских кварталов

$$d(x_i, x_j) = \sum_{k=1}^p \left| x_i^{(k)} - x_j^{(k)} \right|$$



d) Расстояние Чебышева:

$$\text{distance}(x, c) = \text{Maximum } |x_i - c_i|$$

Расстояние между классами

Для проведения кластеризации необходимо также ввести понятие меры близости двух групп объектов.

Пусть:

S_i – i -ая группа (класс, классификатор);

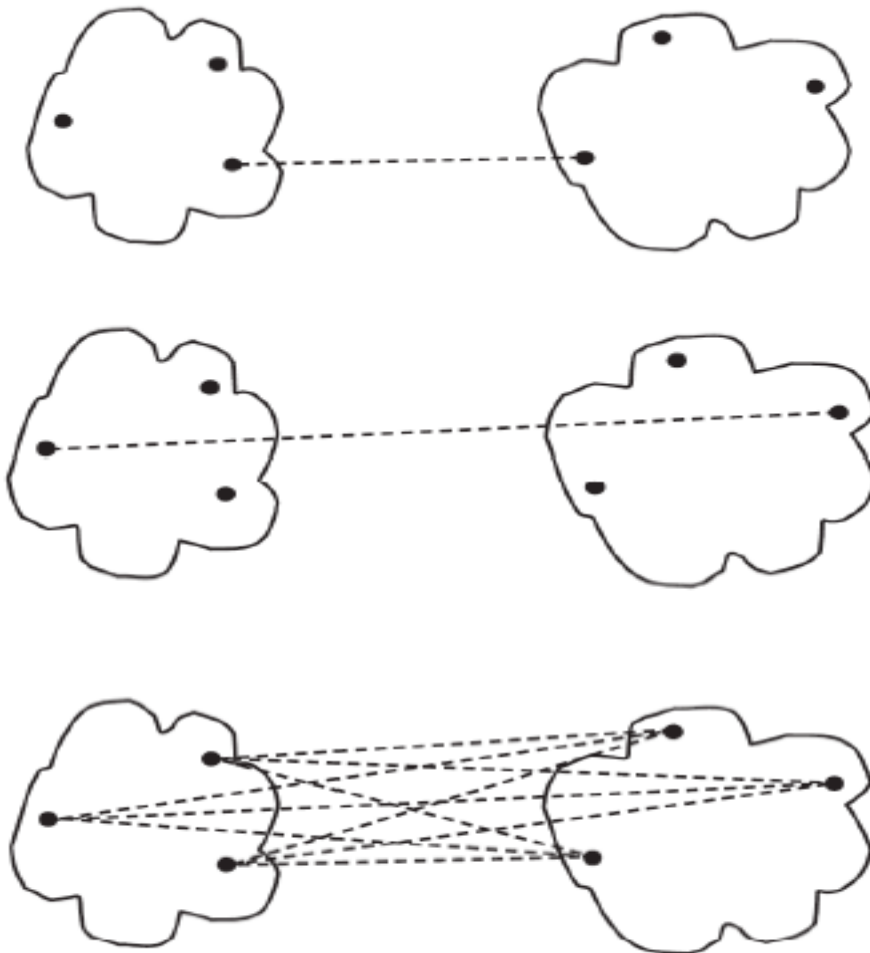
n_j – число объектов, образующих группу S_j ;

$\bar{X}(i)$ – среднее арифметическое вектора наблюдений, т.е. «центр тяжести» i -й группы;

$\rho(S_l, S_m)$ – расстояние между группами.

а) Расстояние, измеряемое по принципу «ближнего соседа».

$$\rho_{\min}(S_l, S_m) = \min_{\substack{x_i \in S_l, x_j \in S_m}} d(x_i, x_j)$$



Расстояния между классами а), б) и с)

b) Расстояние, измеряемое по принципу «дальнего соседа».

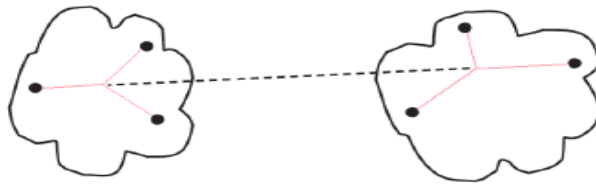
$$\rho_{\max}(S_l, S_m) = \max_{x_i \in S_l, x_j \in S_m} d(x_i, x_j)$$

c) Расстояние, измеряемое по принципу «средней связи».

$$\rho_{\text{ср}}(S_l, S_m) = \frac{1}{n_l n_m} \sum_{x_i \in S_l} \sum_{x_j \in S_m} d(x_i, x_j)$$

d) Расстояние, измеряемое по центрам тяжести групп.

$$\rho(S_l, S_m) = d(\bar{X}(l), \bar{X}(m))$$

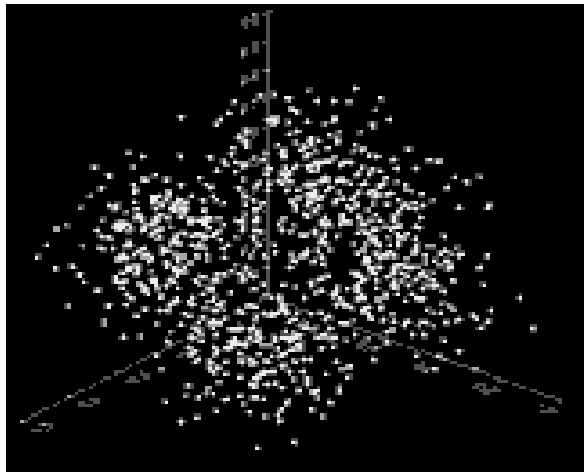


Определяется как арифметическое среднее всевозможных попарных расстояний между представителями рассматриваемых групп.

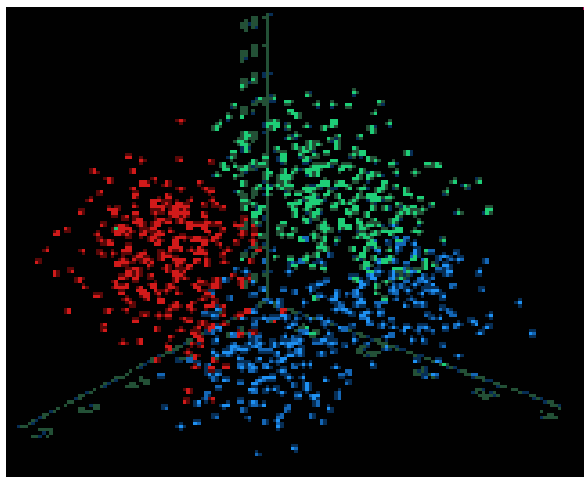
Кластер и кластеризация

Кластер — это множество объектов, близких между собой по некоторой мере сходства. В пространстве переменных кластеры представляют собой скопления точек (объектов) различной формы.





Так выглядят объекты до кластеризации



Объекты после кластеризации

Типы кластерных структур



*Внутрикластерные расстояния,
как правило, меньше межкластер-
ных*

ленточные кластеры

кластеры с центром

Обзор методов кластерного анализа

Следует отметить, что в результате применения различных методов кластерного анализа могут быть получены кластеры различной формы. Например, возможны кластеры "цепочного" типа, когда кластеры представлены длинными "цепочками", кластеры удлиненной формы и т.д., а некоторые методы могут создавать кластеры произвольной формы.

Различные методы могут стремиться создавать кластеры определенных размеров (например, малых или крупных) либо предполагать в наборе данных наличие кластеров различного размера.

Некоторые методы кластерного анализа особенно чувствительны к шумам или выбросам, другие - менее.

В результате применения различных методов кластеризации могут быть получены неодинаковые результаты, это нормально и является особенностью работы того или иного алгоритма.

Данные особенности следует учитывать при выборе метода кластеризации. На сегодняшний день разработано более сотни различных алгоритмов кластеризации.

Кластера – это непрерывные области некоторого пространства, с относительно высокой плотностью точек, отделенных от других таких же областей областями с относительно низкой плотностью точек.

Наиболее важными свойствами кластеров являются плотность, дисперсия, размеры, форма и отделимость. Отделимость характеризует степень перекрытия кластеров и насколько далеко они расположены в пространстве.

Разработанные кластерные методы образуют 6 основных семейств:

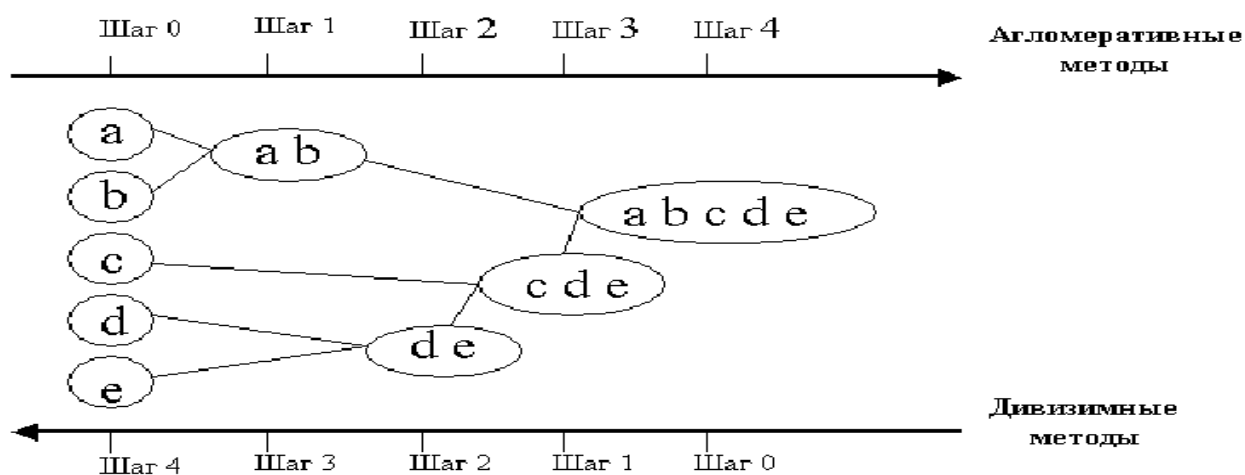
1. Иерархические

- a. агломеративные;
 - b. дивизимные;
2. Итеративные методы группировки;
 3. Методы поиска модальных значений плотности;
 4. Факторные методы;
 5. Методы сгущений;
 6. Методы, использующие теорию графов.

К наиболее распространенным методам решения задачи кластеризации относят:

- метод k-средних (работает только с числовыми атрибутами),
- иерархический кластерный анализ (работает также с символьными атрибутами),
- метод SOM.

Сложностью кластеризации является необходимость ее оценки.

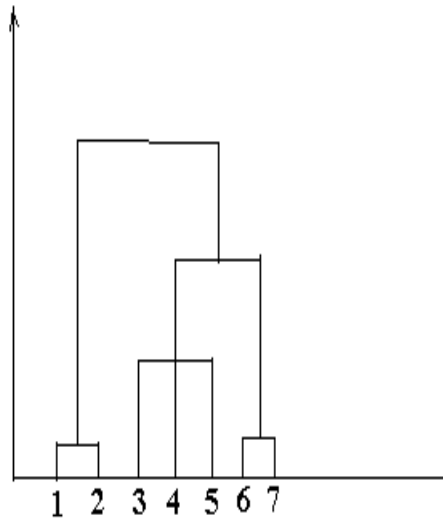


Дендрограмма агломеративных и дивизимных методов

Агломеративная иерархическая кластеризация

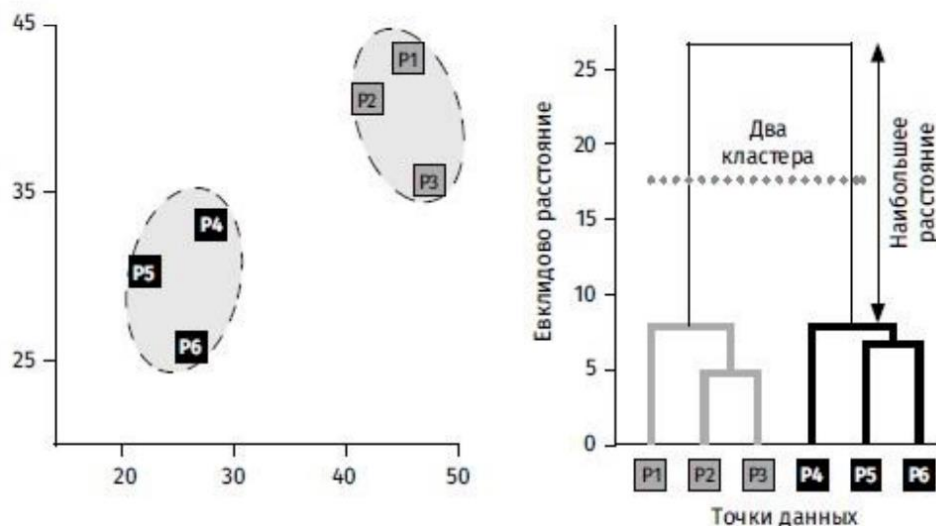
Рассмотрим более подробно метод иерархической классификации.

1. Иерархические агломеративные методы просматривают матрицу сходства размерностью $N \times N$ (где N – число объектов) и последующую, объединяют схожие объекты. Именно поэтому они называются агломеративными или объединяющими.
2. Последовательность объединений кластеров можно представить в виде древовидной диаграммы, называемой дендрограммой.



Для полной кластеризации этим методом на основе матрицы сходства размерность $N \times N$ требуется ровно $N-1$ шагов. На 1-ом шаге объекты рассматриваются как самостоятельные кластеры. На последнем шаге все события (объекты) объединяются в одну большую группу самостоятельные кластеры. На последнем шаге все события (объекты) объединяются в одну большую группу.

3. Для понимания иерархических агломеративных методов не нужны обширные знания.



Несмотря на простоту методов, они обладают некоторыми недостатками:

- Вычисление и хранение большой матрицы сходств (например, для 500 объектов потребуется матрица из 125000 элементов).

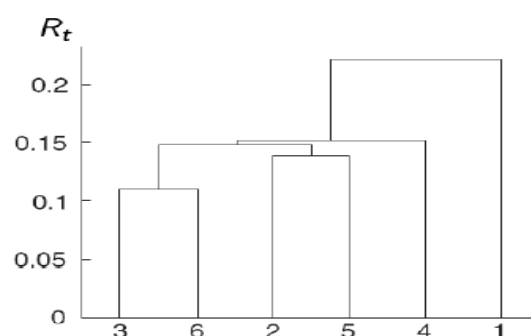
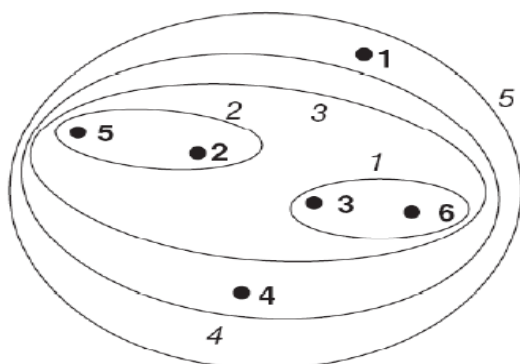
- Объекты распределяются по кластерам за один проход, поэтому плохое начальное разбиение множества данных не может быть изменено на последующих шагах процесса кластеризации.
- Все методы (за исключением метода одиночной связи) могут порождать разные решения в результате простого переупорядочивания объектов в матрице сходства и, кроме того, их результаты изменяются, если некоторые объекты исключить из рассмотрения. Устойчивость – важное свойство «естественной» группировки по сравнению с теми группами, которые исчезают, если некоторые объекты переупорядочены или исключены из анализа.

Иерархические алгомеративные методы различаются главным образом по правилам построения кластеров. Существует много различных правил группировки, каждое из которых порождает специфический иерархический метод.

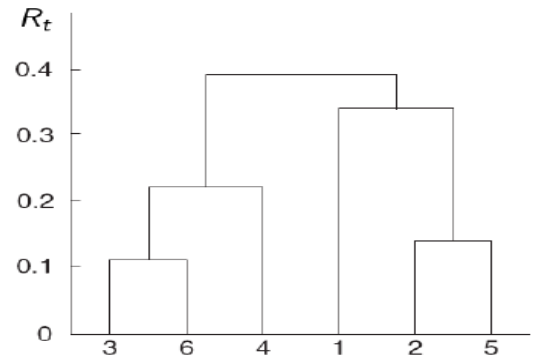
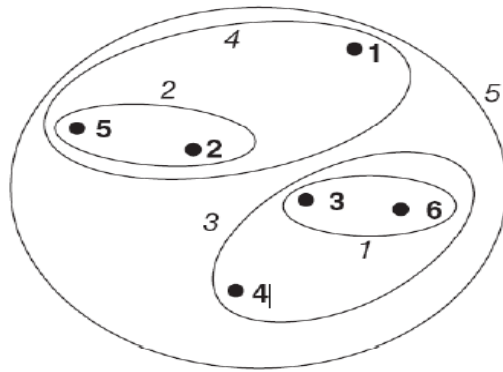
Рассмотрим наиболее распространенные:

1. метод одиночной связи;
2. метод полной связи;
3. метод средней связи;
4. метод Уорда (Ward).

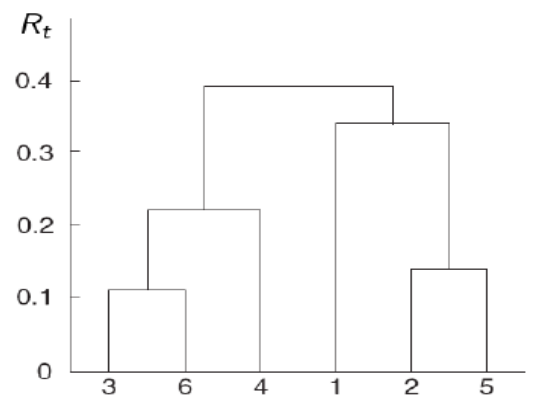
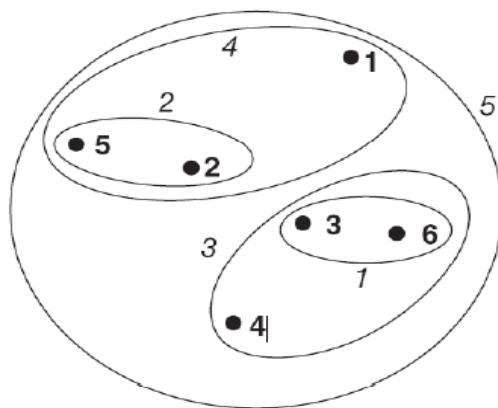
1. В *методе одиночной связи* на первом шаге объединяются два объекта, имеющие между собой максимальную меру сходства. На следующем шаге к ним присоединяется объект с максимальной мерой сходства с одним из объектов кластера, т.е. используется принцип «ближнего соседа». Таким образом процесс продолжается дальше. Итак, для включения объекта в кластеризацию требуется максимальное сходство с одним членом кластера. Отсюда и название метода одиночной связи. Нужна только одна связь, чтобы присоединить объект к кластеризации. Недостатки этого метода это образование слишком больших продолговатых кластеров.



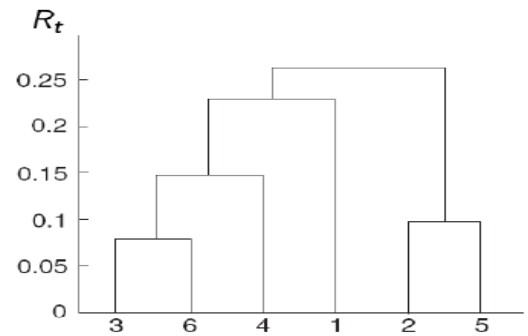
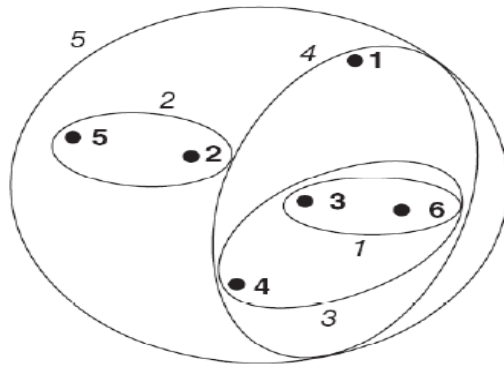
2. *Метод наиболее удаленных соседей или полной связи* позволяет устранить этот недостаток. В качестве правила построения кластеров используется принцип «дальней связи»: расстояние между объектом-кандидатом и самым дальним объектом кластера не должно превышать некоторого уровня.



3. В *методе средней связи* вычисляется среднее сходство рассматриваемого объекта со всеми объектами в одном существующем кластере, а затем, если найденное среднее значение сходства достигает или превосходит некоторый заданный пороговый уровень сходства, объект присоединяется к этому кластеру. В других вариантах метода средней связи вычисляется сходство между центрами тяжести двух кластеров, подлежащих объединению.



4. *Метод Уорда* (Ward's method). В качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения (Ward, 1963). В отличие от других методов кластерного анализа для оценки расстояний между кластерами, здесь используются методы дисперсионного анализа. На каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному увеличению целевой функции, т.е. внутригрупповой суммы квадратов. Этот метод направлен на объединение близко расположенных кластеров и "стремится" создавать кластеры малого размера.



Математические характеристики кластера

Кластер имеет следующие математические характеристики:

1. Центр кластера — это среднее геометрическое место точек в пространстве переменных.

$$x_{kj} = \frac{\sum_{i=1}^n w_j x_{ij}}{I_k}$$

2. Дисперсия кластера - это мера рассеяния точек в пространстве относительно центра кластера:

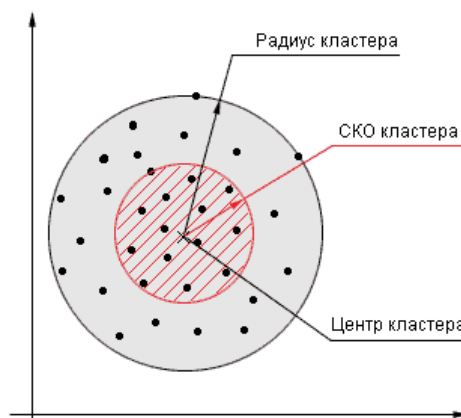
$$D_k = \frac{\sum_{i=1}^{I_1} \sum_{j=1}^n w_j (x_{ij} - \bar{x}_{kj})^2}{I_k - 1}$$

3. Среднеквадратичное отклонение (СКО) объектов относительно центра кластера:

$$S_k = \sqrt{D_k}$$

4. Радиус кластера - максимальное расстояние точек от центра кластера:

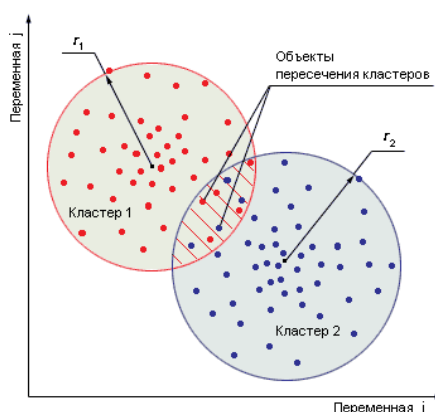
$$R_k = \max \sqrt{\sum_{j=1}^n w_j (x_{ij} - \bar{x}_{kj})^2}$$



Спорные объекты и перекрывающиеся кластеры

Спорный объект - это объект, который по мере сходства может быть отнесен к нескольким кластерам.

Размер кластера может быть определен либо по радиусу кластера, либо по среднеквадратичному отклонению объектов для этого кластера. Объект относится к кластеру, если расстояние от объекта до центра кластера меньше радиуса кластера. Если это условие выполняется для двух и более кластеров, объект является спорным. Неоднозначность данной задачи может быть устранена экспертом или аналитиком.



Методы кластерного анализа. Итеративные методы.

При большом количестве наблюдений иерархические методы кластерного анализа не пригодны. В таких случаях используют неиерархические методы, основанные на разделении, которые представляют собой итеративные методы дробления исходной совокупности. В процессе деления новые кластеры формируются до тех пор, пока не будет выполнено правило остановки.

Такая неиерархическая кластеризация состоит в разделении набора данных на определенное количество отдельных кластеров. Существует два подхода. Первый заключается в определении границ кластеров как наиболее плотных участков в многомерном пространстве исходных данных, т.е. определение кластера там, где имеется большое "сгущение точек". Вторым подходом является минимизация меры различия объектов.

Неиерархические алгоритмы основаны на оптимизации некоторой целевой функции, определяющей оптимальное в определенном смысле разбиение множества объектов на кластеры. В этой группе популярны алгоритмы семейства k-средних (k-means, fuzzy c-means, Густафсон-Кесселя), которые в качестве целевой функции используют сумму квадратов взвешенных отклонений координат объектов от центров искомых кластеров. Кластеры ищутся сферической либо эллипсоидной формы.

Алгоритм k-средних (k-means)

Наиболее распространен среди неиерархических методов алгоритм k -средних, также называемый **быстрым кластерным анализом**. В отличие от иерархических методов, которые не требуют предварительных предположений относительно числа кластеров, для возможности использования этого метода необходимо иметь гипотезу о наиболее вероятном количестве кластеров.

Алгоритм k -средних строит k кластеров, расположенных на возможно больших расстояниях друг от друга. Основной тип задач, которые решает алгоритм k -средних, - наличие предположений (гипотез) относительно числа кластеров, при этом они должны быть различны настолько, насколько это возможно. Выбор числа k может базироваться на результатах предшествующих исследований, теоретических соображениях или интуиции.

Общая идея алгоритма: заданное фиксированное число k кластеров наблюдения сопоставляются кластерам так, что средние в кластере (для всех переменных) максимально возможно отличаются друг от друга.

Описание алгоритма

1. Первоначальное распределение объектов по кластерам.

- Произвольно выбирается k объектов, как исходных центров кластеров. На первом шаге эти точки считаются "центрами" кластеров. Каждому кластеру соответствует один центр.
- Выбор начальных центроидов может осуществляться следующим образом:
 - выбор k -наблюдений для максимизации начального расстояния;
 - случайный выбор k -наблюдений;
 - выбор первых k -наблюдений.
- Каждый объект назначается ближайшему центроиду.

2. Итеративный процесс.

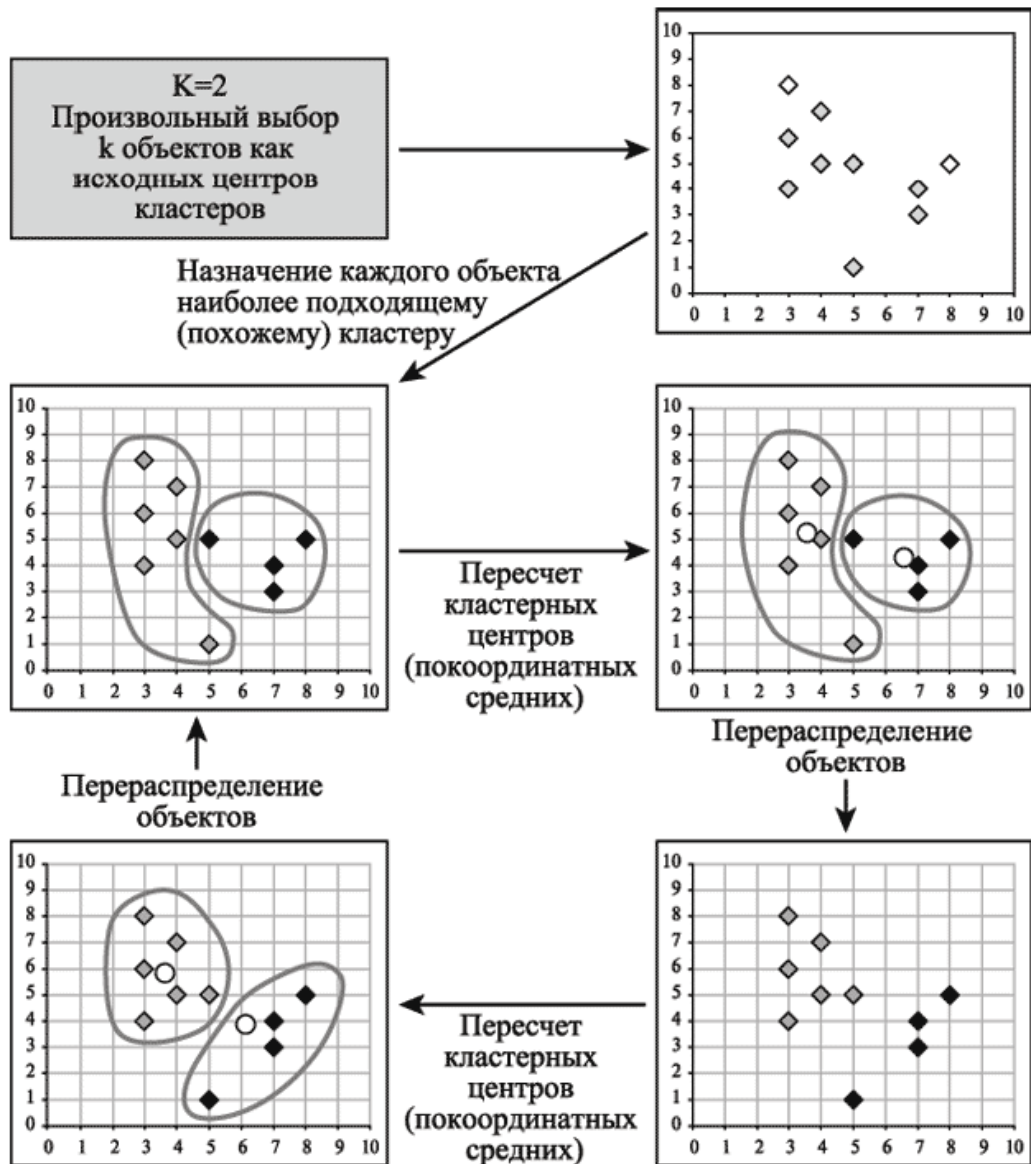
Вычисляются центры кластеров, которыми затем и далее считаются по координатные средние кластеров. Объекты опять перераспределяются.

Процесс вычисления центров и перераспределения объектов продолжается до тех пор, пока не выполнено одно из условий:

- кластерные центры стабилизировались, т.е. все наблюдения принадлежат кластеру, которому принадлежали до текущей итерации;
- число итераций равно максимальному числу итераций.

На рисунке приведен пример работы алгоритма k -средних для k , равного двум.

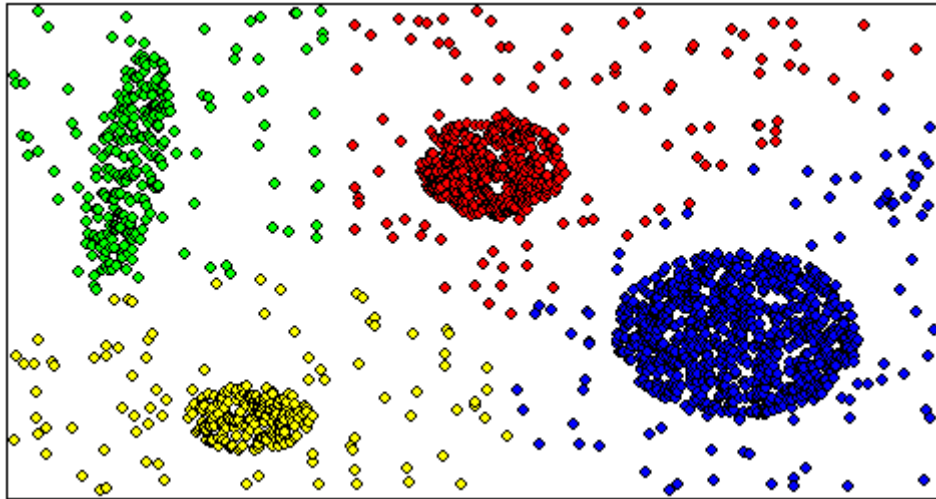
Выбор числа кластеров является сложным вопросом. Если нет предположений относительно этого числа, рекомендуют создать 2 кластера, затем 3, 4, 5 и т. д., сравнивая полученные результаты.



Проверка качества кластеризации методом k -средних

После получения результатов кластерного анализа методом k -средних следует проверить правильность кластеризации (т. е. оценить, насколько кластеры отличаются друг от друга).

Для этого рассчитываются средние значения для каждого кластера. При хорошей кластеризации должны быть получены сильно отличающиеся средние для всех измерений или хотя бы большей их части.

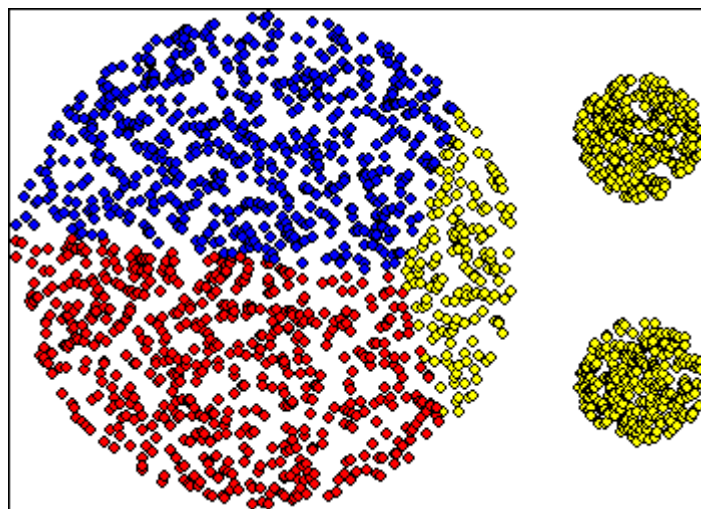


Результат кластеризации алгоритмом k-means

(<http://www.basegroup.ru/library/analysis/clusterization/datamining/>)

Алгоритм оптимизации целевой функции в неиерархических алгоритмах, основанных на расстояниях, носит итеративный характер, и на каждой итерации требуется рассчитывать матрицу расстояний между объектами. При большом числе объектов это неэффективно и требует серьезных вычислительных ресурсов. Вычислительная сложность 1й итерации алгоритма k-means оценивается как $O(kmn)$, где k, m, n – количество кластеров, атрибутов и объектов соответственно. Но итераций может быть очень много! Придется делать много проходов по набору данных.

Имеет массу недостатков в k-means сам подход с идеей поиска кластеров сферической или эллипсоидной формы. Подход хорошо работает, когда данные в пространстве образуют компактные сгустки, хорошо отличимые друг от друга. А если данные имеют вложенную форму, то ни один из алгоритмов семейства k-means никогда не справится с такой задачей. Также алгоритм плохо работает в случае, когда один кластер значительно больше остальных, и они находятся близко друг от друга – возникает эффект "расщепления" большого кластера.



Эффект расщепления большого кластера

Достоинства алгоритма k-средних:

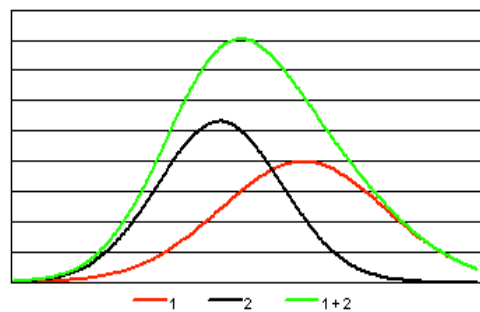
- простота использования;
- быстрота использования;
- понятность и прозрачность алгоритма.

Недостатки алгоритма k-средних:

- алгоритм слишком чувствителен к выбросам, которые могут исказить среднее. Возможным решением этой проблемы является использование модификации алгоритма - алгоритм k-медианы;
- алгоритм может медленно работать на больших базах данных. Возможным решением данной проблемы является использование выборки данных.

EM-алгоритм (Expectation-Maximization)

Среди неиерархических алгоритмов, не основанных на расстоянии, следует выделить EM-алгоритм (Expectation-Maximization). В нем вместо центров кластеров предполагается наличие функции плотности вероятности для каждого кластера с соответствующим значением математического ожидания и дисперсией. В смеси распределений (рис. 2) ведется поиск их параметров (средние и стандартные отклонения) по принципу максимума правдоподобия. Алгоритм EM и есть одна из реализаций такого поиска. Проблема заключается в том, что перед стартом алгоритма выдвигается гипотеза о виде распределений, которые оценить в общей совокупности данных сложно.



Распределения и их смесь

Еще одна проблема появляется тогда, когда атрибуты объекта смешанные – одна часть имеет числовой тип, а другая часть – категориальный. Например, пусть требуется вычислить расстояние между следующими объектами с атрибутами (Возраст, Пол, Образование):

- (1) {23, муж, высшее}
- (2) {25, жен, среднее}.

Первый атрибут является числовым, остальные – категориальными. Если мы захотим воспользоваться классическим иерархическим алгоритмом с какой-либо мерой сходства, нам придется каким-то образом произвести дискредитацию атрибута "Возраст". Например, так:

(1) {до 30 лет, муж, высшее}

(2) {до 30 лет, жен, среднее}.

При этом часть информации, мы, безусловно, потеряем. Если же мы будем определять расстояние в евклидовом пространстве, то возникнут вопросы с категориальными атрибутами. Понятно, что расстояние между "Пол муж" и "Пол жен" равно 0, т.к. значения этого признака находятся в шкале наименований. А атрибут "Образование" можно измерить как в шкале наименований, так и в шкале порядка, присвоив каждому значению определенные баллы. Какой вариант выбрать? А что делать, если категориальные атрибуты важнее числовых? Решение этих проблем ложится на плечи аналитика. Кроме того, при использовании алгоритма k-средних и ему подобных возникают трудности с пониманием центров кластеров у категориальных атрибутов, априорным заданием количества кластеров.

Оценка качества кластеризации

Оценка качества кластеризации может быть проведена на основе следующих процедур:

- ручная проверка;
- установление контрольных точек и проверка на полученных кластерах;
- определение стабильности кластеризации путем добавления в модель новых переменных;
- создание и сравнение кластеров с использованием различных методов. Разные методы кластеризации могут создавать разные кластеры, и это является нормальным явлением. Однако создание схожих кластеров различными методами указывает на правильность кластеризации.

Лекция 8. Поиск ассоциативных правил

В последнее время неуклонно растет интерес к методам обнаружения знаний в базах данных. Объемы современных баз данных, которые весьма внушительны, вызвали устойчивый спрос на новые масштабируемые алгоритмы анализа данных. Одним из популярных методов обнаружения знаний стали алгоритмы поиска ассоциативных правил.

Обучение ассоциативным правилам (далее Associations rules learning — ARL) или поиск ассоциативных правил— это метод обучения машин на базе правил обнаружения интересующих нас связей между переменными в большой базе данных. Метод предлагается для установления сильных правил, обнаруженных в базе данных с помощью некоторых мер интересности.

Этот основанный на правилах подход генерирует также новые правила по мере анализа дополнительных данных. Конечной целью, исходя из достаточно большого набора данных, помочь машине имитировать выделение признаков человеческим и создать возможность нахождения абстрактных ассоциаций из новых неклассифицированных данных.

Ассоциативные алгоритмы относятся к обучению без учителя и определяют вероятность того, что несколько предметов окажутся вместе в определённом наборе данных. В основном они используются при анализе рыночной корзины.

Самый распространённый среди них **алгоритм добычи данных Apriori**. Обычно он применяется в транзакционных базах данных. С его помощью можно найти часто встречающиеся наборы элементов, а затем сформировать из них определённые ассоциативные правила.

Например, если человек покупает молоко и хлеб, то, скорее всего, он также возьмёт яйца. Такие прогнозы строятся исходя из предыдущих покупок разных посетителей. После чего создаются ассоциативные правила в соответствии с конкретным показателем доверенности, который определяется алгоритмом на основе того, как часто встречаются эти товары вместе.

Транзакционная или операционная база данных (Transaction database) представляет собой двумерную таблицу, которая состоит из номера транзакции (TID) и перечня покупок, приобретенных во время этой транзакции.

TID	Приобретенные покупки
100	хлеб, молоко, печенье
200	Молоко, сметана
300	Молоко, хлеб, сметана, печенье
400	Колбаса, сметана
500	Хлеб, молоко, печенье, сметана

Исследование компании Teradata

В 1992 году группа по консалтингу в области ритейла компании Teradata провела исследование **1.2 миллиона транзакций в 25 магазинах** для ритейлера Osco Drug (Drug Store — формат разнокалиберных магазинов у дома). После анализа всех этих транзакций самым сильным правилом получилось **«Между 17:00 и 19:00 чаще всего пиво и подгузники покупают вместе»**. Такое правило показалось руководству Osco Drug настолько контринтуитивным, что ставить подгузники на полках рядом с пивом они не стали. Хотя объяснение паре пиво-подгузники вполне себе нашлось: когда оба члена молодой семьи возвращались с работы домой (как раз

часам к 5 вечера), жены обычно отправляли мужей за подгузниками в ближайший магазин. И мужа, недолго думая, совмещали приятное с полезным — покупали подгузники по заданию жены и пиво для собственного вечернего времяпрепровождения. Существует много мнений, насколько история истинна.

Кроме примера выше об анализе рыночной корзины, ассоциативные правила используются ныне во многих других областях, включая

- Web mining,
- обнаружение вторжений,
- непрерывное производство и
- биоинформатику.

Ассоциативные правила

Обучение ассоциативным правилам обычно не учитывает порядок элементов внутри транзакции или по транзакциям.

Пусть имеется база данных, состоящая из покупательских транзакций. Каждая транзакция – это набор товаров, купленных покупателем за один визит. Такую транзакцию еще называют рыночной корзиной.

Задача поиска ассоциативных правил ставится следующим образом:

- Пусть дан набор $I = \{i_1, i_2, \dots, i_n\}$ из n двоичных атрибутов, называемых объектами, то есть это множество (набор) товаров, называемых также элементами.
- Пусть дан набор $D = \{t_1, t_2, \dots, t_m\}$ транзакций, называемый базой данных или коллекцией.

Каждая транзакция в D имеет уникальный ID (номер) транзакции и состоит из подмножества объектов из I .

- Бинарная коллекция

$$I = \{i_1, i_2, \dots, i_n\} = \{\text{хлеб, вода, молоко, масло, ...}\}$$

- $t_1 = [1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 0, \dots]$ первая транзакция (*itemset*)
- $t_2 = [1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, \dots]$ вторая транзакция (*itemset*)
- $t_3 = [0, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, \dots]$...
- ...
- $t_m = [1, 0, 0, 1, 0, 0, 0, 1, 1, 0, 1, \dots]$

Любое правило состоит из двух различных наборов объектов, известных также как **наборы объектов** X и Y , где X называется *первым операндом* или *левой частью*, а Y — *вторым операндом* или *правой частью*.

Правило определяется как импликация вида:

$$X \Rightarrow Y, \text{ где } X, Y \subseteq I.$$

Примером правила для супермаркета может служить

$$\{\text{масло, хлеб}\} \Rightarrow \{\text{молоко}\},$$

что означает, что, если куплены масло и хлеб, покупатель также купит и молоко.

Каждая транзакция представляет собой бинарный вектор, где $t[k]=1$, если i_k элемент присутствует в транзакции, иначе $t[k]=0$.

Мы говорим, что транзакция T содержит некоторый набор элементов X из I , если $X \subseteq T$.

Ассоциативным правилом называется импликация $X \Rightarrow Y$, где $X \subseteq I$, $Y \subseteq I$ и $X \cap Y = \emptyset$.

Правило $X \Rightarrow Y$ имеет **поддержку** s (*support*), если $s\%$ транзакций из D , содержат $X \Rightarrow Y$, то есть удовлетворяют этому правилу, $\text{supp}(X \Rightarrow Y) = \text{supp}(X \cup Y)$.

Достоверность правила показывает какова вероятность того, что из X следует Y .

Правило $X \Rightarrow Y$ справедливо с достоверностью (*confidence*) c , если $c\%$ транзакций из D , содержащих X , также содержат Y , $\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$.

Покажем на конкретном примере:

- 75% транзакций, содержащих хлеб, также содержат молоко. 3% от общего числа всех транзакций содержат оба товара.
- 75% – это **достоверность** (*confidence*) правила,
- 3% – это **поддержка** (*support*), или 'Хлеб' 'Молоко' с вероятностью 75%.

Другими словами, целью анализа является установление следующих зависимостей: если в транзакции встретился некоторый набор элементов X , то на основании этого можно сделать вывод о том, что другой набор элементов Y также должен появиться в этой транзакции. Установление таких зависимостей дает нам возможность находить очень простые и интуитивно понятные правила.

Алгоритмы поиска ассоциативных правил предназначены для нахождения всех правил $X \Rightarrow Y$, причем поддержка и достоверность этих правил должны быть выше некоторых наперед определенных порогов, называемых соответственно **минимальной поддержкой** (*minsupport*) и **минимальной достоверностью** (*minconfidence*).

Задача нахождения ассоциативных правил разбивается на две подзадачи:

- Нахождение всех наборов элементов, которые удовлетворяют порогу *minsupport*. Такие наборы элементов называются часто встречающимися.
- Генерация правил из наборов элементов, найденных согласно п.1. с достоверностью, удовлетворяющей порогу *minconfidence*.

Значения для параметров *минимальная поддержка* и *минимальная достоверность* выбираются таким образом, чтобы ограничить количество найденных правил.

Если поддержка имеет большое значение, то алгоритмы будут находить правила, хорошо известные аналитикам или настолько очевидные, что нет никакого смысла проводить такой анализ. С другой стороны, низкое значение поддержки ведет к генерации огромного количества правил, что, конечно, требует существенных вычислительных ресурсов. **Тем не менее, большинство интересных правил находится именно при низком значении порога поддержки.** Хотя слишком низкое значение поддержки ведет к генерации статистически необоснованных правил.

Поиск ассоциативных правил совсем не тривиальная задача, как может показаться на первый взгляд. Одна из проблем – алгоритмическая сложность при нахождении часто встречающихся наборов элементов, т.к. с ростом числа элементов в I ($|I|$) экспоненциально растет число потенциальных наборов элементов.

Пример из области супермаркета

Чтобы проиллюстрировать концепцию, используем маленький пример из области супермаркета. Множество объектов I — это молоко, хлеб, масло, пиво, памперсы, и в таблице ниже показана маленькая база данных, содержащая объекты, в которой значение 1 означает наличие объекта в соответствующей транзакции, а значение 0 означает отсутствие объекта в транзакции.

Пример базы данных с 5 транзакциями и 5 элементами

ID транзакции	молоко	хлеб	масло	пиво	памперсы
1	1	1	0	0	0
2	0	0	1	0	0
3	0	0	0	1	1
4	1	1	1	0	0
5	0	1	0	0	0

Таким образом, датасет представляет собой разреженную матрицу со значениями $\{1,0\}$. Это будет **бинарный датасет**. Существуют и другие виды записи – **вертикальный датасет** (показывает для каждого отдельного *item* вектор транзакций, где он присутствует) и **транзакционный датасет** (примерно как в кассовом чеке).

Данные преобразовали, как найти правила?

Существует целый ряд ключевых понятий в ARL, которые нам помогут эти правила вывести.

Полезные концепции

Чтобы выбрать вызывающее интерес правило из множества всех возможных правил, используются ограничения на различные меры значимости и содержательности. Наиболее известными ограничениями являются **минимальный порог поддержки и доверия**.

Пусть X будет набором объектов, $X \Rightarrow Y$ будет ассоциативным правилом, а T — набором транзакций данной базы данных.

Поддержка (Support)

Поддержка — это показатель, насколько часто набор объектов обнаруживается в базе данных.

Поддержка набора X по отношению к T определяется как отношение числа транзакций t в базе данных, содержащих набор X , к общему числу транзакций.

$$\text{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

где X — itemset, содержащий в себе i -items, а $|T|$ — количество транзакций.

В нашем примере набор данных $X=\{\text{пиво, памперсы}\}$ имеет поддержку $1/5 = 0,2$, поскольку он обнаруживается в 20 % всех транзакций (1 из 5 транзакций). Аргумент функции $\text{supp}()$ является множеством предусловий и потому становится более ограничивающим по мере расширения (в отличие от более охватывающего).

Доверие, достоверность (Confidence)

Доверие — это показатель, насколько часто правило оказывается верным. Значение доверия правила $X \Rightarrow Y$ по отношению к набору транзакций T является отношением числа транзакций, которые содержат как набор X , так и набор Y , к числу транзакций, содержащих набор X .

Доверие определяется как:

$$\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$$

Например, правило $\{\text{масло, хлеб}\} \Rightarrow \{\text{молоко}\}$ имеет доверие $0,2/0,2=1,0$ в базе данных, что означает, что для 100% транзакций, содержащих масло и хлеб, правило верно (в 100 % случаев, когда покупается масло и хлеб, молоко покупается также).

Заметим, что $\text{supp}(X \cup Y)$ означает поддержку объектов в X и Y . Это несколько запутывает, поскольку мы обычно думаем в терминах вероятности событий, а не терминах набора объектов. Мы можем переписать $\text{supp}(X \cup Y)$ как вероятность $P(E_X \cap$

ЕУ), где ЕХ и ЕУ являются событиями, что транзакция содержит наборы X и Y соответственно.

Доверие можно понимать как оценку условной вероятности $P(EY|EX)$, вероятности нахождения правой части правила в транзакциях при условии, что транзакции содержат левую часть правила.

Лифт (lift)

Лифт правила определяется как

$$\text{lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}$$

или отношение наблюдаемой поддержки к математическому ожиданию события, если бы X и Y были бы независимы. Например, правило {молоко, хлеб} => {масло} имеет лифт

$$\frac{0,2}{0,4 \times 0,4} = 1,25.$$

Если правило имеет лифт 1, это означает, что событие в левой части независимо от события в правой части. Если два события независимы, никакого правила нельзя вытащить из этих двух событий.

Если лифт > 1, это позволяет нам знать степень, насколько события связаны друг с другом, и делает эти правила потенциально полезными для предсказания следствия в будущих наборах данных.

Если лифт < 1, это означает, что объекты заменяют друг друга. Это означает, что наличие одного объекта имеет отрицательный эффект на присутствие второго объекта, и наоборот.

Значение лифта принимает во внимание как доверие правила, так и общие данные.

Уверенность, убедительность (Conviction)

Уверенность правила определяется как

$$\text{conv}(X \Rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \Rightarrow Y)}.$$

Например, правило {молоко, хлеб} => {масло} имеет уверенность

$$\frac{1 - 0,4}{1 - 0,5} = 1,2$$

и может пониматься как отношение ожидаемой частоты, что X встречается без Y (говоря иначе, частота, что правило даёт неправильное предсказание), если бы X и Y были бы независимыми, и наблюдаемой частоты неверных предсказаний.

Процесс

От ассоциативных правил обычно требуется выполнение определённой пользователем минимальной поддержки и определённого пользователем минимального доверия. Генерация ассоциативного правила обычно разделяется на два шага:

1. Минимальный порог поддержки используется для поиска всех частых наборов объектов в базе данных.
2. Ограничение минимального доверия применяется к этим наборам для формирования правила.

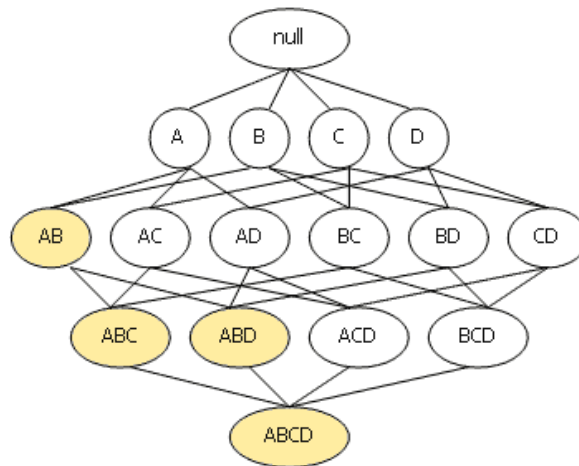
Второй шаг прост и ясен, а первый шаг требует большего внимания.

Свойство анти-монотонности

Выявление часто встречающихся наборов элементов – операция, требующая много вычислительных ресурсов и, соответственно, времени. Примитивный подход к решению данной задачи – простой перебор всех возможных наборов элементов. Это потребует $O(2^{|I|})$ операций, где $|I|$ – количество элементов. Поэтому используют одно из свойств поддержки, гласящее: поддержка любого набора элементов не может превышать минимальной поддержки любого из его подмножеств. Например, поддержка 3-элементного набора {Хлеб, Масло, Молоко} будет всегда меньше или равна поддержке 2-элементных наборов {Хлеб, Масло}, {Хлеб, Молоко}, {Масло, Молоко}.

Дело в том, что любая транзакция, содержащая {Хлеб, Масло, Молоко}, также должна содержать {Хлеб, Масло}, {Хлеб, Молоко}, {Масло, Молоко}, причем обратное не верно. Это свойство носит название **анти-монотонности** и служит для снижения размерности пространства поиска.

Свойству анти-монотонности можно дать и другую формулировку: с ростом размера набора элементов поддержка уменьшается, либо остается такой же. Из всего вышесказанного следует, что любой k-элементный набор будет часто встречающимся тогда и только тогда, когда все его (k-1)-элементные подмножества будут часто встречающимися. Все возможные наборы элементов из I можно представить в виде решетки, начинающейся с пустого множества, затем на 1 уровне 1-элементные наборы, на 2-м – 2-элементные и т.д. На k уровне представлены k-элементные наборы, связанные со всеми своими (k-1)-элементными подмножествами.



Рассмотрим рисунок иллюстрирующий набор элементов $I = \{A, B, C, D\}$. Предположим, что набор из элементов $\{A, B\}$ имеет поддержку ниже заданного порога и, соответственно, не является часто встречающимся. Тогда, согласно свойству антимонотонности, все его супермножества также не являются часто встречающимися и отбрасываются. Вся эта ветвь, начиная с $\{A, B\}$, отмечена желтым фоном. Использование этой эвристики позволяет существенно сократить пространство поиска.

Алгоритмы

В процессе поиска ассоциативных правил может производиться обнаружение всех ассоциаций, поддержка и достоверность для которых превышают заданный минимум.

Простейший алгоритм поиска ассоциативных правил рассматривает все возможные комбинации условий и следствий, оценивает для них поддержку и достоверность, а затем исключает все ассоциации, которые не удовлетворяют заданным ограничениям. Число возможных ассоциаций с увеличением числа предметов растет экспоненциально. Поэтому в процессе генерации ассоциативных правил широко используются методики, позволяющие уменьшить количество ассоциаций, которое требуется проанализировать.

Было предложено много алгоритмов для генерации ассоциативных правил.

Несколько алгоритмов хорошо известны, это Apriori, Eclat и FP-Growth, но они делают только половину работы, поскольку они предназначены для отыскания часто встречающихся наборов объектов. Нужно сделать ещё один шаг после того, как часто встречающиеся наборы найдены в базе данных.

Алгоритм Apriori

В основе алгоритма Apriori лежит понятие частого набора. Под частотой понимается простое количество транзакций, в которых содержится данный предметный набор.

Частый предметный набор – предметный набор с поддержкой больше заданного порога либо равной ему. Этот порог называется **минимальной поддержкой**.

Работа данного алгоритма состоит из нескольких этапов, каждый из этапов состоит из следующих шагов:

1. *формирование кандидатов;*
2. *подсчет кандидатов.*

Формирование кандидатов (candidate generation) - этап, на котором алгоритм, сканируя базу данных, создает множество i -элементных кандидатов (i - номер этапа). На этом этапе поддержка кандидатов не рассчитывается.

Подсчет кандидатов (candidate counting) - этап, на котором вычисляется поддержка каждого i -элементного кандидата. Здесь же осуществляется отсеечение кандидатов, поддержка которых меньше минимума, установленного пользователем (min_sup).

Оставшиеся i -элементные наборы называем **часто встречающимися**.

Чтобы сократить пространство поиска ассоциативных правил, алгоритм Apriori использует свойство антимонотонности. Свойство утверждает, что если предметный набор Z не является частым, то добавление некоторого нового предмета A к набору Z не делает его более частым. Данное полезное свойство позволяет значительно уменьшить пространство поиска ассоциативных правил.

На первом этапе алгоритма Apriori формируются частые однопредметные наборы – множество F_1 .

Для поиска F_k , то есть k -предметных наборов, алгоритм Apriori сначала создает множество F_k кандидатов в k -предметные наборы путем связывания множества F_{k-1} с самим собой. Затем F_k сокращается с использованием свойства антимонотонности.

Предметные наборы множества F_k , которые остались после сокращения, формируют F_k .

После того, как все частые предметные наборы найдены, можно переходить к генерации на их основе ассоциативных правил. Для этого к каждому частому предметному набору s можно применить процедуру, состоящую из 2 шагов.

1. Генерируются все возможные поднаборы s
2. Если поднабор ss является непустым поднабором s , то рассматривается ассоциация $R:ss \rightarrow (s-ss)$, где $s-ss$ представляет собой набор s без поднабора ss . R считается ассоциативным правилом, если удовлетворяет условию заданного минимума

поддержки и достоверности. Данная процедура повторяется для каждого подмножества ss из s .

Рассмотрим работу алгоритма Apriori на примере базы данных D . Минимальный уровень поддержки равен 3.

На первом этапе происходит формирование одноэлементных кандидатов. Далее алгоритм подсчитывает поддержку одноэлементных наборов. Наборы с уровнем поддержки меньше установленного, то есть 3, отсекаются. В нашем примере это наборы e и f , которые имеют поддержку, равную 1. Оставшиеся наборы товаров считаются часто встречающимися одноэлементными наборами товаров: это наборы a , b , c , d .

Далее происходит формирование двухэлементных кандидатов, подсчет их поддержки и отсекаются наборы с уровнем поддержки, меньшим 3. Оставшиеся двухэлементные наборы товаров, считающиеся часто встречающимися двухэлементными наборами ab , ac , bd , принимают участие в дальнейшей работе алгоритма.

Если смотреть на работу алгоритма прямолинейно, на последнем этапе алгоритм формирует трехэлементные наборы товаров: abc , abd , bcd , acd , подсчитывает их поддержку и отсекает наборы с уровнем поддержки, меньшим 3. Набор товаров abc может быть назван часто встречающимся.

Однако алгоритм Apriori уменьшает количество кандидатов, отсекая априори тех, которые заведомо не могут стать часто встречающимися, на основе информации об отсеченных кандидатах на предыдущих этапах работы алгоритма.

Отсечение кандидатов происходит на основе предположения о том, что у часто встречающегося набора товаров все подмножества должны быть часто встречающимися.

Если в наборе находится подмножество, которое на предыдущем этапе было определено как нечасто встречающееся, этот кандидат уже не включается в формирование и подсчет кандидатов.

Так наборы товаров ad , bc , cd были отброшены как нечасто встречающиеся, алгоритм не рассматривал товаров abd , bcd , acd .

При рассмотрении этих наборов формирование трехэлементных кандидатов происходило бы по схеме, приведенной в верхнем пунктирном прямоугольнике. Поскольку алгоритм априори отбросил заведомо нечасто встречающиеся наборы, последний этап алгоритма сразу определил набор abc как единственный трехэлементный часто встречающийся набор (этап приведен в нижнем пунктирном прямоугольнике).

Алгоритм Apriori рассчитывает также поддержку наборов, которые не могут быть отсечены априори. Это так называемая негативная область (*negative border*), к ней принадлежат наборы-кандидаты, которые встречаются редко, их самих нельзя

отнести к часто встречающимся, но все подмножества данных наборов являются часто встречающимися.

Статистически обоснованные ассоциации

Одним из ограничений стандартного подхода к обнаружению ассоциаций является то, что при поиске в большом числе возможных ассоциаций набора объектов, которые могут быть ассоциированными, есть большой риск нахождения большого числа случайных ассоциаций. Это наборы объектов, которые оказываются вместе с неожиданной частотой в данных, но чисто случайно. Например, предположим, что мы рассматриваем набор из 10.000 объектов и ищем правило, содержащее два объекта в левой части и один объект в правой части. Имеется примерно 1.000.000.000.000 таких правил. Если мы применим статистический тест независимости с уровнем 0,05 это означает, что имеется только 5 % шанса принять правило при отсутствии ассоциации.

Если мы предполагаем, что нет никаких ассоциаций, мы должны, тем не менее, ожидать нахождения 50.000.000.000 правил. Статистически обоснованное обнаружение ассоциаций контролирует этот риск, в большинстве случаев сокращая риск нахождения любой случайной ассоциации для заданного пользователем уровня значимости.

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Тематика лабораторных работ

Номер	Наименование лабораторной работы
1	Ввод и предварительный анализ данных
2	Регрессионный анализ
3	Факторный анализ
4	Кластерный анализ

Лабораторная работа 1. Ввод и предварительный анализ данных

Ввод данных

1. В специальной папке STATISTICA на рабочем столе выберите модуль *Работа с данными* и запустите его на выполнение.
2. Используя описание пакета, познакомьтесь с основными элементами интерфейса.
3. Используя команду **File|New Data...**, создайте таблицу следующей структуры:

Таблица 1

Столбец	Переменная	Обозначение	Ед. изм	Шкала	Формат	Комментарий
1	Группа	Группа		Номин., категориз.	8.0	
2	Пол	Пол		Номин., категориз.	4.0	
3	Контракт /Бюджет	К-Б		Номин., категориз.	4.0	
4	Дата рождения	Дата		Порядковая (ординальная)	DATE6, 9.0	Пример: 17/07/81
5	Вес	Вес	Кг	Количественная	6.0	
6	Рост	Рост	См	Количественная	6.0	
7	Объем в талии	Талия	См	Количественная	6.0	
8-11	Оценки	Матем. Прогр. Филос.		Порядковая (ординальная)	6.0	

Примечание: Объектам (случаям) дать имена – фамилии студентов.

4. Введите исходные данные в созданную таблицу (см. табл.1).
5. В каталоге STATISTICA сохраните ее под именем AABV.STA, где AA – номер группы, BV – год поступления, например 2198.sta.

Первичная обработка данных

1. *Обработка пропусков.* Способ обработки пропущенных данных может корректироваться индивидуально для каждого вида анализа. Обычно он может быть установлен из стартовой панели конкретного статистического модуля. Пользователь имеет возможность
 - устранить данные из вычислений;
 - заменить их средним значением;
 - интерполировать данные.
 - Если в вашей таблице нет пропущенных данных, создайте их искусственно.
 - В модуле по работе с данными **Data management** – *Управление данными* при помощи команды **Replace Missing Data by Means** – *Замена пропущенных значений на среднее* произведите обработку пропусков для переменной ВЕС.
 - При необходимости восстановите исходное состояние таблицы.
2. *Проверка значений данных.* Команда **Verify Data Values** – *Проверка значений данных* позволяет проверить введенные данные. Проверяется выполнение некоторых, предварительно заданных пользователем условий. Имеется возможность задания до 4 условий. При это можно установить между ними операции логического НЕ, И, ИЛИ. Заданные условия можно сохранить в специальном файле.
 - Установите условия проверки, например:
 - Условие 1 – $v5 \geq 140$ and $v5 \leq 200$
 - Условие 2 – $v6 \geq 40$ and $v6 \leq 100$
 - Условие 3 – $v8 \leq 5$ and $v9 \leq 5$ and $v10 \leq 5$ and $v11 \leq 5$
 - В секции **Range** задайте номера проверяемых случаев, например от 1 до 25.
 - Проверьте введенные данные, исправьте обнаруженные ошибки.
 - Сохраните условия в специальном файле. При следующем вызове этот файл можно загрузить и использовать для проверки.
3. *Функциональные преобразования* количественных переменных используются для преобразования данных к форме, более соответствующих природе задачи, ее «физическому смыслу». Например, часто встречаются несимметричные

распределения с длинным “хвостом”, что затрудняет статистический анализ. Распределение приобретает симметричный вид, если перейти к логарифму переменной.

Для выполнения функциональных преобразований создадим дополнительную таблицу и выполним в ней заданные действия:

- Используйте команду **Create subset from data file – Создать подмножество данных из файла** для создания новой таблицы из существующей. Включите в эту новую таблицу с 5 по 7 переменные (5-7) и все (**All**) случаи. Необходимые переменные для новой таблицы можно выбрать и предварительно, то есть до выбора команды **Create subset from data file**.
 - Перейдите к новой переменной $\ln(\text{РОСТ})$. Формулы преобразования переменных задаются в диалоговом окне спецификаций переменных.
 - Вернитесь к прежней переменной РОСТ, используя функцию $\text{Exp} (= \text{Exp}(v1))$
4. *Нормировка количественных признаков* используется для приведения их к стандартному виду, удобному для обработки. Обычно при расчете расстояния между объектами нормируют признаки, измеренные в разнородных физических единицах. В STATISTICA значения переменных изменяются на стандартизованные значения по следующей формуле:

Новое значение = (Старое значение – Среднее) / Стандартное отклонение

- Выполните стандартизацию переменных РОСТ и ВЕС по команде **Standardize Variables**.
 - Вернитесь к первоначальной таблице.
5. *Сортировка данных*. Команда **Sort Cases – Сортировка случаев** выполняет иерархическую сортировку данных. На каждом уровне сортировки (их число ограничено 7 уровнями) должен быть задан ключ. В качестве ключа могут быть использованы либо переменные, либо случаи. Для различных ключей допускаются различные направления сортировки: по убыванию и возрастанию. Сортировка может проводиться как по текстовым, так и по числовым значениям переменной.
- Отсортируйте данные, например, по полу и внутри пола по убыванию роста.
 - Вернитесь к прежней сортировке (по случаям, текстовое значение).

Работа с данными в системе STATISTICA

1. Добавьте еще одну переменную в конец таблицы. Имя новой переменной – СРБАЛЛ (средний балл по всем предметам), формат б.2. В диалоговом окне

Спецификации переменной секции **Long name** задайте формулу расчета значений новой переменной и произведите расчет.

2. Переместите новую переменную на другую позицию – после переменной ВЕС.
3. Полученную таблицу добавьте в конец общей таблицы, содержащей обобщенную информацию по всем группам. Таблица ROSTVES.STA хранится в каталоге STATISTICA. Для выполнения указанной операции используйте команду **Merge two Data Files** – *Объединить два файла данных*.
4. Отработайте основные операции над случаями и переменными, приведенные в описании пакета.
5. Изучите операции с выделенным блоком значений.

Предварительный анализ данных. Визуализация данных

1. *Настройка графических компонент*. Перед началом работы с графиками желательно произвести предварительную настройку. Инструменты настройки доступны только при открытом графике, поэтому:
 - Откройте произвольный график, например, двумерную диаграмму рассеяния ВЕС–РОСТ. Для этого выделите произвольную ячейку в столбце РОСТ, на ПИ нажмите клавишу **Quick Stats Graphs** (на кнопке буква S) и затем последовательно – **Scatterplot by...| regular | ВЕС**.
 - Откройте панель настройки **Options|Global Defaults...**
 - Выберите шрифты для всех заголовков в секции TITLES.
 - Выберите шрифты для символов осей в секции AXES.
 - Сохраните установки в специальном файле, чтобы в следующий раз не устанавливать все заново.

Все компоненты графика могут быть настроены при помощи двух основных диалоговых окон – **General Layout** – *Общая разметка* и **Plot Layout** – *Размещение графика*. Из диалогового окна **General Layout** можно перейти в **Plot Layout** и наоборот.

Рекомендация. Иногда проще всего задать параметры настройки, дважды щелкнув на том или ином элементе графика, либо один раз нажать на правую кнопку мыши и выбрать необходимую команду из контекстного меню.

2. *Проекция данных на один признак – гистограмма*. Работа с гистограммой предполагает
 - оценку наличия группировки объектов. Если на гистограмме просматриваются 2-3 горба, то это явный признак наличия групп;

- оценку параметров теоретического распределения. Если гистограмма примерно симметрична, ее “хвосты” не слишком длинные (например, не больше пяти объектов на сто лежат вне интервала $\pm 2s$), то оценками параметров являются среднее значение выборки \bar{x} и выборочное стандартное отклонение s . Значения \bar{x} и s используются в качестве характеристик положения и разброса всех выборок независимо от того, из какого они сделаны распределения. При резко асимметричной гистограмме более удобной характеристикой центра является медиана, которая более устойчива к резким выбросам в данных, чем среднее;
- анализ ‘засоренности’, выявление ошибок в данных. Подозрительные объекты обычно находятся на краях диаграммы.

Выполните следующие действия:

- Постройте гистограммы переменных РОСТ и ВЕС. Найдите основные статистики (БСГ|Values | Stats of РОСТ). Проверьте на наличие групп.
- Постройте гистограммы переменных РОСТ и ВЕС отдельно для студентов и студенток.
- Проведите анализ полученных диаграмм.

3. Проекция на плоскость двух признаков – двумерная диаграмма рассеяния.

- Постройте двумерные гистограммы вида «рост-вес», «вес-объем талии» (см. рис.1).
- Проверьте диаграммы на ошибки и выбросы.
- Исследуйте влияние размерности на вид диаграмм.

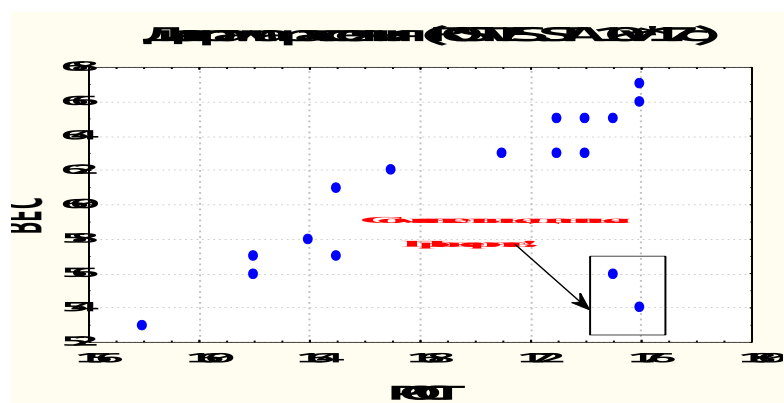


Рис. 1

4. Проекция на плоскость трех признаков – объемная диаграмма рассеяния

- Постройте объемную гистограмму вида «рост-вес, объем талии».


- Изучите возможности вращения и прокрутки объектов. Проверьте работу в режиме анимационного расслоения.

Вероятностный калькулятор

Вероятностный калькулятор – это средство, позволяющее максимально быстро построить график наиболее употребляемых функций и их плотностей, вычислить процентные точки.

- Запустите модуль **Basic Statistics/Tables** из **Переключателя модулей**. Откройте окно **Probability calculator**.
- Изучите стандартные распределения (нормальное, Стьюдента, хи-квадрат, F-распределение, логарифмически-нормальное, биномиальное), задавая различные значения параметров распределения. Как меняются графики распределений при изменении среднего, стандартного отклонения, степени свободы?
- Научитесь рассчитывать вероятность по заданному значению и наоборот. Предполагая, что распределение роста в группах студентов подчиняется нормальному закону, рассчитайте вероятность, что случайно выбранный студент имеет рост выше 200 см.
- Исследуйте на графике нормального распределения правило 2 и 3 сигма. Задавая различные параметры нормального распределения и указывая в строке Z вероятностного калькулятора 2 и 3 стандартных отклонения, убедитесь, правила 2 и 3 сигм действительно справедливы.

Генерация случайных чисел в Statistica

 Исследуйте зависимость стандартной ошибки выборочного среднего и выборочной дисперсии от объема выборки, для чего

1. Создайте выборку объемом N из генеральной совокупности с нормальным распределением $N(\mu, \sigma)$ и произвольными параметрами.
2. По выборке рассчитайте выборочное среднее \bar{x} и выборочное стандартное отклонение s.
3. Повторите действия 1 и 2 несколько k раз.
4. Постройте гистограмму для выборочного среднего \bar{x} и стандартного отклонения s.
5. Рассчитайте стандартную ошибку выборочного среднего.
6. Сравните стандартную ошибку среднего со стандартным отклонением и сделайте выводы.

📖 Исследуйте влияние объема выборки на достоверность результатов статистического анализа, для чего

1. Сформируйте выборку объемом N из генеральной выборки, имеющей стандартизованное нормальное распределение $N(\mu=0, \sigma=1)$.
2. По каждому признаку найдите среднее значение координат выбранных объектов

$$m_j = (x_{j1}^{(i)} + x_{j2}^{(i)} + \dots + x_{jN}^{(i)}) / N, \quad j=1, p.$$

3. В качестве критерия оценки координат центра совокупности возьмем расстояние от полученной по выборке точки с координатами (m_1, m_2, \dots, m_p) до истинного центра (M_1, M_2, \dots, M_p) , т.е. до начала координат

$$\begin{aligned} L &= \sqrt{(M_1 - m_1)^2 + (M_2 - m_2)^2 + \dots + (M_p - m_p)^2} \\ &= \sqrt{m_1^2 + m_2^2 + \dots + m_p^2} \end{aligned}$$

Величина L характеризует ошибку, которую мы допустим, если припишем всей совокупности координаты центра сделанной небольшой выборки.

Повторите указанные действия для различных значений N и P :

$$N = \{10, 30, 100, 300, 1000\} \quad P = \{1, 2, 5, 10, 30, 150\}$$

На основании полученных данных постройте следующие зависимости:

- ◆ Значение ошибки L при оценивании положения центра генеральной совокупности, имеющей стандартизованное p -мерное нормальное распределение в зависимости от объема выборки N для $p=10$ и $p=150$.
- ◆ Зависимость ошибки L от числа признаков для $N=10$ и $N=100$.

Лабораторная работа 2. Регрессионный анализ

Теоретическое введение

Регрессионные методы позволяют выявить связи между переменными, причем особенно эффективно, если эти связи не совершенны или не имеют точного функционального описания между этими переменными. В анализе используются независимые переменные x и одна зависимая переменная y .

Регрессией в прямом смысле называется функция вида

$$y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i < j} b_{ij} x_i x_j + \sum_{i=1}^k b_{ii} x_i^2 + \dots + \varepsilon \quad (1)$$

где $b_0, b_{ij}, b_{ii}, \dots$ - известные коэффициенты регрессии;

ε - невязка (ошибка, отклонение), обусловленная недостаточной пригодностью модели и ошибкой эксперимента. Эти причины обычно являются смешанными.

Регрессия (1) называется множественной.

В том случае, если исследуется влияние одной переменной на результат эксперимента, то выражение (1) упрощается к виду

$$y = b_0 + b_1 x_1 + \varepsilon \quad (2)$$

Выражение (2) представляет собой линейную регрессию.

Кроме того, часто используется нелинейное представление регрессии (логарифмическое, экспоненциальное, гиперболическое и другое задаваемое исследователем). В этом случае переменная x подвергается соответствующему нелинейному преобразованию.

Для нахождения коэффициентов уравнений (1) и (2) используется метод наименьших квадратов. Сущность метода заключается в том, чтобы минимизировать сумму квадратов отклонений

$$SS = \sum_{i=1}^n [y_i - y_i^0]^2 \rightarrow \min \quad (3)$$

где y_i^0 - значение результата, вычисленное по уравнению (1) в точке x_i ;

y_i - экспериментальное значение результата в этой же точке.

Существует и другой смысл регрессии. В этом смысле под термином регрессии понимается метод обработки статистических данных, позволяющий подобрать некоторую сглаживающую функцию по упорядоченным данным. При этом полученное функциональное описание сглаженной кривой необязательно имеет физический смысл.

Регрессия может использоваться в разных целях:

- для получения функциональной кривой описывающей исследуемый процесс;
- для свертывания информации заданной в табличном виде в вид функциональной зависимости;
- для прогнозирования результата исследования в промежутках между дискретным описанием исследуемого процесса (задача интерполяции);

- для предсказания результата за границами исследования (задача прогнозирования).

Корректность выводов при регрессионном анализе обеспечивается при выполнении следующих условий:

1. Результаты наблюдений y_1, y_2, \dots, y_n являются независимыми, нормально распределенными случайными величинами.
2. Ошибки измерения независимых переменных x_1, x_2, \dots, x_n пренебрежимо малы по сравнению с ошибкой определения y .
3. Переменные x_1, x_2, \dots, x_n - линейно - независимые величины.

Выполнение первого условия осуществляется с использованием критериев Колмогорова - Смирнова и Шапиро-Вилкоксона в тех случаях, когда число параллельных наблюдений по каждому случаю (Case) не менее трех.

Второе условие требует, чтобы ошибки измерения факторов x_i были пренебрежимо малы по сравнению с ошибкой определения y_i .

Третье условие, налагающее ограничения на взаимную связь между значениями факторов, проявляется в процессе определения коэффициентов регрессии b . Может оказаться, что число уравнений (Case), из которых находятся b , будет меньше количества определяемых коэффициентов (количество переменных + 1).

Выполнение регрессионного анализа с использованием пакета STATISTICA включает в себя несколько этапов:

Предварительный анализ. Проводится с целью определения выполнения рассмотренных выше условий и предварительного анализа вида предполагаемой регрессии. В простых случаях этого исследования может быть достаточно для формирования выводов. Этот вид анализа выполняется при использовании блока «Описательные статистики» и при построении графиков **LinePlots (XY Trace)**.

Проведение регрессионного анализа. Проводится с использованием блока Line Regression. В результате анализа определяются коэффициенты уравнения регрессии при заданных условиях, уровень их значимости (p) и значения t - критерия для них, коэффициент корреляции R , значения F - критерия и уровень его значимости (p).

Анализ полученных результатов и формирование выводов. В том случае, если полученное уравнение регрессии адекватно описывает исследуемый процесс, что подтверждается высоким уровнем значимости F - критерия, делается вывод о соответствии уравнения регрессии исследуемому процессу. Если значение F - критерия не соответствует требуемому уровню значимости, гипотезу об адекватности уравнения регрессии исследуемому процессу отвергают. Анализ повторяют снова, используя различные методы преобразования переменных. Кроме того, для получения более обоснованных статистических выводов проводят анализ остатков. Это

позволяет выявить случаи (Case), которые описываются полученным уравнением регрессии, провести их анализ и сделать обоснованные выводы.

А. Примеры использования модуля регрессионного анализа в Statistica

Ниже рассмотрены два примера одномерной и многомерной линейной регрессии, иллюстрирующие работу модуля «Множественная регрессия».

Пример 1. Линейная регрессия

Формулировка задачи

В качестве примера рассмотрим реальные данные. Пусть даны оптовые цены на марочные вина в зависимости от года закладки вина (табл.3). Цены указаны в долларах за одну бутылку.

Таблица 3

ГОД	ЦЕНА	ГОД	ЦЕНА
1890	50,00	1941	10,00
1900	35,00	1944	5,99
1920	25,00	1948	8,98
1931	11,98	1950	6,98
1934	15,00	1952	4,99
1935	13,00	1955	5,98
1940	6,98	1960	4,98

Естественно предположить, что между возрастом вина и его стоимостью имеется некоторая зависимость. Интуитивно понятно, что чем больше выдержка, тем вино должно быть дороже. Данные показывают, что в общем эта тенденция выполняется, однако имеются и исключения. Итак, мы хотим приблизительно выразить зависимость между этими двумя переменными. Зачем это нужно? Например, имея такую формулу, можно прогнозировать стоимость вин на следующем аукционе (в 1973 году) или, если вы планируете продавать вино, года закладки которого нет в таблице (например, 1945), вы можете грамотно назначить на него цену.

Математическая постановка задачи

В нашем примере независимая переменная – год закладки (или выдержка, которая равна возрасту вина), а зависимая переменная – его цена на аукционе. Иногда используется и другая терминология. Зависимая переменная часто называется откликом, а независимые переменные – предикторами, контролируруемыми

переменными или факторами. Эта терминология подчеркивает, что ряд переменных оказывает влияние на одну переменную – отклик.

В результате исследований невозможно найти точные значения параметров β_1 , β_0 , однако можно получить их оценки, которые мы будем обозначать через b_1 , b_0 и которые являются приближениями к значениям неизвестных параметров. После того, как оценки получены, мы можем записать уравнение связи в виде

$$\hat{y} = b_1x + b_0 \quad (4)$$

Уравнение (4) позволяет нам найти оценку отклика при любом значении контролируемой переменной.

Запуск STATISTICA

Запустите Windows 95. Нажмите кнопку *Пуск* и в меню *Программы* выберите папку, которая содержит систему STATISTICA. В этой папке выберите ярлык программы STATISTICA и дважды щелкните на нем мышью.

Выбор статистического модуля

После запуска программы на экране появится **Переключатель модулей – Module Switcher**, при помощи которого можно выбрать необходимый для работы модуль. Выберите модуль **Множественная регрессия – Multiple Regression**. Для этого подведите указатель мыши к названию этого модуля и дважды щелкните левой кнопкой мыши.

После запуска модуля на экране откроется основное окно системы STATISTICA. При запуске системы в нее автоматически загружается последний файл, с которым вы работали в ней. Если вы запускаете STATISTICA в первый раз, то по умолчанию в ней открывается файл с исходными данными, который называется *Adstudy.sta*. Одновременно с этим появляется *Стартовая панель* модуля, содержащая основные операции, доступные в запущенном модуле, которая также позволяет определить различные параметры анализа. В *Стартовой панели* можно открыть необходимый файл данных для анализа, приписать веса переменным, отобрать (если это требуется) необходимое подмножество наблюдений и выбрать переменные для анализа (в нашем случае – зависимую и независимые переменные). Напомним, что в STATISTICA реализован принцип постоянной логической подсказки. Если вы не знаете, что нужно делать на следующем шаге обработки, то просто нажмите на клавишу ENTER. STATISTICA сама отправит вас к нужному диалоговому окну. Например, если вы не выбрали переменные для анализа, то откроется диалоговое окно выбора переменных, в котором вам будет предложено выбрать эти переменные; если вы не задали значения каких-либо параметров, то они будут заданы по умолчанию и т.д.

Создание электронной таблицы с исходными данными

Исходные данные в системе STATISTICA организованы в виде таблицы. Большинство реальных данных могут быть структурированы в табличную форму. Если читатель имеет опыт работы с электронными таблицами (например, MS Excel), то ему будет просто освоиться с электронными таблицами в STATISTICA. Электронная таблица в STATISTICA состоит из строк и столбцов. Столбцы таблицы называются **Variables** – *Переменные*, а строки **Cases** – *Наблюдения*. В качестве переменных выступают исследуемые величины, а случаи – это значения, которые принимают переменные и которые изменяются в процессе наблюдения. В нашем примере в качестве переменных могут естественно выступать *Год* закладки вина и его *Цена* на аукционе.

Ввод исходных данных и дополнительной информации

Переменные в электронной таблице могут принимать как текстовые, так и численные значения. Электронная таблица с данными из нашего примера приведена на рис. N. В первом столбце содержится переменная *Год* (год закладки), во втором – переменная *Цена* – цена бутылки (в долларах). Кроме значений переменных, таблица может содержать дополнительную информацию (название таблицы, комментарии об источнике данных и т.д.)

Перед непосредственным применением той или иной статистической процедуры часто возникает необходимость преобразования данных. Так, например, может потребоваться вычислить *Возраст* (выдержку) вина, которая определяется как разность года аукциона (1972) и года закладки вина.

Преобразование исходных данных

В электронных таблицах STATISTICA вы можете выполнить все необходимые преобразования. Например, перейдем от переменной *Год* и *Цена* к новым переменным *Возраст* и *Цена_Лог*, которые связаны с исходными данными при помощи формул $Возраст = 1972 - Год$, $Цена_Лог = \ln(Цена)$.

После этого таблица будет содержать четыре переменные и примет вид, изображенный на рис 2.1.

Формулы преобразований переменных задаются в диалоговом окне спецификаций переменной. Для его вызова достаточно дважды щелкнуть мышью на имени переменной в электронной таблице с исходными данными.

Визуализация данных

Теперь имеет смысл отобразить данные на графике. STATISTICA включает в себя большое количество разнообразных категорий и типов графиков. Это всевозможные графики на плоскости и в пространстве, включая научные графики в различных системах координат, деловые графики и диаграммы, специализированные статистические графики (включая гистограммы, матричные, категоризированные

графики, диаграммы рассеяния и др.), пиктографики. Графические средства системы *STATISTICA* доступны в любом модуле и на любом шаге статистического анализа.

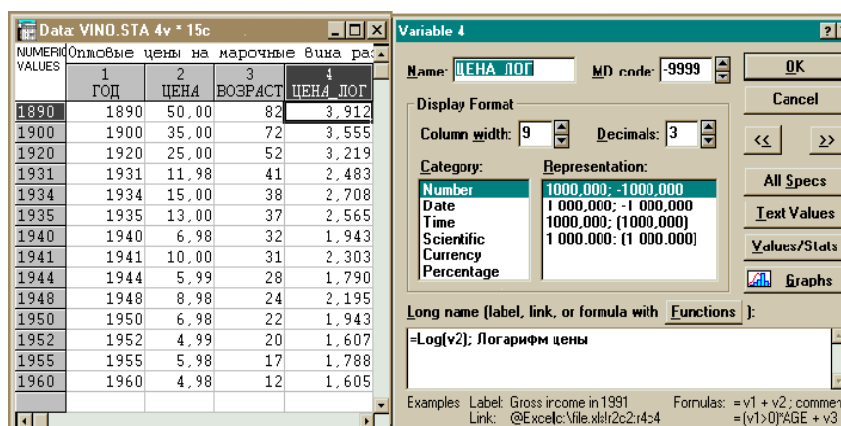


Рис 2.1 Преобразование переменных в таблице и диалоговое окно, в котором задаются все спецификации переменных

Для вызова графических возможностей системы можно воспользоваться меню **Graphics – Графика** и выбрать необходимый тип графика. В нашем примере мы воспользуемся двумерными диаграммами рассеяния. В диалоговом окне при помощи кнопки **Variables – Переменные** выберите необходимые переменные, которые вы хотите отобразить графически и необходимый тип графика. После нажатия на кнопку ОК график будет выведен в отдельном окне на рабочем пространстве системы.

После построения этих графиков на рабочем пространстве системы будет открыто три окна. В одном из них будет электронная таблица с исходными данными, а в двух других – графики. Пользователь имеет возможность сохранить необходимые ему графики в различных форматах или вывести их на принтер. Если нет необходимости сохранять графики, можно просто закрыть окно с ними.

Замечание

Преобразование логарифма стабилизирует дисперсию и часто применяется в статистике. Его можно интерпретировать и следующим образом: чем выше абсолютное значение переменной, тем выше и уровень случайных ошибок. При логарифмировании все ошибки становятся примерно одинаковыми. Поэтому мы будем искать линейную зависимость не между *Возрастом* и *Ценой*, а между *Возрастом* и логарифмированной ценой, получая при этом более устойчивые оценки параметров модели. Впоследствии, когда модель будет построена, можно перейти к исходным величинам. Итак, наша задача построить модель вида:

$$\text{ЦЕНА_ЛОГ} = b_1 * \text{ВОЗРАСТ} + b_0, \tag{5}$$

где b_1 – неизвестный коэффициент; b_0 – свободный член (также неизвестен).

При этом модель для цены будет иметь вид:

$$\text{ЦЕНА} = \exp(b_1 * \text{ВОЗРАСТ} + b_0), \quad (6)$$

Мы не только оценим неизвестные параметры, но исследуем значимость регрессии и адекватность построенной модели исходным данным.

Вызов стартовой панели модуля и определение анализа

Для начала статистического анализа вам необходимо вызвать *Стартовую панель* модуля. Это основное диалоговое окно модуля, в котором необходимо задать различные опции анализа. Если *Стартовая панель* модуля закрыта, то откройте ее. Для этого войдите в меню **Analysis – Анализ** и выберите команду **Startup Panel – Стартовая панель**.

Выбор переменных для анализа

Далее необходимо выбрать переменные для анализа. В нашем случае имеется одна зависимая переменная *Цена_Лог* и одна независимая переменная *Возраст*. Для их задания воспользуйтесь кнопкой **Variables – Переменные** из *Стартовой панели*.

В открывшемся окне **Select dependent and independent variable list – Выбор зависимой и списка независимых переменных** в качестве зависимой переменной выберите переменную с номером 4 – *Цена_Лог*, а в качестве независимой переменной переменную с номером 3 – *Возраст*. Нажмите кнопку **ОК** в правом верхнем углу. Вы вновь окажетесь в *Стартовой панели* модуля **Множественная регрессия**.

Задание дополнительных параметров анализа

Заметьте, что в *Стартовой панели* вы можете задать и дополнительные опции и параметры анализа. Например, вы можете выбрать определенное подмножество наблюдений для анализа, приписать вес переменным – эти опции относятся к исходным данным. Вы также можете задать и опции, которые относятся непосредственно к статистической процедуре: задать правило обработки пропущенных данных, выбрать метод анализа по умолчанию и др. Мы выбрали опцию - **Расчет с расширенной точностью** и выбор метода по умолчанию. Выбор таких опций не является необходимым.

Вывод результатов и их анализ

В стартовой панели нажмите на кнопку **ОК**. Система произведет вычисления и через секунду окно результатов появится на вашем экране.

Окно результатов анализа имеет следующую простую структуру: верхняя часть окна – информационная, нижняя содержит функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа.

Информационная часть

Рассмотрим вначале информационную часть окна. В ней содержится краткая информация о проведенном анализе, а именно:

- *Dep. Var* – Имя зависимой переменной. В нашем случае – *Цена_Лог*.
- *No. of Cases* – Число наблюдений, по которым построена регрессия. В примере число равно 14.
- *Multiple R* – Коэффициент множественной корреляции (Эта статистика полезна в множественной регрессии, когда вы хотите описать зависимости между переменными).
- *R-square (R2)* – *RI* – Квадрат коэффициента множественной корреляции, обычно называемый коэффициентом детерминации. Коэффициент детерминации является одной из основных статистик в данном окне, он показывает долю общего разброса (относительно выборочного среднего зависимой переменной), которая объясняется построенной регрессией.
- *Adjusted R-square: adjusted RI* – Скорректированный коэффициент детерминации, определяемый как:

$$\text{Adjusted R-square} = 1 - (1 - R^2) * (n / (n-p)),$$

где n – число наблюдений в модели, p – число параметров модели (число независимых переменных плюс 1 из-за свободного члена).

- *Std. Error of estimate* – Стандартные ошибки оценки. Эта статистика является мерой рассеяния наблюдаемых значений относительно регрессионной прямой.
- *Intercept* – Оценка свободного члена регрессии. Значение коэффициента b_0 в уравнении регрессии.
- *Std. Error* – Стандартная ошибка оценки свободного члена. Стандартная ошибка коэффициента b_0 в уравнении регрессии.
- *t(df) and p-value* – Значение t-критерия и уровень p . Т-критерий используется для проверки гипотезы о равенстве нулю свободного члена регрессии.
- *F* – значения F-критерия
- *df* – число степеней свободы F-критерия
- *p* – уровень значимости

В информационной части прежде всего необходимо смотреть на значение коэффициента детерминации. В нашем примере $RI = 0.929$. Это значит, что построенная регрессия объясняет 92,9% разброса значений *Цена_Лог* относительно среднего. Это хороший результат.

Далее вы смотрите на значение F-критерия и уровень его значимости p . F-критерий используется для проверки значимости регрессии. В данном случае для проверки

гипотезы, утверждающей, что между зависимой переменной *Цена_Лог* и независимой переменной *Возраст* нет линейной зависимости, т.е. $b_1 = 0$, против альтернативы b_1 не равен 0. В данном примере большое значение F-критерия =157.0486 и даваемый в окне уровень значимости $p = 0.0000$ показывают, что построенная регрессия высоко значима.

Функциональные кнопки

Нажмите далее на кнопку **Regression Summary – Краткие результаты регрессии**. Вы увидите электронную таблицу с результатами анализа.

В третьем столбце таблицы вы видите оценки неизвестных параметров модели:

$$b_0=1.142891;$$

$$b_1=0.034652.$$

Рассмотрим вторую часть информационного окна. В этой части система сама говорит нам о значимых регрессионных коэффициентах, высвечивая строку ЦЕНА_ЛОГ $\beta = 0,964$. В данном случае **beta** есть стандартизированный коэффициент b_1 , т.е. коэффициент при независимой переменной ВОЗРАСТ. Перейдем в функциональную часть окна результатов.

Прежде всего нажмем кнопку **Итоговый результат регрессии – Regression summary**. На экране появится электронная таблица вывода – **spreadsheet**, в которой представлены итоговые результаты оценивания регрессионной модели.

Это стандартная таблица вывода регрессионного анализа. В первом столбце таблицы даны значения коэффициентов **beta** – стандартизованные коэффициенты регрессионного уравнения, во втором – стандартные ошибки **beta**, в третьем – точечные оценки параметров модели:

Свободный член $b_0 = 1.143891$.

Коэффициент $b_1 = 0.034652$.

Далее, стандартные ошибки для b_0 , b_1 , значения статистик t-критерия и т.д.

Итак, искомая модель зависимости логарифма цены от возраста имеет вид:

$$\text{Цена_Лог} = 0.034652 * \text{Возраст} + 1.143891 \quad (7)$$

Требуемая регрессия построена. После перехода к исходным переменным модель примет вид:

$$\text{Цена} = \exp(0.034652 * \text{Возраст} + 1.143891) \quad (8)$$

График приведен на рис.2.2.

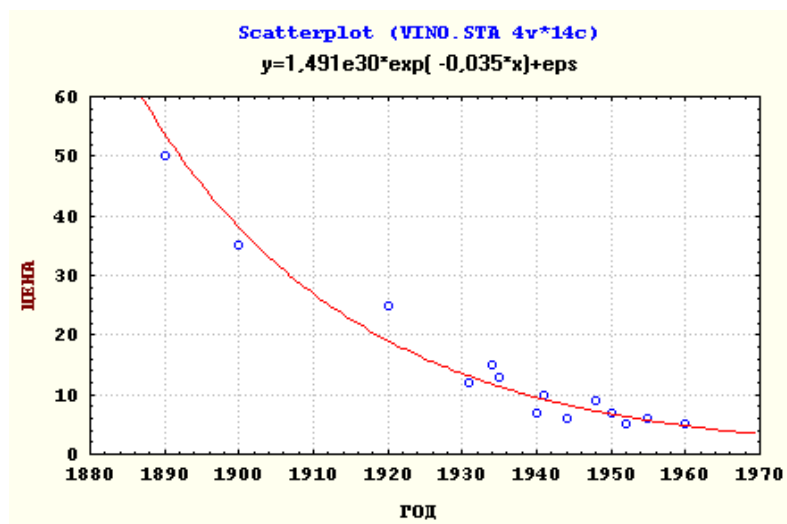


Рис 2.2. График зависимости цены от года закладки

Оценка адекватности модели. Исследование остатков.

После того как доказана адекватность модели, полученные результаты могут быть использованы по назначению. Анализ адекватности основывается на анализе остатков. Что такое остатки модели? Оценки представляют собой разности между наблюдаемыми значениями и модельными (предсказанными), т. е. значениями, подсчитанными по модели с оцененными параметрами (см. рис. 2.3.).

Data: VINO.STA 6v * 14c						
NUMERIC VALUES	Опловые цены на марочные вина различных годов вымуч					
	1 ГОД	2 ЦЕНА	3 ВОЗРАСТ	4 ЦЕНА ЛОГ	5 ПРЕДСКАЗ	6 ОСТАТКИ
1890	1890	50,00	82	3,91	3,99	-,0733
1900	1900	35,00	72	3,56	3,64	-,0835
1920	1920	25,00	52	3,22	2,95	,2731
1931	1931	11,98	41	2,48	2,56	-,0814
1934	1934	15,00	38	2,71	2,46	,2474
1935	1935	13,00	37	2,56	2,43	,1389
1940	1940	6,98	32	1,94	2,25	-,3097
1941	1941	10,00	31	2,30	2,22	,0845
1944	1944	5,99	28	1,79	2,11	-,3241
1948	1948	8,98	24	2,19	1,98	,2195
1950	1950	6,98	22	1,94	1,91	,0368
1952	1952	4,99	20	1,61	1,84	-,2295
1955	1955	5,98	17	1,79	1,73	,0554
1960	1960	4,98	12	1,61	1,56	,0457

Рис 2.3. Значения остатков

Пересчитаем значение *Цена_Лог* исходя из построенной модели для различных значений независимой переменной *Возраст*. Эти значения и называются *Predicted values* – *Предсказанные значения* или модельные, т.е. значения, предсказанные с

помощью модели. Очевидно, эти значения будут отличаться от значений *Цена_Лог*, имеющихся в исходном файле *vine.sta*. Разность между исходными (наблюдаемыми) значениями зависимой переменной и предсказанными значениями называются остатками.

В модуле **Множественная регрессия** системы *STATISTICA* остатки исследуются в специальном окне **Анализ остатков**. Исследуя остатки, вы можете оценить степень адекватности модели. Для этого нажмите в окне **Результаты множественной регрессии** кнопку **Residual analysis – Анализ остатков**. Нажав данную кнопку, вы раскроете окно **Анализ остатков**. С помощью функциональных кнопок в данном окне можно всесторонне просмотреть остатки модели как в графическом виде, так и в электронных таблицах.

Вначале для оценки адекватности модели лучше всего использовать визуальные методы и затем, если потребуется, перейти к статистическим.

Для оценки адекватности модели рассмотрим график остатков, например, на нормальной вероятностной бумаге (**Normal plot of resids**) (рис 2.4).

Из графика остатков на нормальной вероятностной бумаге видно, что они достаточно хорошо ложатся на прямую, которая соответствует нормальному закону. Поэтому предположение о нормальном распределении ошибок выполнено.

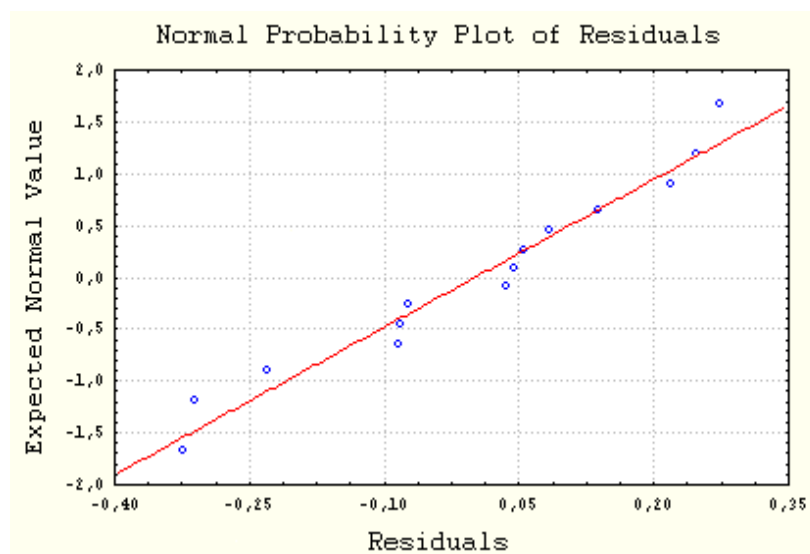


Рис 2.4. График остатков на нормальной вероятностной бумаге

Исследуем зависимость остатков от наблюдаемых значений (**Obs & residuals**), а также от предсказанных значений (**Pred & residuals**) (рис.2.5). Остатки хаотично разбросаны относительно прямой, в их поведении нет закономерностей. Нет оснований говорить, что остатки коррелированы между собой, нет также резко выделяющихся остатков. Отсюда можно заключить, что модель достаточно адекватно описывает данные.

Замечание. Следует заметить, что мы имеем очень небольшое число данных – всего 16. Поэтому мы используем графические методы оценки адекватности модели. В сложных задачах графические и статистические методы оценки адекватности должны использоваться совместно, естественно дополняя друг друга. Статистические процедуры анализа остатков собраны в левой части окна в группе *Statistics – Статистики*

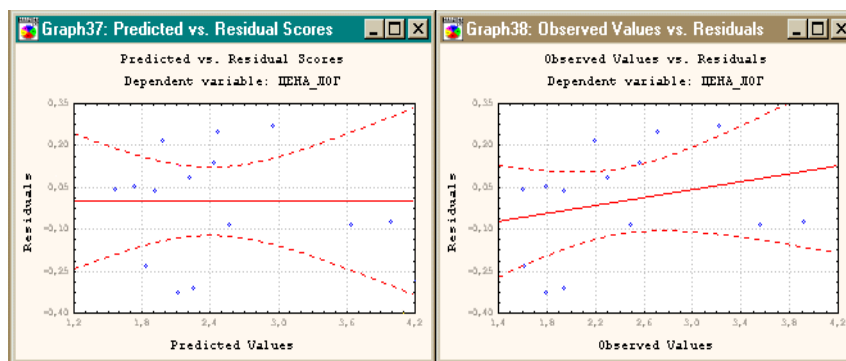


Рис 2.5 Зависимость остатков от предсказанных и наблюдаемых значений

Вывод результатов анализа в файл с отчетом

Возможен вывод результатов анализа не только в виде электронных таблиц и графиков на рабочее пространство системы. Можно также создать специальный файл с отчетом. Для него на рабочем пространстве системы откроется специальное окно, в которое можно «распечатать» любой документ, например таблицу или график.

Пример 2. Множественная регрессия

Рассмотрим теперь более сложный пример множественной регрессии. При этом в модель будет включена не одна, а несколько независимых переменных.

Описание примера

В таблице ниже приведены данные о капитальных затратах на строительство атомных электростанций с реактором водяного охлаждения.

Данные собраны для 32 различных станций США. Требуется: оценить зависимость между ценой станции и рядом параметров, приведенных в таблице, предсказать величину капитальных затрат на строительство новой станции, попробовать выделить наиболее значимые величины, влияющие на цену станции.

Данные имеют следующую структуру:

Таблица 4

№	C	D	T1	T2	S	PR	NE	CT	BW	N	PT
1	460.05	68.58	14	46	687	0	1	0	0	14	0

2	452.99	67.33	10	73	1065	0	0	1	0	1	0
3	443.22	67.33	10	85	1065	1	0	1	0	1	0
4	652.32	68.00	11	67	1065	0	1	1	0	12	0
5	642.23	68.00	11	78	1065	1	1	1	0	12	0
6	345.39	67.92	13	51	514	0	1	1	0	3	0
7	272.37	68.17	12	50	822	0	0	0	0	5	0
8	317.21	68.42	14	59	457	0	0	0	0	1	0
9	457.12	68.42	15	55	822	1	0	0	0	5	0
10	690.19	68.33	12	71	792	0	1	1	1	2	0
11	350.63	68.58	12	64	560	0	0	0	0	3	0
12	402.59	68.75	13	47	790	0	1	0	0	6	0
13	412.18	68.42	15	62	530	0	0	1	0	2	0
14	495.58	68.92	17	52	1050	0	0	0	0	7	0
15	394.36	68.92	13	65	850	0	0	0	1	16	0
16	423.32	68.42	11	67	778	0	0	0	0	3	0
17	712.27	69.50	18	60	845	0	1	0	0	17	0
18	289.66	68.42	15	76	530	1	0	1	0	2	0
19	881.24	69.17	15	67	1090	0	0	0	0	1	0
20	490.88	68.92	16	59	1050	1	0	0	0	8	0
21	567.79	68.75	11	70	913	0	0	1	1	15	0
22	665.99	70.92	22	57	828	1	1	0	0	20	0
23	621.45	69.67	16	59	786	0	0	1	0	18	0
24	608.80	70.08	19	58	821	1	0	0	0	3	0
25	473.63	70.42	19	44	538	0	0	1	0	19	0

26	697.14	71.08	20	57	1130	0	0	1	0	21	0
27	207.51	67.25	13	63	745	0	0	0	0	8	1
28	288.48	67.17	9	48	821	0	0	1	0	7	1
29	284.88	67.83	12	63	886	0	0	0	1	11	1
30	280.36	67.83	12	71	886	1	0	0	1	11	1
31	217.38	67.25	13	72	745	1	0	0	0	8	1
32	270.71	67.83	7	80	886	1	0	0	1	11	1

Здесь:

C – цена в млн долларах, приведенная к курсу 1976 г.

D – срок разрешения на строительство.

T1 – время между обращением за разрешением и получением разрешения на строительство.

T2 – время между получением оперативной лицензии и разрешением на строительство.

S – номинальная мощность электростанции, МВт.

PR – наличие в той же самой местности ранее построенной электростанции на РВО. Если значение равно 1, то имеется уже построенная станция.

NE – характеристика района, в котором строится станция.

CT – использование нагревательной башни. Если равно 1, то используется, если 0 – нет.

BW – использование силовой установки производства фирмы Babcock-Wilcox. Если значение равно 1, то используется установка этой фирмы, 0 – нет.

N – суммарное число электростанций, построенное архитектором-инженером станции.

PT – электростанция, строящаяся под частичным надзором. **PT=1**, если надзор есть, **PT=0**, если надзора нет.

Математическая постановка задачи

Для исследования данной задачи воспользуемся методами регрессионного анализа. В отличие от предыдущего, в этом примере имеется несколько независимых переменных, поэтому применяется метод множественной регрессии.

Воспользуемся векторными обозначениями. Обозначим через y – вектор наблюдений, состоящий из n элементов, через x – матрицу независимых переменных, размером m на n , где m – число независимых переменных, а n – число наблюдений.

В этих обозначениях задача может быть сформулирована следующим образом:

$$y = x\beta + \varepsilon, \quad (9)$$

где ε есть независимые случайные ошибки со средним 0 , которые интерпретируются как ошибки наблюдений, а β – вектор неизвестных параметров, которые необходимо оценить. Оценки параметров β обозначим через b .

В данном примере зависимая переменная – цена станции, а независимые – D , $T1$, $T2$, S , PR , NE , CT , BW , N , PR (т.е. все остальные переменные, перечисленные в таблице). Зависимость между переменными предполагается линейной.

Создание электронной таблицы с исходными данными

Ввод исходных данных. Текстовые и численные значения

Переменные в электронной таблице могут принимать как текстовые, так и численные значения. Текстовые значения вводятся аналогично численным. Необходимо поместить указатель на ячейку в таблице, щелкнуть левой кнопкой мыши и ввести требуемое значение с клавиатуры. Для переменных, которые принимают текстовые значения, в STATISTICA используется так называемое соглашение «двойной записи», при котором каждому текстовому значению приписывается некоторый численный эквивалент. Для просмотра переменных, принимающих текстовые значения, переключитесь в режим просмотра текстовых значений при помощи кнопки ABC на панели инструментов электронной таблицы. Например, для повышения наглядности восприятия можно ввести для переменных PR и BW текстовые значения. Для переменной PR – ДА (1) будет обозначать наличие уже построенной в этой местности станции на РВО, а НЕТ (0) – ее отсутствие. Аналогично и для переменной BW введены текстовые значения ИСП и НЕИСП для обозначения использования установок фирмы BW или нет. Для просмотра этих значений нажмите на кнопку ABC123 **Менеджера текстовых значений**.

Преобразование исходных данных

В электронных таблицах STATISTICA вы можете выполнить все необходимые преобразования. Такая задача часто возникает в процессе обработки данных. В систему STATISTICA включено большое количество общих математических и специализированных статистических функций. Для некоторых из переменных мы применим, аналогично предыдущему примеру, преобразование логарифмирования. Формулы преобразования задаются в диалоговом окне спецификаций переменной, которое вызывается двойным щелчком на имени переменной в строке заголовка

электронной таблицы. Вам, возможно, потребуется вставить дополнительные строки и столбцы в таблицу. Воспользуйтесь для этого кнопками **Vars Cases** для вызова соответствующих команд по работе с переменными и наблюдениями.

При помощи кнопки вы можете просмотреть спецификации всех переменных в электронной таблице с исходными данными.

Поставим задачу построения линейной регрессии между зависимой переменной $\text{Log}_C = \ln(C)$ и независимыми переменными D, PR, NE, CT, BW, PT, Log_N , Log_S , Log_{T1} , Log_{T2} .

Предварительный анализ и визуализация данных

Построим ряд специализированных статистических графиков для более полного исследования исходных данных. Для этого поместите указатель мыши на ту переменную в таблице, которую необходимо отобразить графически, щелкните правую кнопку мыши и из появившегося контекстного меню выберите необходимый тип графика. В этом случае в диалоговом окне определения графика при помощи кнопки **Variables – Переменные** выберите необходимые переменные, которые вы хотите отобразить графически, и необходимый тип графика. Постройте гистограмму для переменной T1 и график диапазона значений для C.

Вызов стартовой панели модуля и определение процедуры анализа

Для начала статистического анализа вам необходимо вызвать *Стартовую панель* модуля. Это основное диалоговое окно модуля, в котором необходимо задать различные опции анализа. Если *Стартовая панель* модуля закрыта, то откройте ее. Для этого зайдите в модуль **Analysis – Анализ** и выберите команду **Startup Panel – Стартовая панель**.

Выбор переменных для анализа

Далее необходимо выбрать переменные для анализа. В нашем примере имеется одна зависимая переменная Log_C и набор независимых переменных. Для их задания воспользуйтесь кнопкой **Variables – Переменные** из *Стартовой панели*.

В открывшемся окне **Select dependent and independent variable list – Выбор зависимой переменной и списка независимых переменных** выберите необходимые переменные. Для выбора переменной щелкните мышью на ее имени. Для выбора нескольких переменных удерживайте при этом клавишу CTRL. Нажмите кнопку **OK** в правом верхнем углу. Вы вновь окажетесь в *Стартовой панели* модуля **Множественная регрессия**.

Задание дополнительных параметров анализа

В *Стартовой панели* можно задать и дополнительные опции и параметры анализа. Например, вы можете выбрать определенное подмножество наблюдений для анализа, приписать веса переменным – эти опции относятся к исходным данным. Вы также можете задать и опции, которые относятся непосредственно к

статистической процедуре: задать правило обработки пропущенных данных, выбрать метод анализа по умолчанию и др. Отменим выбор метода анализа по умолчанию (**Perform default analysis**). После нажатия на кнопку **OK** появится диалоговое окно определения метода (рис 2.6.).

В прокручиваемом списке методов выберите один из пошаговых регрессионных методов, например, **Forward stepwise**, значения остальных параметров оставьте неизменными. Нажмите **OK**.

Замечание

Метод пошаговой регрессии состоит в том, что на каждом шаге в модель включается, либо исключается какая-то независимая переменная. Таким образом, выделяется множество наиболее “значимых” переменных. Это позволяет сократить число переменных, которые описывают зависимость.

В данном случае выбран пошаговый метод **включения**. При использовании этого метода в регрессионное уравнение последовательно включаются независимые переменные, пока уравнение не станет удовлетворительно описывать входные данные. Включение переменных определяется при помощи **F**-критерия.

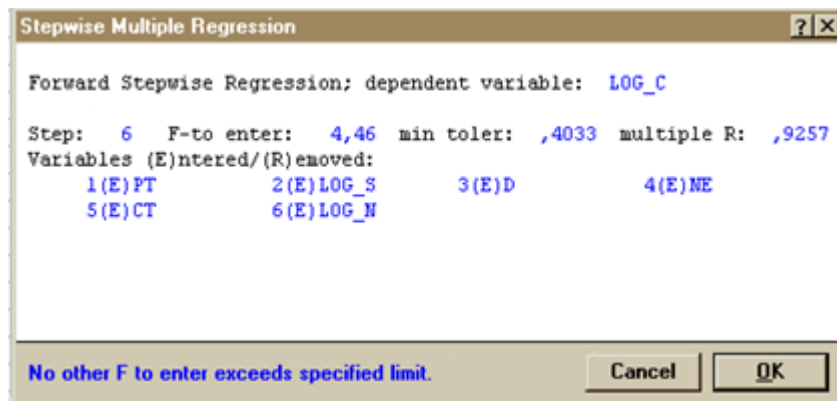


Рис 2.6. Окно результатов анализа. Отмечены переменные, которые были включены в модель

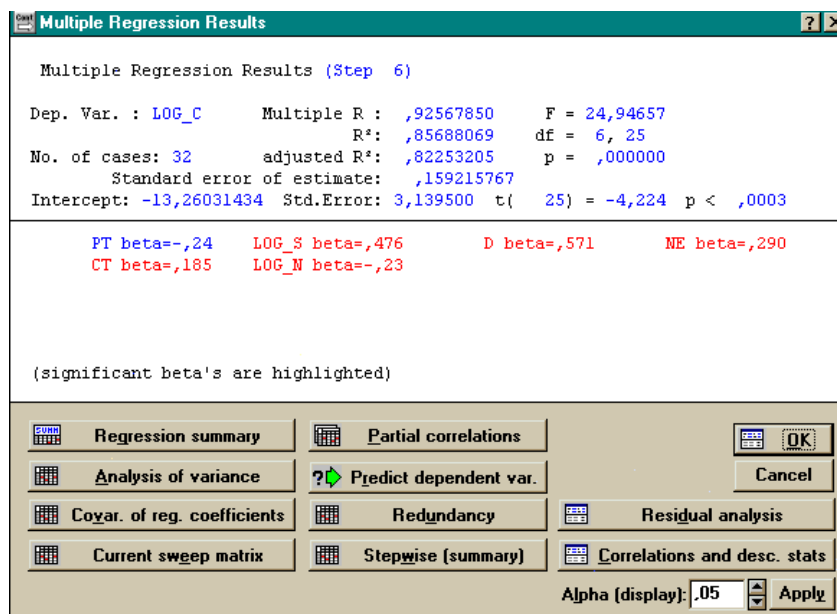


Рис 2.7 Окно с результатами анализа. Красным цветом выделены значимые коэффициенты регрессии

Вывод результатов и их анализ

В стартовой панели нажмите на кнопку **OK**. Система произведет вычисления и на экране появится окно результатов (рис.2.6.)

Нажав на кнопку **OK**, вы откроете основное окно анализа результатов (рис.2.7.).

Окно результатов анализа имеет следующую простую структуру: верхняя часть окна – информационная, нижняя содержит функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа. Оно описано в предыдущем примере.

В информационной части вы прежде всего смотрите на значение коэффициента детерминации. В нашем примере $R^2 = 0.857$. Это значит, что построенная регрессия объясняет 85.7% разброса значений относительно среднего.

Далее вы смотрите на значение F-критерия и уровень его значимости p . F-критерий используется для проверки значимости регрессии.

Щелкните далее на кнопку **Regression summary** – Краткие результаты регрессии. Вы увидите следующую электронную таблицу с результатами анализа:

Regression Summary for Dependent Variable: LOG_C					
Continue...		R= ,92567850 RI= ,85688069 Adjusted RI= ,82253205 F(6,25)=24,947 p<,00000 Std.Error of estimate: ,1592			
N=32	BETA	St. Err. of BETA	B	St. Err. of B	t(25)
Intercept			-13,2603	3,139500	-4,22370
PT	-,237240	,119142	-,2261	,113549	-1,99123
LOG_S	,475980	,078180	,7234	,118820	6,08825
D	,570609	,116206	,2124	,043259	4,91030
NE	,289875	,086299	,2490	,074137	3,35896
CT	,185358	,079778	,1404	,060425	2,32343
LOG_N	-,227767	,107867	-,0876	,041475	-2,11156

Рис 2.8 Краткие результаты регрессии

В третьем столбце расположены искомые коэффициенты. Итак, искомая регрессия имеет вид:

$$\text{Log}_C = -13.2603 + 0.2261 \cdot \text{PT} + 0.7234 \cdot \text{Log}_S + 0.2124 \cdot \text{D} + 0.249 \cdot \text{NE} + 0.1404 \cdot \text{CT} - 0.0876 \cdot \text{Log}_N$$

Качественно построенное уравнение можно интерпретировать следующим образом:

- Стоимость строительства растет с увеличением мощности станции (S), при использовании нагревательной башни и при строительстве в NE районе.
- Стоимость уменьшается с возрастанием опыта инженера-архитектора и при строительстве под частичным надзором.

Итак, на рассмотренных примерах мы проследили технологию обработки данных и стиль работы в *STATISTICA* и увидели, что даже несложные модели линейной регрессии позволяют в реальных задачах получать содержательные результаты.

Остатки

Ниже приведены графики остатков данной модели:

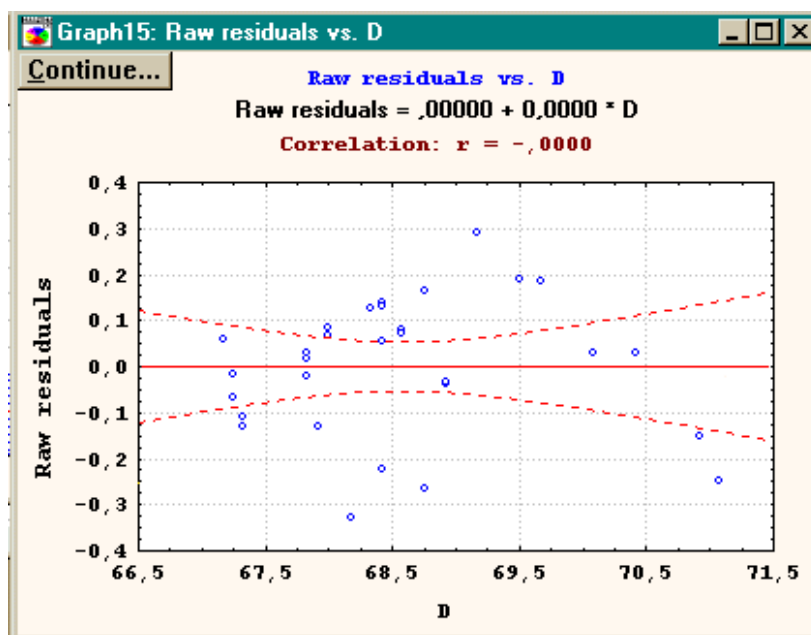


Рис. 2.9. Остатки как функция от D

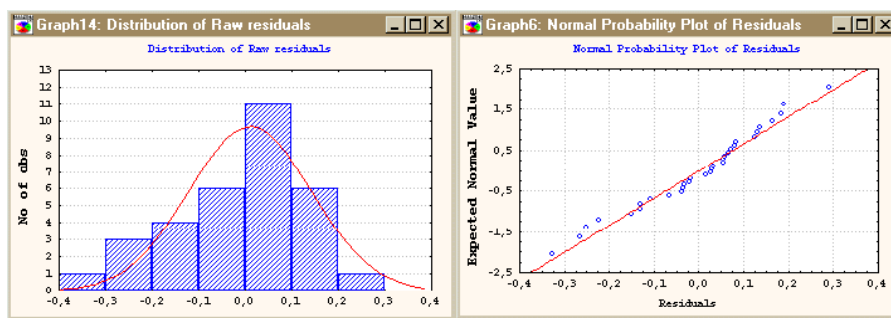


Рис 2.10 Гистограмма остатков

Рис 2.11 График остатков на НВБ

Б. Исследование статистической зависимости Вес-Рост

Постановка задачи

В первой лабораторной работе вами были собраны статистические данные измерения веса и роста студентов соответствующих групп. Результаты измерений должны быть сведены в одну общую таблицу. Найти уравнение регрессии, описывающей вес студента в зависимости от его роста. Статистические выводы обосновать.

Последовательность решения задачи

1. Подготовить таблицу.
2. Оценить корректность условий проведения регрессионного анализа
3. Провести предварительный анализ и определить предполагаемый вид функции. Выяснить, нужны ли дополнительные преобразования.

4. Провести регрессионный анализ. Результаты анализа позволят ответить на следующие вопросы:
 - Какова степень статистической связи между двумя переменными?
 - Адекватна ли модель исходным данным?
 - Значимы ли коэффициенты модели?
5. Исследовать остатки.

Лабораторная работа 3. Факторный анализ

Теоретическое введение

Цель факторного анализа - выделение из выборки величин, так называемых факторов, количество которых значительно меньше участвующих в анализе переменных. Эти факторы с достаточной точностью воспроизводят наблюдаемые корреляции между переменными и не доступны для непосредственного измерения. Не всегда эти факторы имеют практический смысл, так как в основу факторного анализа положены алгебраические методы преобразования корреляционных матриц. В психологических исследованиях методами факторного анализа объясняют взаимозависимость психологических явлений, например характер, способности, потребности и успеваемость учащихся.

Пусть R - корреляционная матрица, диагональные элементы этой матрицы равны 1. Для $n=4$, где n - количество переменных, корреляционная матрица имеет вид

$$R = \begin{bmatrix} 1 & x_{12} & x_{13} & x_{14} \\ x_{21} & 1 & x_{23} & x_{24} \\ x_{31} & x_{32} & 1 & x_{34} \\ x_{41} & x_{42} & x_{43} & 1 \end{bmatrix}$$

На первом этапе факторного анализа корреляционная матрица R преобразуется в редуцированную матрицу R_h , главное отличие которой от исходной корреляционной матрицы заключается в том, что ее диагональные элементы не равны 1 и называются **оценками общности**. Преобразование корреляционной матрицы R в R_h составляет проблему общности.

$$R_k = \begin{bmatrix} r_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & r_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & r_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & r_{44} \end{bmatrix}$$

На втором этапе факторного анализа путем алгебраического преобразования матрицы R_h осуществляется выделение m факторов, $m \ll n$. Это преобразование происходит при минимальной потере информации. Если обозначить матрицу факторов как \mathbf{A} , а через \mathbf{A}^T ее транспонированную матрицу, в которой изменены столбцы на строки и строки на столбцы, то проблему **выделения факторов** можно представить в виде

$$R = A A^T$$

или

$$R_k = \begin{bmatrix} r_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & r_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & r_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & r_{44} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} [a_1 \ a_2 \ a_3 \ a_4]$$

В приведенном выражении показана матрица \mathbf{A} с выделенным одним фактором. Аналогично представляются матрицы и с несколькими факторами. Значения a_1, a_2, \dots, a_n - называются факторными нагрузками. Факторные нагрузки определяют в какой степени выделенный фактор описывается каждой переменной и имеют смысл коэффициента корреляции переменной и фактора.

Очень редко полученные на этом этапе результаты могут найти практическое применение в силу того, что нет четкой связи выделенного фактора с переменными. Такая связь считается удовлетворительной, если факторные нагрузки по абсолютному значению превышают 0.7-0.8. В связи с этим возникает третья проблема факторного анализа - **проблема вращения**.

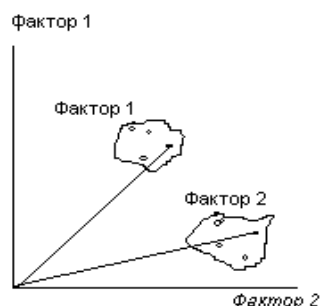


Рис.3.1.

После выделения факторов их взаимное положение в пространстве переменных можно представить геометрически в следующем виде (рис.3.1)

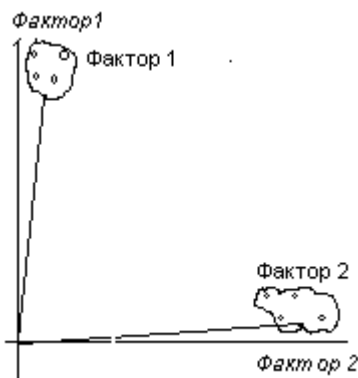


Рис.3.2.

После выполнения вращения положение факторов в пространстве изменяется. Вместе с этим изменяются и факторные нагрузки (рис.3.2).

Для вращения используются различные методы, в том числе метод главных компонент, варимаксный метод и другие. В основе этих методов положены различные способы выбора системы координат, приводящие к расположению факторов в ортогональных (взаимно перпендикулярных) плоскостях.

Наконец, последняя проблема применения факторного анализа заключается в оценке значений факторов и их содержательной интерпретации.

А. Пример использования модуля «Факторный анализ»

В файле *factor.sta* собраны результаты опросов 100 взрослых людей относительно степени их удовлетворенности жизнью. Это типичный пример данных, возникающих при социологических опросах. В файле даны значения следующих 10 переменных:

1. *Work 1* – удовлетворенность работой, – первая компонента,
2. *Work 2* – удовлетворенность работой, – вторая компонента,
3. *Work 3* – удовлетворенность работой, – третья компонента,
4. *Hobby 1* – удовлетворенность свободным временем, – первая компонента,
5. *Hobby 2* – удовлетворенность свободным временем, – вторая компонента,
6. *Home 1* – удовлетворенность домашней жизнью, – первая компонента,
7. *Home 2* – удовлетворенность домашней жизнью, – вторая компонента,
8. *Home 3* – удовлетворенность домашней жизнью, – третья компонента,
9. *Miscel 1* – общая удовлетворенность, – первая компонента,
10. *Miscel 2* – общая удовлетворенность, – вторая компонента.

Многомерность каждой переменной объясняется тем, что рассматриваются различные аспекты удовлетворенности, например, вы можете быть удовлетворены зарплатой, получаемой на работе, но не удовлетворены коллективом или тем, сколько времени тратите, чтобы до нее добраться, и т.д. Фактически здесь имеется 4 многомерных переменных.

Щелкните в этом окне на кнопку **Shrink** — *Сократить*. Далее щелкните на кнопку **Ok** — *Да* и вернитесь в стартовое окно модуля.

Файл открыт и переменные для анализа выбраны. Теперь можно начать анализ по выявлению главных факторов, влияющих на удовлетворенность человека жизнью.

Начало анализа

Щелкнув в стартовом окне модуля на кнопку **Ok** — *Да*, вы начнете анализ выбранных переменных.

STATISTICA обработает пропущенные значения тем способом, какой вы ей указали, вычислит корреляционную матрицу и предложит на выбор несколько методов факторного анализа.

Вычисление корреляционной матрицы, если она не задается сразу, — первый этап факторного анализа.

Итак, пусть вы имеете исходный данные, как в файле `factor.sta`.

После щелчка на кнопку **Ok** — *Да* перед вами появится окно **Define Method of Factor Extraction** — *Определить метод выделения факторов*. Данное окно имеет следующую структуру.

Первая, верхняя часть окна является информационной: здесь сообщается, что пропущенные значения обработаны методом **casewise**. Обработано 100 случаев и 100 случаев принято для дальнейших вычислений. Корреляционная матрица вычислена для 10 переменных.

Вторая, нижняя часть диалогового окна **Define Method of Factor Extraction** — *Определить метод выделения факторов* содержит 4 функциональные кнопки и опции выбора метода — левая часть окна, — а также поля, в которых проводятся установки для итеративного вычисления общностей.

Обратите внимание на поля в правой части окна:

- **Maximum no. of Factors** — *Максимальное число факторов*;
- **Minimum eigenvalue** — *Минимальное собственное значение*.

Эти поля определяют число факторов, которые будут выделены системой. Собственные значения меньше указанного в поле игнорируются.

Рассмотрим функциональные клавиши окна.

Из 4 функциональных кнопок — две стандартные: кнопка **Ok** — *Да* нажимается для запуска вычислительной процедуры, когда выбран метод и проведены установки в полях. Кнопка **Cancel** — *Отменить* прекращает работу в окне и возвращает в стартовое окно модуля.

Инициировав кнопку **Review corr_s/means/SD** — *Просмотреть корреляции/средние/стандартные отклонения*, вы откроете окно *Просмотреть описательные статистики*.

С помощью кнопки **Perform multiple regression** — *Выполнить множественную регрессию* вы можете выполнить множественную регрессию, не выходя из модуля.

Группа опций, объединенных под заголовком **Extraction Method** — *Метод выделения*, позволяет выбрать метод обработки.

В зависимости от критерия оптимальности возможен анализ либо методом **Principal components** — *Методом главных компонент*, либо одним из методов, объединенных в группу **Principal factor analysis** — *Анализ главных (общих) факторов*.

Система предлагает следующие методы в группе **Principal factor analysis** — *Анализ главных факторов*:

Communalities=multiple R2** — *Общности равны квадрату коэффициента множественной корреляции*.

Iterated communalities (MINRES) — *Итеративных общностей (минимальных остатков)*.

Centroid method — *Центроидный метод*.

Principal axis method — *Метод главных осей*.

В окне **Define Method of Factor Extraction** — *Определить метод выделения факторов* щелкните мышью по кнопке **Review corr_s/means/SD** — *Просмотреть корреляции/средние /стандартные отклонения*.

Перед вами появится окно просмотра описательных статистик для анализируемых данных, где можно посмотреть средние, стандартные отклонения, корреляции, ковариации, построить различные графики.

Щелкните по кнопке **Correlations** — *Корреляции*.

Вы увидите на экране корреляционную матрицу выбранных ранее переменных.

Вы можете сохранить корреляционную матрицу, инициировав кнопку **Save correlations** — *Сохранить корреляции*. Далее в модуле **Факторный анализ** при дальнейших исследованиях этих данных можно сразу работать с корреляционной матрицей.

Методическое замечание: часто на главной диагонали корреляционной матрицы вместо единиц ставятся квадраты коэффициентов множественной корреляции или

другие оценки общностей. По определению квадрат коэффициента множественной корреляции случайной величины X со случайными величинами $Y(1), \dots, Y(l)$ равен

$$\min E(X - (a(1)*X(1) + \dots + a(l)*X(l)))^2,$$

где минимум берется по всевозможным наборам значений $a(1), \dots, a(l)$.

Щелкните на кнопку **C**ontinue — *Продолжить* в верхнем левом углу таблицы и вернитесь в окно **D**efine Method of Factor Extraction — *Определить метод выделения факторов*.

Выберите опцию **P**roincipal components — *Главные компоненты* и щелкните на **O**k — *Да*.

Система быстро произведет вычисления, и на экране появится окно **F**actor **A**nalysis **R**esults — *Результаты факторного анализа*.

Структура окна «Результаты факторного анализа»

В верхней части окна **F**actor **A**nalysis **R**esults — *Результаты факторного анализа* дается информационное сообщение:

Number of variables — *Число анализируемых переменных* 10,

Method — *Метод анализа: главные компоненты*,

log(10) determination of correlation matrix — *Десятичный логарифм детерминанта корреляционной матрицы*: -4.1096,

Number of factor extraction — *Число выделенных факторов*: 2,

Eigenvalues — *Собственные значения*: 6.11837, 1.80068.

В нижней части окна находятся функциональные кнопки, позволяющие всесторонне просмотреть результаты анализа численно и графически.

Обратите внимание на кнопку **F**actor **r**otation — *Вращение факторов*.

Щелкнув по данной кнопке, вы откроете окно **F**actor **R**otation — *Вращение факторов*.

В окне **F**actor **R**otation вы можете выбрать различные повороты осей.

Если пространство общих факторов найдено, то с помощью поворота системы координат в принципе можно получить бесчисленное множество решений.

Важно найти интерпретируемое решение!

Вы можете искать нужный поворот эмпирически, однако STATISTICA предлагает несколько полезных процедур. Попробуйте оценить их возможности.

Возможны следующие методы:

- **V**arimax — *варимакс*
- **B**iquartimax — *биквартимакс*

- **Quartimax** — *квартимакс*
- **Equamax** — *эквимакс*

Дополнительный термин в названии методов — *normalized* – *нормализованные* — указывает на то, что факторные нагрузки в процедуре нормализуются, т.е. делятся на корень квадратный из соответствующей общности. Термин **raw** — *исходные* показывает, что вращаемые нагрузки не нормализованы.

Например, щелкнув по кнопке **Varimax normalized** — *Варимакс нормализованный*, вы проведете вращение факторных нагрузок методом варимакс, тогда как опция **Varimax Raw** — *Варимакс исходных* означает применение метода *варимакс* к не-нормализованным (исходным) факторам.

Щелкните на кнопку **Plot of loadings, 2D** — *Двумерный график нагрузок* и посмотрите результаты факторного анализа на графике (рис 3.3.).

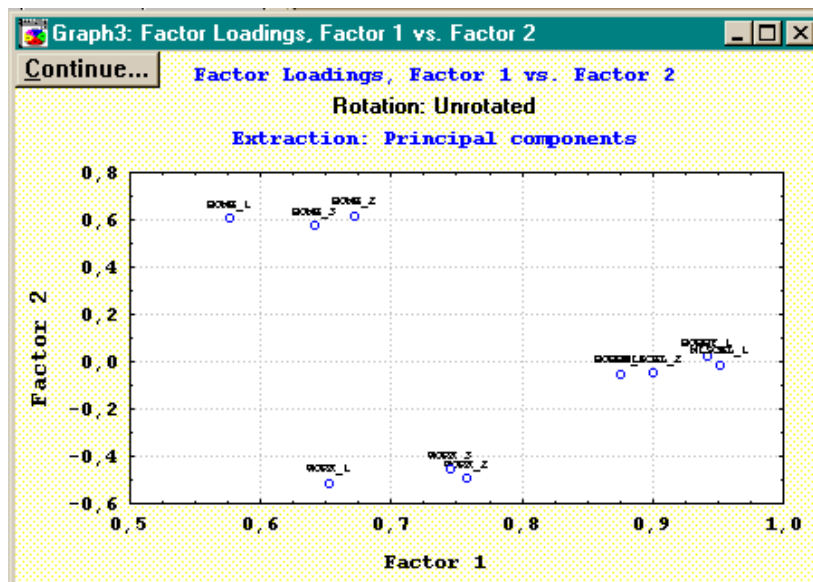


Рис. 3.3. Факторное решение для данных «Удовлетворенность жизнью»

Щелкнув по кнопке **Factor loadings** — Факторные нагрузки, вы сможете посмотреть нагрузки численно:

Variable	Factor 1	Factor 2
WORK_1	.652601	-,514217
WORK_2	.756976	-,494770
WORK_3	.745706	-,456680
HOBBY_1	.941630	,021835
HOBBY_2	.875615	-,051643
HOME_1	.576062	,604977
HOME_2	.671289	,617962
HOME_3	.641532	,573925
MISCEL_1	.951516	-,013513
MISCEL_2	.900333	-,048154
Expl. Var	6,118369	1,800682
Prp. Totl	,611837	,180068

Рис.3.4. Таблица факторных нагрузок

Рассматривая решение на графике, вы видите, что его трудно проинтерпретировать. Не понятно, какой смысл придать двум выделенным факторам и как в этих терминах описывать удовлетворенность жизнью индивидуума.

В таких случаях целесообразно прибегнуть к повороту осей, надеясь получить решение, которое можно проинтерпретировать в предметной области.

Щелкните кнопку **Factor rotation** — *Вращение факторов*.

В появившемся окне **Factor Rotation** — Вращение факторов иницируйте кнопку **Varimax normalized** — *Варимакс нормализованный*.

Система произведет вращение факторов методом нормализованного варимакса, и окно **Factor Analysis Results** — *Результаты факторного анализа* снова появится на мониторе. Вновь иницируйте в этом окне кнопку **Plot of loadings, 2D** — *Двумерный график нагрузок*. Вы опять увидите двумерный график нагрузок:

Этот график отличается от предыдущего. Теперь найденное решение уже можно интерпретировать. Посмотрим еще нагрузки численно, иницировав кнопку **Factor loadings** — Факторные нагрузки.

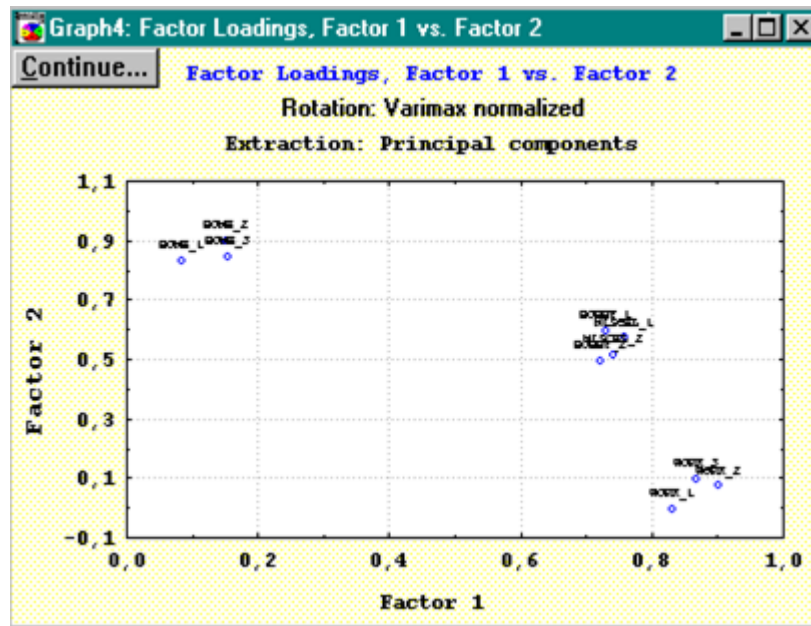


Рис. 3.5. Факторное решение после поворота осей

На этом графике выделено два общих фактора Factor 1, Factor 2: Factor 1 отвечает за удовлетворение, получаемое опрошенными на работе, удовлетворенность работой (значения переменных Work 1 — Work 3 максимально большие по этой оси и малы по другой). Относительно Factor 2 можно сказать, что он измеряет удовлетворенность домашней жизнью.

Глядя на эти результаты, вы можете сделать вывод, что общая удовлетворенность человека определяется лишь двумя факторами: удовлетворенностью работой и удовлетворенностью домом.

Variable	Factor 1	Factor 2
WORK_1	.830827	-.005746
WORK_2	.901325	.073641
WORK_3	.869058	.096808
HOBBY_1	.730235	.594896
HOBBY_2	.723177	.496371
HOME_1	.083802	.831157
HOME_2	.151040	.899830
HOME_3	.154555	.846798
MISCEL_1	.759727	.573045
MISCEL_2	.740557	.514289
Expl. Var	4.493483	3.425568
Prp. Totl	.449348	.342557

Рис.3.6. Таблица факторных нагрузок после применения метода главных компонент и вращения факторов

Методические замечания

1. Выбор числа факторов. Не существует общего решения о выборе числа факторов, однако существует несколько полезных процедур, которыми пользуются на практике.

Критерий Кайзера. Сначала мы можем оставить только факторы, соответствующие собственным значениям, большим 1. Это означает, что мы оставляем факторы, которые дают дисперсию не меньше, чем исходные переменные. Используя этот критерий, в нашем примере мы должны сохранить 2 фактора (главные компоненты).

Критерий каменистой осыпи. Графический метод, называемый критерием каменистой осыпи (the scree test). Мы можем показать собственные значения на простом линейном графике. Предлагается найти точку на графике, справа от которой собственные значения уменьшаются гладко. Справа от этой точки возможно найти только "факторную каменистую осыпь". Согласно этому критерию, мы должны сохранить 2 или 3 фактора в нашем примере.

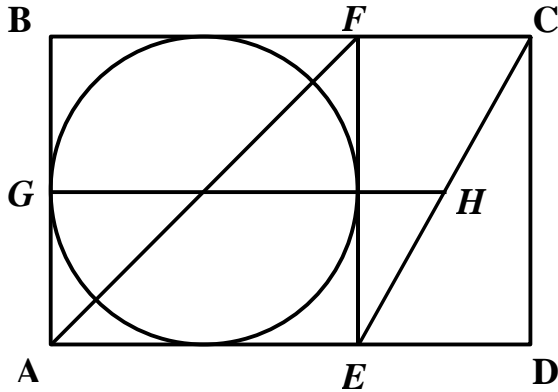
Первый метод (критерий Кайзера) иногда сохраняет слишком многие факторы, а вторая техника (критерий каменистой осыпи) — слишком мало.

2. Критерий согласия для необходимого числа факторов. Если оценки нагрузок строятся методом максимального правдоподобия, то имеется возможность использовать статистический критерий проверки гипотезы о числе факторов в модели. Критерий доступен по клавише **Goodness of Fit Test – Критерий согласия** в окне **Factor Analysis Results — Результаты факторного анализа**. Эта клавиша доступна лишь в тех случаях, когда решение найдено методом максимального правдоподобия.

Б. Задание для самостоятельного анализа

Описание задачи

Пусть объектом исследования будет обычный прямоугольник, описываемый восьмью признаками. Скрытыми факторами будут являться высота и длина прямоугольника.



Признаки несут следующую смысловую нагрузку:

x_1 = периметр ABCD

x_2 = длина окружности

x_3 = периметр AFE

x_4 = отрезок ED

x_5 = периметр EFCD

x_6 = диагональ AF

x_7 = периметр ABFE

x_8 = средняя линия трапеции ABCE

Таким образом, каждый из признаков, входящий в исследуемый набор, может быть представлен как функция двух общих факторов и специфического фактора:

1. $X_1=2F_1+2F_2+U_1$
2. $X_2=\pi F_1+U_2$
3. $X_3=(2+\sqrt{2})F_1+U_3$
4. $X_4=F_1-F_2+U_4$
5. $X_5=2F_2+U_5$
6. $X_6=\sqrt{2}F_1+U_6$
7. $X_7=4F_1+U_7$
8. $X_8=0.5F_1+0.5F_2+U_8$

Априорно известно, что признаки распределены по нормальному закону. Требуется провести факторный анализ и определить скрытые определяющие факторы. Создайте таблицу данных из 8 переменных и 100 объектов, используя специальную функцию генерации случайных нормально распределенных чисел. Проведите факторный анализ.

Порядок проведения анализа

1. Расчет корреляционной матрицы
2. Выделение первоначальных ортогональных факторов
3. Вращение факторов
4. Оценка значений факторов

Со держание отчета:

1. Корреляционная матрица
2. Результаты факторного анализа
3. Таблица собственных значений
4. Таблица общностей
5. Остаточная матрица корреляций
6. Таблица факторных нагрузок до вращения
7. График факторных нагрузок до вращения
8. Таблица факторных нагрузок после вращения
9. График факторных нагрузок после вращения
10. Оценки общих факторов

Проведите аналогичный анализ, включив факторы «высота» и «длина» прямоугольника в список общих переменных.

Лабораторная работа 4. Кластерный анализ

Анализ данных, основанный на выделении кластеров. Под кластером понимают группу переменных в статистическом анализе с общими признаками. В кластерном анализе используются следующие типы классификации:

1. **Исключающие (неисключающие).** В исключающей классификации один элемент может появляться в только в одном подмножестве, а в неисключающих – в нескольких;
1. **Внутренние (внешние).** В первом случае классификация основывается на заданном наборе переменных. При внешней классификации используются другие переменные, которые объединяют несколько переменных ;
2. **Иерархические (неиерархические).** В иерархической классификации группы выбираются таким образом, чтобы каждая была возможно более однородной. Неиерархические методы классификации, как правило, не используются;

3. **Агломеративные (дивизивные).** Различия между этими классификациями состоит в направлении объединения подмножеств. В агломеративной классификации переменные объединяются в подмножества возрастающего объема. В дивизивной классификации переменные постепенно разделяются до тех пор, пока не будет получена заданная степень разделения;
4. **Монотетические (политетические)** . В монотетической классификации деление производится на основании одного признака, имеющего максимальную информативность, тогда как в политетической классификации все признаки учитываются в равной степени.

В пакете STATISTICA реализованы три метода кластерного анализа:

- **Дерево кластеров (Tree Clustering).** Метод основан на иерархической классификации. В психологических исследованиях этот метод применяется наиболее часто. В этом методе реализуется 7 различных правил объединения (по выбору):
 - Простое связывание переменных (Single Linkage);
 - Полное связывание (Complete Linkage);
 - Невзвешенное парно-групповое среднее (Unweighted Pair-group Average);
 - Взвешенное парно-групповое среднее (Weighted Pair-group Average);
 - и др. правила.

По умолчанию используется простое связывание переменных.

В качестве меры сходства используется Евклидова метрика (Euclidean Distances), Квадрат Евклидовой метрики (Squared Euclidean Distances), Манхэттенская метрика (Manhattan Distances) и др.

2. **Алгоритм K внутри групповых средних (K - means Clustering Results)** . Метод основан на выявлении кластера путем определения минимума суммы квадратов расстояний всех точек , входящих в кластерную область, до центра кластера. Качество алгоритма зависит от выбора исходных центров кластеров и их числа.
3. **Влочная кластеризация (Block Clustering).** Метод основан на моделировании кластеров переменных и вариант (Case). Использует эвристический параметр Treshold.

Постановка задачи

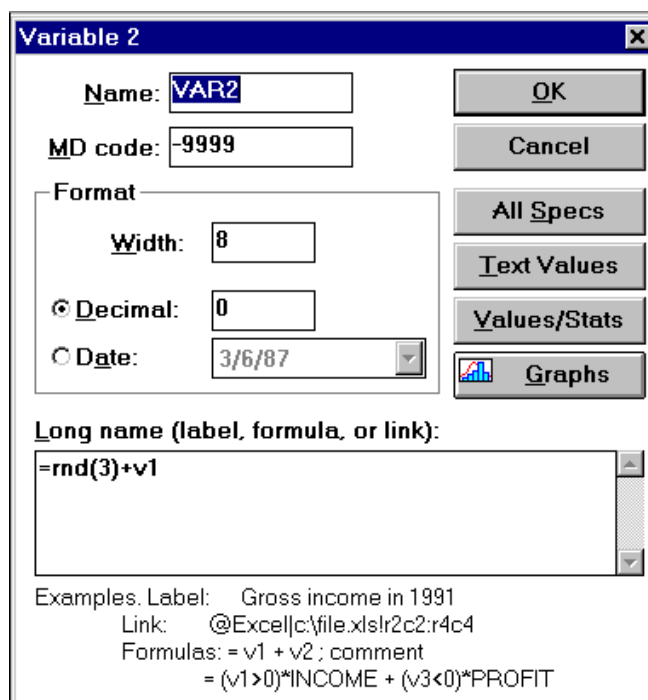
Сгенерировать выборку случайных величин, образующих несколько кластеров, по следующим законам:

1. $VAR1 = RND(7)$, где функция $RND(7)$ представляет собой псевдослучайное число распределенное равномерно в интервале $[0,7]$. Эти числа получаются по специальному алгоритму, производящему некоторые простые арифметические действия над исходным числом. При каждом последующем обращении базовое число изменяется. Это приводит к появлению псевдослучайной последовательности чисел.
1. $VAR2 = V1 + RND(3)$, где $V1$ - сокращенное название переменной $VAR1$, а $RND(3)$ - случайное число, распределенное равномерно в интервале $[0,3]$. Это предполагает, что между переменной $V1$ и $V2$ существует статистическая связь, которая должна проявиться как кластерное образование.
2. $VAR3 = RND(7)$. Хотя переменные $VAR1$ и $VAR3$ генерируются по одним и тем же законам и имеют одинаковые статистики (выборочное среднее, выборочную дисперсию и т.д.) тем не менее в каждой конкретной реализации (Case) они существенно различаются.
3. $VAR4 = V3 + RND(3) - RND(3)$. Переменная $V4$ также статистически связана с переменной $V3$, но по другому закону.
4. $VAR5 = V1 + RND(3)$. Эта переменная статистически зависима от переменной $V1$.

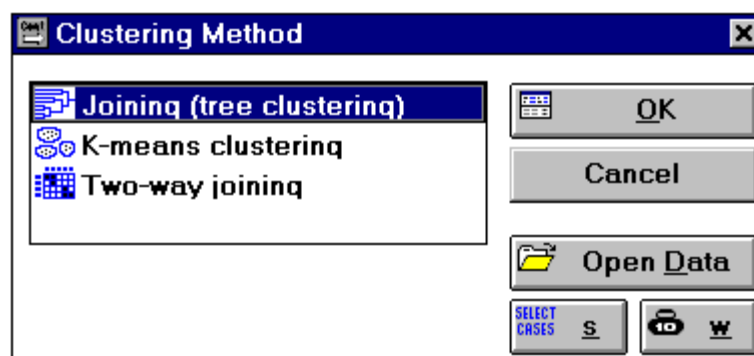
Таким образом, в результате кластерного анализа должно проявиться по крайней мере 2 кластерных образования, хотя по сгенерированной выборке их обнаружить невозможно.

Последовательность решения задачи

1. Запустить пакет STATISTICA.
1. В появившемся окне статистического анализа выбрать Clustering Method и нажать клавишу Switch To (Переключить).
2. В появившемся верхнем горизонтальном меню выбрать раздел File, а затем New Data.. В результате этих действий должна появиться электронная таблица с стандартным обозначением переменных $VAR1, VAR2, \dots, VAR10$ и 10 строками.
3. Преобразовать таблицу. С помощью мышки выделить переменные $VAR6-VAR10$ и удалить их выбрав в меню VAR функцию DELETE. Нажатием на кнопку Cases выбрать функцию ADD (Добавить строки таблицы). Затем в появившемся окне заполнить необходимые данные о количестве добавляемых строк и после какой строки это выполнить (90, 10).

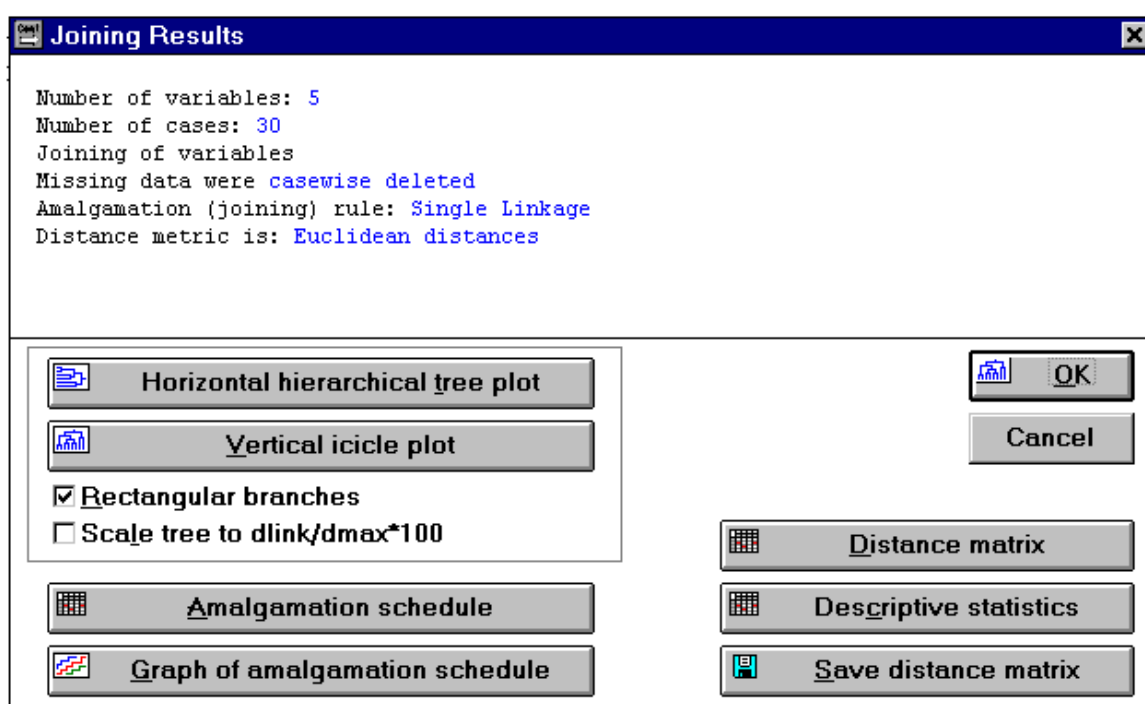
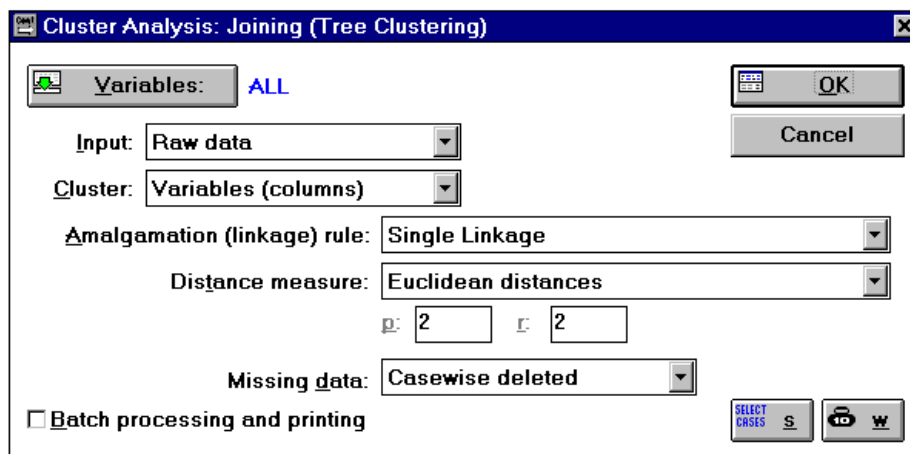


1. Ввести формулы для моделирования значений переменных по исходному заданию. Для этого выделить с использованием мышки переменную и нажать комбинацию клавиш Ctrl+F2. В появившемся окне (рис. 33) установить поле Decimal равным 0, а в поле Long Name записать формулу для моделирования переменной по заданным исходным данным. Точно также повторить для других переменных. После нажатия клавиши Ok дать подтверждение на расчет значений переменной в электронной таблице.
1. Раскрыть панель для выбора метода анализа, выбрав в горизонтальном меню раздел Analysis (рис.34). Выбрать метод Tree Clustering и нажать Ok.



1. Ввести необходимые исходные данные для проведения анализа (рис.35). Переменные ALL (Все), способ выделения кластеров по указанным переменным - Variable(columns), метод выделения кластеров Single Linkage, мера сходства и различия признаков - Euclidian Distances. После ввода всех данных нажать клавишу Ok.

Результаты анализа можно увидеть через промежуточную форму (рис.36).

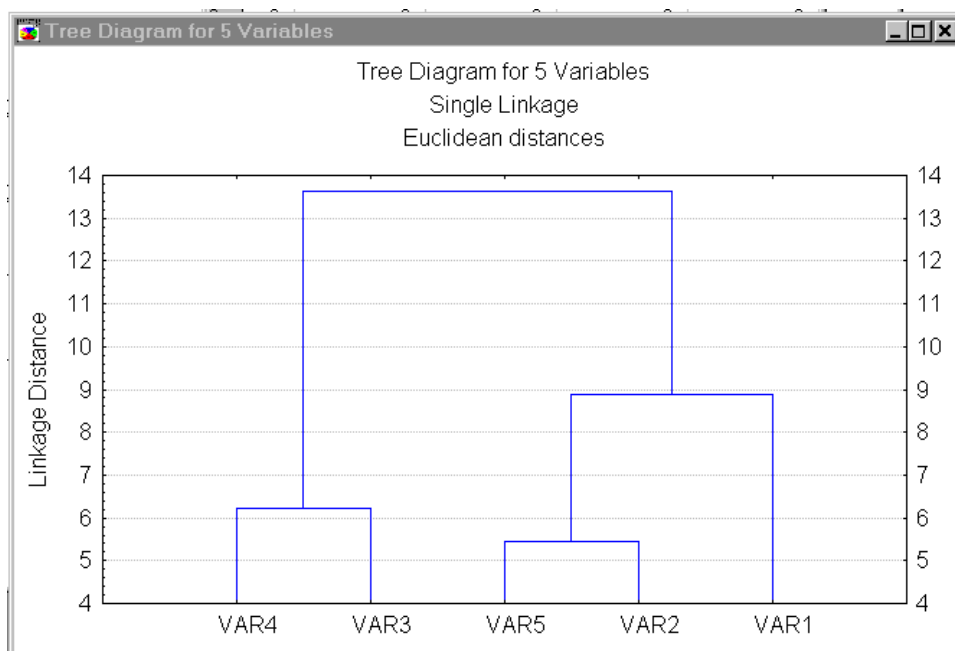


Кнопки Horizontal hierarchical tree plot и Vertical icicle plot позволяют наглядно представить дерево кластеров в иерархическом виде (рис.37). Переключатель Rectangular Branches меняет форму представления дерева кластеров с прямоугольной на треугольную. Переключатель Scale tree масштабирует изображения относительно максимальной длины связи. Следующие две кнопки Amalgamation (Образования) позволяют показать меры сходства в табличном и графическом видах.

Выводы

Результаты анализа дерева кластеров показывают, что заданная исходная модель данных практически оказалась выделенной в виде двух кластеров с

переменными VAR3 и VAR4 , а также VAR5, VAR2 и VAR1. Кроме того, выделено кластер VAR5 и VAR2, что связано с незначительными отклонениями в их математических моделях. Это показывает и описательная статистика.



Задание

1. Повторить расчеты, изменив закон формирования $VAR5 = v1 + RND(5)$. Что при этом измениться ? Сделать выводы.
2. Как изменяются результаты выполнения кластерного анализа, если изменить масштаб одной или двух переменных, умножив их значения на 10 ? Выводы обосновать.
3. Провести анализ с результатами данных таблицы 5 и сделать статистически обоснованные выводы.
4. Провести анализ с результатами данных таблицы 6 и сделать статистически обоснованные выводы.
5. Измените способ группирования данных в кластеры не по переменным (столбцам), а по строкам (Case). Как в этом случае интерпретируются результаты, например таблицы 5 ?
6. Попробуйте проверить какой-либо тест- опросник на предмет выделения заложенных в него шкал на вашей группе. Сделайте статистически обоснованные выводы. (Не выполнять в этом 2000 году)

Выберите какой—либо подходящий файл статистических данных из каталога с примерами, входящего в состав пакета Statistica, и проведите самостоятельный кластерный анализ.

САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩИХСЯ

Виды самостоятельной работы распределяются в течение семестра. Подготовка к промежуточной аттестации ведется в установленные календарным учебным графиком сроки.

Рекомендуемый перечень учебной литературы:

1. Мхитарян В.С. Статистика. В 2-х частях. Часть 1. Учебник и практикум для академического бакалавриата. Издательство: Юрайт, 2020
2. Дуброва. Анализ данных: Учебник. - Москва: Издательство Юрайт, 2018.
3. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009.
4. Соловьев В.И. Анализ данных в экономике: Теория вероятностей, прикладная статистика, обработка и анализ данных в Microsoft Excel. Учебник. издательство: КноРус, 2021
5. Миркин. Введение в анализ данных [Электронный ресурс]: Учебник и практикум. - Москва: Издательство Юрайт, 2019. - 174
6. Воронов В. И., Воронова Л. И., Усачев В. А.. Data Mining - технологии обработки больших данных [Электронный ресурс]: Учебное пособие. - Москва: Московский технический университет связи и информатики, 2018. - 47 с.
7. Федин Ф. О., Федин Ф. Ф.. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс]: Учебное пособие. - Москва: Московский городской педагогический университет, 2012. - 204 с. – Режим доступа: <http://www.iprbookshop.ru/26444.html>
8. Нейтан Яу - Искусство визуализации в бизнесе. Как представить сложную информацию простыми способами
9. Джин Желязны - Говори на языке диаграмм. Пособие по визуальным коммуникациям
10. Методы бизнес-анализа / Пол Тернер, Джеймс Кадл, Дебра Пол
11. Ключевые показатели эффективности. 75 показателей, которые должен знать каждый менеджер / Бернارد Марр.

Ниже приведены ресурсы информационно-телекоммуникационной сети «Интернет», полезные для самостоятельной работы.

1. Бизнес-аналитика. Введение https://sf.education/vvedenie-v-biznes-analitiku?admitad_uid=b045cb655ef2605764eff6e800a29960&utm_source=admitad&utm_campaign=442763&utm_medium=cpa
2. Анализ данных - Мхитарян В.С. Учебные материалы для студентов https://studme.org/93298/statistika/analiz_dannyh
3. Студия машинного обучения AZURE. <https://azure.microsoft.com/ru-ru/services/machine-learning-studio/>
4. Платформа машинного обучения ML.NET от компании Microsoft. <https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet>
5. Видео об обработке и анализе данных для начинающих <https://docs.microsoft.com/ru-ru/dotnet/machine-learning/resources/basics>
6. Школа прикладного бизнес-анализа <https://babok-school.ru/blogs/dfd-diagram-practical-example/>

ТЕСТОВЫЕ ЗАДАНИЯ

Тестовые задания по теме «Введение в бизнес-аналитику. Предварительный анализ данных»

1. Как называется совокупность подходов и методов для автоматизации анализа текстовых документов, включая задачи установления степени сходства текстов, категоризации документов, формирования аннотаций и пр.?
 1. Генетические алгоритмы (Genetic algorithms)
 2. Вычислительный интеллект (Computational intelligence)
 3. Анализ текстов (Text analysis)
 4. Кластер-анализ (Cluster analysis)
2. К классу описательных задач относятся:
 1. кластеризация и классификация;
 2. кластеризация и поиск ассоциативных правил;
 3. классификация и регрессия;
 4. классификация и поиск ассоциативных правил.
3. К классу предсказательных задач относятся:
 1. кластеризация и классификация;
 2. кластеризация и поиск ассоциативных правил;
 3. классификация и регрессия;
 4. классификация и поиск ассоциативных правил.
4. К классу задач supervised learning (обучение с учителем) относятся:
 1. кластеризация и классификация;
 2. кластеризация и поиск ассоциативных правил;
 3. классификация и регрессия;
 4. классификация и поиск ассоциативных правил.
5. К классу задач unsupervised learning (обучение без учителя) относятся:
 1. кластеризация и классификация;
 2. кластеризация и поиск ассоциативных правил;
 3. классификация и регрессия;
 4. классификация и поиск ассоциативных правил.
6. Задача классификации сводится к ...
 1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристикам;
 3. определение по известным характеристикам объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
7. Задача регрессии сводится к ...
 1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристиками;
 3. определение по известным характеристикам объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик в всем множестве анализируемых данных.
8. Задача кластеризации заключается в ...
 1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристиками;

3. определение по известным характеристиками объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
9. Целью поиска ассоциативных правил является ...
1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристиками;
 3. определение по известным характеристиками объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
10. Модели классификации описывают ...
1. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 2. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 3. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 4. группы, на которые можно разделить объекты, данные о которых подвергаются анализу.

Тестовые задания по теме «Исследование зависимостей»

1. Регрессивные модели описывают ...
 1. правила или набор правил, в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
 2. функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
 3. функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
 4. группы, на которые можно разделить объекты, данные о которых подвергаются анализу
2. Задача регрессии сводится к ...
 1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристикам;
 3. определение по известным характеристикам объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
3. Многомерный анализ – это:
 1. одновременный анализ по нескольким измерениям;
 2. одновременный анализ по нескольким параметрам;
 3. одновременный анализ по нескольким данным.
4. Коэффициент парной корреляции между процентом охвата населения прививками и заболеваемостью на 10 000 населения равен (-0,86). Можно сделать следующие выводы:
 1. связь между изучаемыми явлениями отсутствует, т.к. коэффициент корреляции отрицательный
 2. связь между изучаемыми явлениями обратная (отрицательная)
 3. связь между изучаемыми явлениями сильная и обратная
 4. связь между изучаемыми явлениями слабая и обратная

5. связь между изучаемыми явлениями средняя и обратная.
5. При каком значении коэффициента корреляции связь можно считать умеренной?
1. $r = 0,47$;
 2. $r = 0,71$.
 3. $r = 0,21$;
6. Верно или нет следующее утверждение (и почему): если линия регрессии имеет большой наклон, то корреляция между переменными также должна быть большой?
1. Да, потому что при большем наклоне увеличение одной переменной ведет к более сильному увеличению другой переменной
 2. Да, потому что коэффициент корреляции пропорционален углу наклона линии регрессии.
 3. Нет, корреляция не зависит угла наклона.
7. Определить коэффициент корреляции двух переменных: агрессивности (x_a) и IQ у школьников (y_{iq}) по полученным данным тестирования

№	X_{agr}	Y_{iq}
1	24	100
2	27	115
3	26	117
4	21	119
5	20	134
6	31	94
7	26	105
8	22	103
9	20	111
10	18	124
11	30	122
12	29	109
13	24	110
14	26	86

1. 0,27
 2. -0,42
 3. -0,64
 4. 0,81
8. Знания 10 студентов проверены по двум тестам А и В. Оценки по стобалльной системе сведены в таблицу

A	95	90	86	84	75	70	62
	60	57	50				
B	92	93	83	80	55	60	45
	72	62	70				

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками по двум тестам.

1. 0,27
 2. 0,42
 3. 0,64
 4. 0,81
9. По данным задачи №8 определить коэффициент ранговой корреляции Кендалла
1. 0,28
 2. 0,47
 3. 0,63

4. 0,81
10. Если величина коэффициента корреляции Пирсона между переменными равна $-0,75$ то
1. это очень слабая корреляция и в большинстве случаев мы не берем ее в расчет;
 2. это высокая корреляция и мы принимаем ее в расчет.
11. Что означает совсем низкое или нулевое значение коэффициента корреляции двух количественных признаков?
1. наличие неизвестного вида связи
 2. наличие квадратичной зависимости
 3. отсутствие линейной связи
 4. наличие линейной связи
12. Термин регрессия в статистике понимают как: а) функцию связи, зависимости; б) направление развития явления вспять; в) функцию анализа случайных событий во времени; г) уравнение линии связи
1. а, б
 2. в, г
 3. а, г
13. Для анализа связи между двумя номинальными признаками составляют следующую таблицу. Строки таблицы соответствуют категориям одного признака, а столбцы — категориям другого признака. Элемент на пересечении строки и столбца — количество объектов, обладающих соответствующими категориями и того и другого признаков. Такая таблица называется таблицей _____
14. Рассмотрим модель «мешок слов» в задаче выявления корреляции по данным следующей таблицы.

Статья	Ключевые слова									
	пить	равный	греть	играть	легкий	цена	свобода	талант	налог	женский
F1	1	2	0	1	2	0	0	0	0	2
F2	0	0	0	1	0	1	0	2	0	2
F3	0	2	0	0	0	0	0	1	0	2
F4	2	1	0	0	0	2	0	2	0	1
E1	2	0	1	2	2	0	0	1	0	0
E2	0	1	0	3	2	1	2	0	0	0
E3	1	0	2	0	1	1	0	3	1	1
E4	0	1	0	1	1	0	1	1	0	0
H1	0	0	2	0	1	2	0	0	2	0
H2	1	0	2	2	0	2	2	0	0	0
H3	0	0	1	1	2	1	1	0	2	0
H4	0	0	1	0	0	2	2	0	2	0

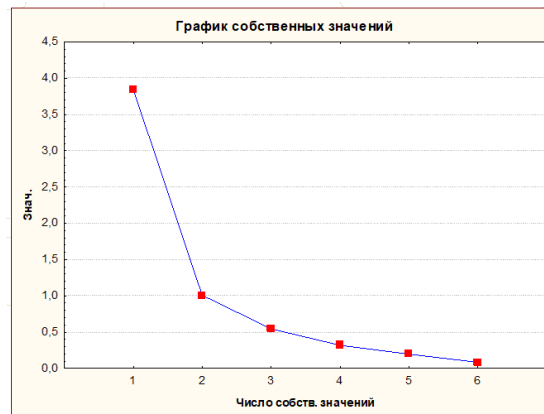
Объектами являются газетные статьи, разделенные на три категории в соответствии с темами «Феминизм», «Развлечения» и «Домохозяйство». Каждая статья характеризуется своим набором ключевых слов, представленных в соответствующей строке таблицы. Чтобы уменьшить эффект случайности отбора статей в таблицу данных, примем, что «мешок» содержит по одному появлению каждого ключевого слова, независимо от того, появилось ли оно в статьях данной категории или нет. Какова вероятность слова «играть» в категории H?

1. $2/41$
2. $7/41$
3. $4/41$

Тестовые задания по теме «Снижение размерности признакового пространства»

1. Укажите действия, позволяющих уменьшить число факторов:
 1. Устранение дублирующей информации при наличии сильно коррелированных признаков
 2. Редукция слабоинформативных (маломеняющихся для различных объектов) признаков
 3. Агрегирование (объединение) нескольких признаков в один
2. Математический метод нахождения главных осей заключается в
 1. вычислении собственных чисел и собственных векторов ковариационной матрицы S .
 2. вычислении детерминанта ковариационной матрицы S
3. Даны четыре примера (наблюдения) в трехмерном пространстве признаков: $A(1;4;10)$, $B(2;5;6)$, $C(1;3;8)$ и $D(2;4;8)$. В результате применения метода главных компонент исходное пространство признаков свели к двумерному пространству признаков на плоскости. Найдите евклидово расстояние между примерами C и D в редуцированном пространстве с точностью до одного знака после запятой: _____
4. Какова идея метода главных компонент?
 - 1) поиск гиперплоскости заданной размерности, такой чтобы ошибка проектирования выборки на данную гиперплоскость была минимальной
 - 2) поиск проекции на гиперплоскость с сохранением большей части дисперсии в данных
 - 3) проекция данных на гиперплоскость с критической ошибкой проектирования
5. Укажите верное утверждение
 1. Метод главных компонент использует меньшее количество компонент, в отличие от метода независимых компонент
 2. Метод главных компонент добивается ортогональности между полученными компонентами, а метод независимых компонент - не ортогональности
 3. Метод независимых компонент работает с коррелированными данными, в отличие от метода главных компонент
 4. Метод главных компонент применяется в основном для задач, где необходимо разделять сигналы, а метод независимых компонент - для визуального разделения данных
6. Каковы недостатки метода главных компонент?
 1. координаты объектов в новом пространстве определены не однозначно
 2. проблема с вычислением собственных векторов ковариационной матрицы, при большом количестве данных
 3. существует произвол в выборе координат объектов в новом пространстве
 4. общая сложность алгоритма
7. Выберите сферы применения метода главных компонент
 1. Визуализация данных
 2. Построение деревьев решений
 3. Обработка изображений

4. Выявление максимальной избыточности
5. Отбор признаков
8. Укажите достоинство использования метода главных компонент
 1. Простой алгоритм
 2. Координаты объектов в новом пространстве определены однозначно
 3. Легкость с вычислением собственных векторов ковариационной матрицы в случае большого количества данных
 4. все перечисленное
9. Согласно критерию каменистой осыпи число общих факторов равно _____

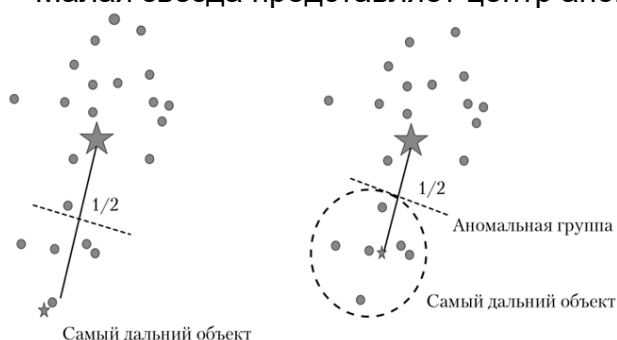


10. По критерию Кайзера отбираются только факторы с собственными значениями
 1. равными или большими 1
 2. меньшими 1
 3. меньше натурального логарифма

Тестовые задания по теме «Классификация многомерных наблюдений»

1. Кластеризация — ...
 1. это установление зависимости непрерывной выходной переменной от входных переменных
 2. эта группировка объектов (Наблюдений, событий) на основе данных, описывающих свойства объектов
 3. выявление закономерностей между связанными событиями
 4. это установление зависимости дискретной выходной переменной от входных переменных.
2. Задача кластеризации заключается в ...
 1. нахождения частых зависимостей между объектами или событиями;
 2. определения класса объекта по его характеристикам;
 3. определение по известным характеристикам объекта значение некоторого его параметра;
 4. поиска независимых групп и их характеристик во всем множестве анализируемых данных.
3. Расставьте в правильном порядке шаги алгоритма К-средних
 1. Предобработка
 2. Инициализация аномального центра
 3. Обновление аномального центра
 4. Обновление аномальной группы
 5. Выдача результатов

4. Поставьте в правильном порядке шаги алгоритма иK-средних (t), где t — порог разрешения, т. е. задаваемое пользователем минимальное количество объектов в аномальной группе, необходимое, чтобы она могла восприниматься как генератор отдельного кластера.
 1. Условие остановки
 2. Метод K-средних
 3. Настройка
 4. Аномальная группа
 5. Отбрасывание малых кластеров
5. На рисунке изображены первая (слева) и финальная (справа) итерации извлечения аномальной группы из структуры некоторого множества объектов. Малая звезда представляет центр аномальной группы.



Как называется точка, обозначенная большой звездой?

1. центральная точка
 2. реперная точка
 3. нормальная точка
 4. аномальная точка
6. Спорный объект кластеризации — это объект, который по мере сходства ...
 1. может быть отнесен к нескольким кластерам
 2. не может быть отнесен ни к одному кластеру
 3. может быть отнесен более чем к двум кластерам
 7. Процедура, которая приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некоей величине, отражающей определенные свойства, это — ...
 1. стандартизация
 2. нормирование
 3. оба ответа верны
 8. Работа кластерного анализа опирается на следующие предположения (выберите неверный ответ):
 1. рассматриваемые признаки объекта в принципе допускают желательное разбиение объектов на кластеры
 2. правильность выбора масштаба или единиц измерения признаков
 3. отнесение всех объектов к одному из predetermined признаков
 9. Иерархические агломеративные методы характеризуются ...
 1. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
 2. делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп

3. сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
10. Объект относится к кластеру, если ...
 1. расстояние от объекта до центра кластера меньше радиуса кластера
 2. расстояние от объекта до центра кластера меньше диаметра кластера
 3. расстояние от объекта до центра кластера больше радиуса кластера
11. Иерархические дивизимные методы характеризуются ...
 1. последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров
 2. делением одного кластера на меньшие кластеры, в результате образуется последовательность расщепляющих групп
 3. сопоставлением фиксированного числа кластеров наблюдения кластерам так, что средние в кластере максимально возможно отличаются друг от друга
12. При применении кластерного анализа переменные ...
 1. должны измеряться в сравнимых шкалах
 2. могут измеряться в каких угодно шкалах
 3. должны быть только числовыми
13. Характеристикой каких групп методов являются последовательное объединение исходных элементов и соответствующее уменьшение числа кластеров?
 1. иерархические агломеративные методы
 2. иерархические дивизимные (делимые) методы
 3. и тех, и других
14. Деление одного кластера на меньшие кластеры, в результате чего образуется последовательность расщепляющих групп. Характеристика каких групп методов описана выше?
 1. иерархические агломеративные методы
 2. иерархические дивизимные (делимые) методы
 3. и тех, и других

ЭКЗАМЕН

Вопросы к экзамену

1. Основные задачи, решаемые в анализе данных.
2. Этапы решения задач. Постановка задачи. Ввод данных в обработку.
3. Этапы решения задач. Качественный анализ. Количественное описание данных. Интерпретация результатов.
4. Многомерные случайные величины. Моменты второго порядка случайной величины.
5. Многомерные случайные величины. Выборочные ковариация и корреляция.
6. Введение в многомерный стохастический анализ. Исследование зависимостей. Основные прикладные цели.
7. Классификация статистических переменных. Виды статистического анализа.
8. Этапы корреляционного анализа. Причинный и статистический характер связи. Виды характеристик связи. Природа статистической взаимосвязи.
9. Количественные характеристики статистической взаимосвязи. Диаграммы рассеивания. Ковариация.
10. Коэффициент парной корреляции. Корреляционная матрица.
11. Частные коэффициенты корреляции. Коэффициент множественной корреляции
12. Корреляционная функция.
13. Характеристики многомерной статистической связи.
14. Связь порядковых переменных. Коэффициент Спирмена. Коэффициент Кендалла.
15. Анализ связей между классификационными переменными. Таблицы сопряженности. Проверка гипотезы о наличии связи.
16. Регрессионный анализ. Общая постановка задачи. Основные этапы регрессионного анализа
17. Выбор модели регрессии.
18. Уравнение регрессии. Оценка параметров модели
19. Анализ регрессии. Адекватность модели. Дисперсия адекватности и дисперсия воспроизводимости.
20. Проверка значимости коэффициентов регрессионной модели.
21. Исследование регрессионных остатков
22. Оценка точности регрессионной модели.
23. Постановка задачи МГК
24. Вычисление главных компонент.
25. Факторный анализ (ФА). Общая постановка задачи. Цели ФА.
26. Порядок выполнения факторного анализа. Основные этапы факторного анализа.
27. Определение главных факторов – факторизация. Задание числа факторов.
28. Факторный анализ. Вращение факторов.
29. Факторный анализ. Оценка значений общих факторов.
30. Кластерный анализ (КА) Общая постановка задачи. Этапы кластерного анализа.
31. Кластерный анализ. Меры сходства. Расстояние между отдельными объектами. Расстояние между классами в кластерном анализе.
32. Обзор методов кластерного анализа.
33. Методы кластерного анализа. Иерархические методы.
34. Методы кластерного анализа. Итеративные методы. Алгоритм k-средних.

35. Методы ассоциативного поиска.

**МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ
ЛАБОРАТОРНЫХ РАБОТ**

Дисциплина (модуль)	<u>Методы управления знаниями и принятия решений</u> <i>наименование дисциплины (модуля)</i>
Уровень образования	<u>магистратура</u> <i>(СПО/бакалавриат/магистратура/специалитет/подготовка кадров высшей квалификации)</i>
Квалификация	<u>Магистр</u> <i>Техник/Бакалавр/Магистр/Инженер/ Исследователь. Преподаватель-исследователь</i>

г. Ульяновск, 2021

Лабораторная работа 1

Разработка онтологии предметной области с использованием инструментального средства онтологического проектирования **OntoStudio**

1. Для выбранной предметной области (*соответствующей теме магистерской диссертации*) выделить не менее 30 понятий (концептов).
2. Дать определения этим понятиям (*то есть составить тезаурус*).
3. На множестве понятий ввести отношения и функции интерпретации для построения онтологии по предметной области. Построить онтологию, используя языки и инструментальные средства онтологического проектирования (например, **OntoStudio**).
4. Осуществить поиск информации по разработанной предметной онтологии.
5. В отчет по [лабораторной работе](#) включить обзор по методам и средствам онтологического проектирования и возможностям выбранного инструментального средства (*например, **OntoStudio***), а также файлы с онтологией для выбранной предметной области и сеть понятий, полученную в результате [визуализации](#) онтологии.
6. Подготовить отчет для защиты лабораторной работы.
7. Также необходимо быть готовым устно ответить на контрольные вопросы.

Требования

Выбрать предметную область для создания онтологии (в случае если вопрос с темой магистерской диссертации не решен, а предмет нужно сдать). Номер темы определяется последней цифрой номера зачетной книжки:

- 1) Алгоритмизация
- 2) Операционные системы
- 3) Программирование
- 4) Алгоритмические языки
- 5) Параллельные вычислительные процессы
- 6) СУБД
- 7) Internet
- 8) Телекоммуникации
- 9) Экспертные системы
- 10) Имитационное моделирование

Методические указания

Лабораторная работа может быть выполнена с помощью программного продукта **OntoStudio**, который представляет собой визуальный инструментарий для создания и редактирования предметных онтологий. Сформированная онтология может быть выражена на языке RDF.

Коротко рассмотрим структуру онтологии, с которой оперирует OntoStudio. Ядром онтологии служит иерархия концептов (абстрактных понятий предметной области (ПрО)). Иерархические отношения соответствуют типу род-вид и используются в механизмах наследования. Для описания прочих, неиерархических типов связей между концептами предназначены бинарные отношения. Концептам приписываются атрибуты, рассматриваемые как отношения определенного типа между концептом и значением. Реализации концептов (экземпляры) представляют конкретные сущности. Онтология также содержит [аксиомы](#), под которыми понимаются правила, справедливые в моделируемой ПрО. Подобные правила выражаются на основе отношений. Правило может включать одно

или несколько отношений. Например: «ЕСЛИ X <является отцом> Y И Y <имеет пол> <мужской>, ТО Y <является сыном> X ». В этом примере X и Y – концепты, <является отцом> и <является сыном> – отношения между ними, а <имеет пол> – отношение, задающее для Y атрибут «пол». Различные типы аксиом используются для описания ограничений целостности, накладываемых на онтологию, функциональных связей, логических зависимостей и отношений между фактами.

Разработка онтологии с помощью редактора *OntoStudio*

Проиллюстрируем пример создания предметной онтологии с помощью редактора *OntoStudio*. В начале определяется уникальное имя онтологии. Далее создаются концепты ПрО, их атрибуты и отношения между концептами (рис. 1).

Редактор *OntoStudio* позволяет задавать **отношения** четырех видов:

1. **простое** – отношение между парой концептов от первого концепта ко второму;
2. **симметричное** – отношение между парой концептов, не зависящее от их порядка (т. е. отношение работает в обе стороны);
3. **транзитивное** – данное отношение означает, что если два концепта A и B связаны отношением R , а также B и C связаны отношением R , тогда A и C связаны отношением R ;
4. **инверсное** – отношение, обратное к указанному отношению, оба отношения задаются для одной пары концептов.

После создания концептов, их атрибутов и отношений между концептами, редактор *OntoStudio* позволяет создавать экземпляры концептов (определять интерпретацию концептов онтологии) и данные указанных экземпляров. Данными экземпляра являются

атрибуты и ссылки (отношения) на экземпляры других концептов (рис. 2).

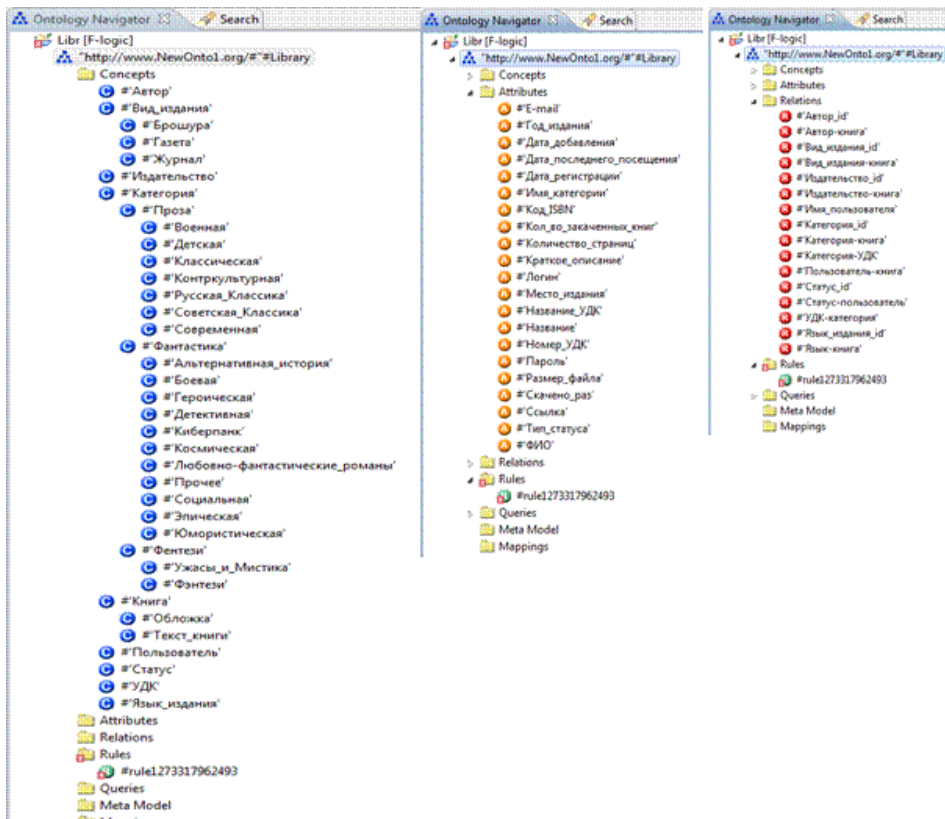


Рис. 1. Описание концептов, атрибутов и отношений в *OntoStudio*

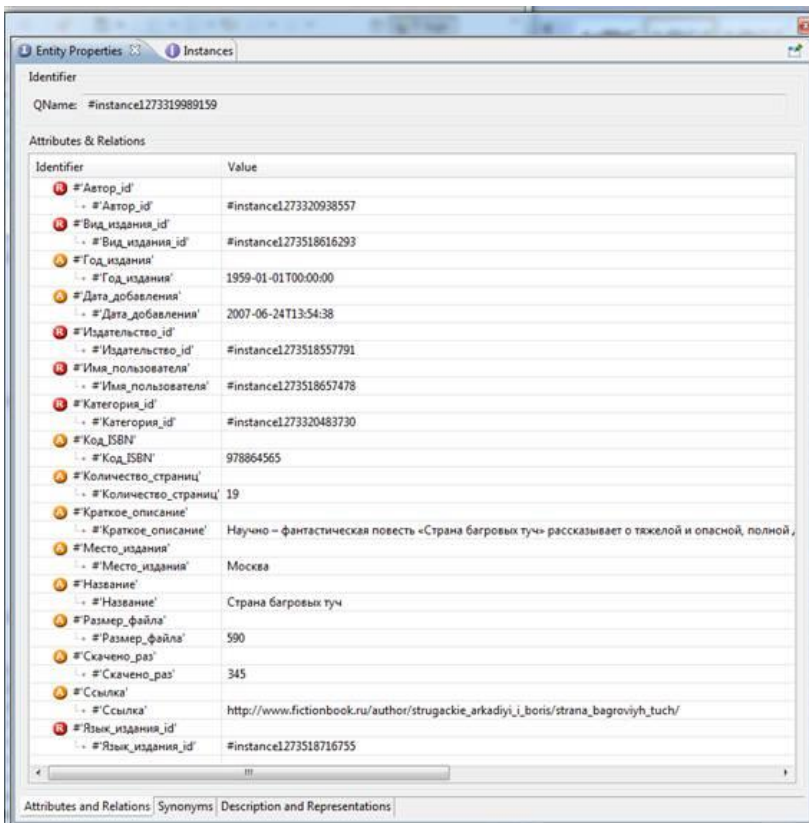


Рис. 2. Пример описания экземпляра концепта

После заполнения онтологии экземплярами концептов и формирования их данных становится возможным осуществлять поиск информации. С этой целью следует создать запрос по интересующему пользователя концепту, а затем выполнить его.

Например, сформируем запрос для поиска всех книг написанных братьями Стругацкими. Для этого следует выбрать концепт «Книга» и в выпадающем меню выбрать пункт «*Newquery*». В появившемся окне необходимо отметить галочками те атрибуты, которые нужны

для вывода результата. После нажатия на кнопки «Автор» следует задать имя автора книг и выполнить запрос (рис. 3).

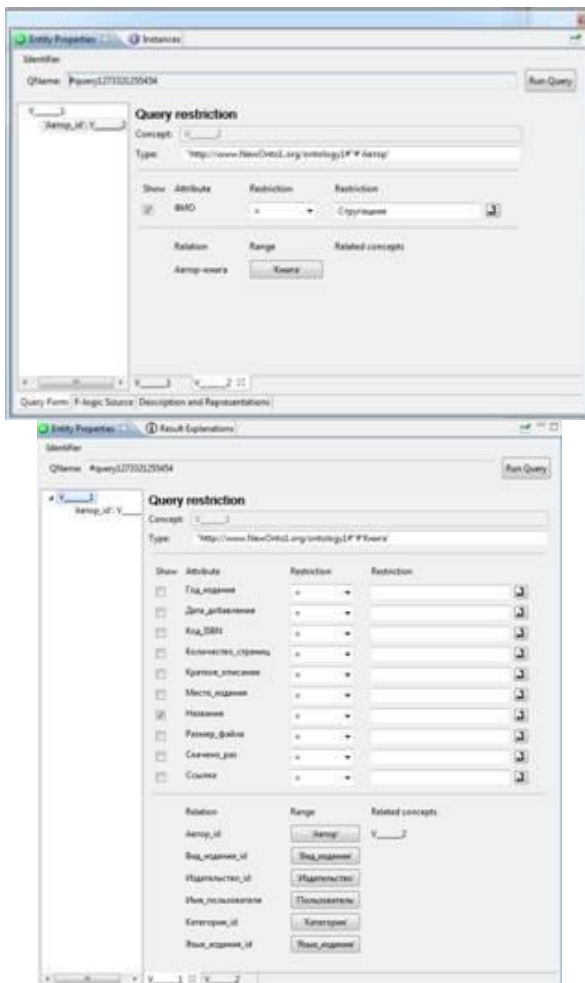


Рис. 3. Пример построения поискового запроса

Результат его выполнения приведен на рис. 4.

W__1__	W__2__
"Страна багровых туч"	"Стругащие"
"Обитаемый остров"	"Стругащие"

Рис. 4. Результат выполнения поискового запроса

Редактор *OntoStudio* содержит средства для создания сети понятий, чтобы наглядно продемонстрировать концепты и отношения между ними (рис. 5). В данную сеть можно также включить атрибуты концептов.

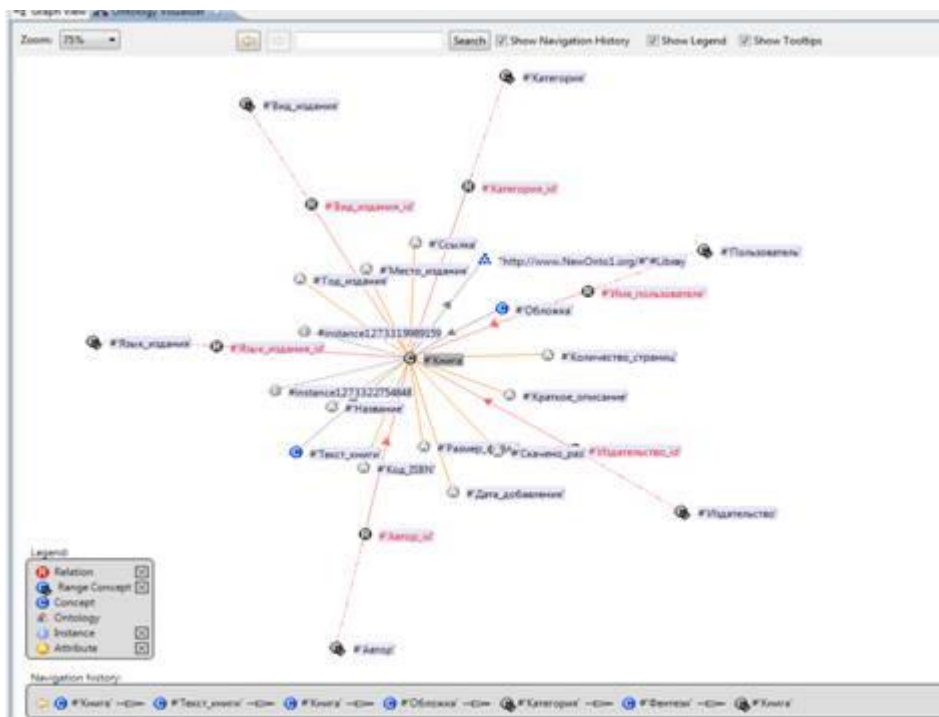


Рис. 5. Визуализация предметной онтологии в *OntoStudio*

Таким образом, программный продукт *OntoStudio* очень удобен для построения онтологии ПрО. Он содержит различные инструменты для рассмотрения всей ПрО как единого целого и обеспечивает возможность создания конкретных экземпляров концептов ПрО.

Ссылка для скачивания *OntoStudio*

(Важно! Устанавливается Trial-версия на 90 дней)

<http://www.softpedia.com/get/Science-CAD/OntoStudio.shtml>

Контрольные вопросы

1. Охарактеризуйте различные интерпретации понятия «онтология».
2. Как представляется модель онтологии?
3. Что такое модель расширенной онтологии? Охарактеризуйте ее компоненты.
4. Какие этапы построения онтологии предусмотрены стандартом *IDEF5*?
5. Каково назначение онтологии верхнего уровня? Приведете примеры таких онтологий.
6. Каково назначений онтологии предметного уровня? Приведете примеры таких онтологий.
7. Перечислите основные возможности редактора онтологий *OntoStudio*.

Лабораторная работа № 2

Онтологический инжиниринг знаний в системе PROTÉGÉ

Для начала необходимо понять, для чего должна быть использована онтология и как примерно выглядел бы ее детальный и общий вариант. Затем среди многих альтернатив вы должны будете выбрать ту, которая будет лучше всего работать для намеченной задачи, а также будет наиболее интуитивна, расширяема, поддерживаема. При этом необходимо помнить, что онтология представляет предметную область в реальном окружающем мире, и потому понятия в онтологии также должны отражать эту реальность. После того как вы сделаете черновой вариант онтологии, вы можете проверить ее и откорректировать, используя Protégé, или путем обсуждения с экспертами в исследуемой области (как результат, почти наверняка придется пересмотреть черновой вариант). Такой процесс итеративной коррекции, вероятней всего, будет продолжаться на протяжении всего жизненного цикла онтологии.

Что такое онтология?

Онтология описывает основные концепции (положения) предметной области и определяет отношения между ними. Процесс построения онтологий состоит из создания следующих блоков:

- Классов и их свойств (classes, properties).
- Свойств каждой концепции, описывающих различные функциональные возможности и атрибуты концепции (слоты (slots), иногда называемые роли).
- Ограничения по слотам (также известных как аспекты/границы (slotfacets), иногда называемые ограничения ролей).

Онтология вместе с множеством индивидуальных экземпляров классов составляют базу знаний.

В литературе по искусственному интеллекту содержится много определений понятия онтологии, многие из которых противоречат друг другу. В этой статье онтология – формальное явное описание понятий в рассматриваемой предметной области (классов, иногда их называют понятиями), свойств каждого понятия, описывающих различные свойства и атрибуты понятия (слотов (иногда их называют ролями или свойствами)), и ограничений, наложенных на слоты (фацетов, иногда их называют ограничениями ролей). Онтология вместе с набором индивидуальных экземпляров классов образует базу знаний. В действительности, трудно определить, где кончается онтология и где начинается база знаний.

Зачем создавать онтологию?

В последние годы разработка онтологий – явное формальное описание терминов предметной области и отношений между ними – переходит из мира лабораторий по искусственному интеллекту на рабочие столы экспертов по предметным областям. Во всемирной паутине онтологии стали обычным явлением. Онтологии в сети варьируются от больших таксономий, категоризирующих веб-сайты (как на сайте Yahoo!), до категоризаций продаваемых товаров и их характеристик (как на сайте Amazon.com). Во многих дисциплинах сейчас разрабатываются стандартные онтологии, которые могут использоваться экспертами по предметным областям для совместного использования и аннотирования информации в своей области. Например, в области медицины созданы большие стандартные, структурированные словари, такие как SNOMED и семантическая сеть Системы Унифицированного Медицинского Языка (theUnifiedMedicalLanguageSystem). Также появляются обширные общецелевые онтологии. Например, Программа ООН по развитию (theUnitedNationsDevelopmentProgram) и компания Dun&Bradstreet объединили усилия для разработки онтологии UNSPSC, которая предоставляет терминологию товаров и услуг (<http://www.unspsc.org/>). Онтология определяет общий словарь для

ученых, которым нужно совместно использовать информацию в предметной области. Она включает машинно-интерпретируемые формулировки основных понятий предметной области и отношения между ними. Почему возникает потребность в разработке онтологии? Вот некоторые причины:

- Для совместного использования людьми или программными агентами общего понимания структуры информации.
- Для возможности повторного использования знаний в предметной области.
- Для того чтобы сделать допущения в предметной области явными.
- Для отделения знаний в предметной области от оперативных знаний.
- Для анализа знаний в предметной области.

Совместное использование людьми или программными агентами общего понимания структуры информации является одной из наиболее общих целей разработки онтологий. К примеру, пусть несколько различных веб-сайтов содержат информацию по медицине или предоставляют информацию о платных медицинских услугах, оплачиваемых через Интернет. Если эти веб-сайты совместно используют и публикуют одну и ту же базовую онтологию терминов, которыми они все пользуются, то компьютерные агенты могут извлекать информацию из этих различных сайтов и накапливать ее. Агенты могут использовать накопленную информацию для ответов на запросы пользователей или как входные данные для других приложений. Обеспечение возможности использования знаний предметной области стало одной из движущих сил недавнего всплеска в изучении онтологий. Например, для моделей многих различных предметных областей необходимо сформулировать понятие времени. Это представление включает понятие временных интервалов, моментов времени, относительных мер времени и т.д. Если одна группа ученых детально разработает такую онтологию, то другие могут просто повторно использовать ее в своих предметных областях. Кроме того, если нам нужно создать большую онтологию, мы можем интегрировать несколько существующих онтологий, описывающих части большой предметной области. Мы также можем повторно использовать основную онтологию, такую как UNSPSC, и

расширить ее для описания интересующей нас предметной области. Создание явных допущений в предметной области, лежащих в основе реализации, дает возможность легко изменить эти допущения при изменении наших знаний о предметной области. Жесткое кодирование предположений о мире на языке программирования приводит к тому, что эти предположения не только сложно найти и понять, но и также сложно изменить, особенно непрограммисту. Кроме того, явные спецификации знаний в предметной области полезны для новых пользователей, которые должны узнать значения терминов предметной области. Отделение знаний предметной области от оперативных знаний – это еще один вариант общего применения онтологий. Мы можем описать задачу конфигурирования продукта из его компонентов в соответствии с требуемой спецификацией и внедрить программу, которая делает эту конфигурацию независимой от продукта и самих компонентов. После этого мы можем разработать онтологию компонентов и характеристик ЭВМ и применить этот алгоритм для конфигурирования нестандартных ЭВМ. Мы также можем использовать тот же алгоритм для конфигурирования лифтов, если мы предоставим ему онтологию компонентов лифта. Анализ знаний в предметной области возможен, когда имеется декларативная спецификация терминов. Формальный анализ терминов чрезвычайно ценен как при попытке повторного использования существующих онтологий, так и при их расширении. Часто онтология предметной области сама по себе не является целью. Разработка онтологии сродни определению набора данных и их структуры для использования другими программами. Методы решения задач, доменнезависимые приложения и программные агенты используют в качестве данных онтологии и базы знаний, построенные на основе этих онтологий.

Как создать онтологию?

Строго говоря, единого универсального подхода к созданию онтологий, который бы привел к однозначно успешному результату не существует. Процесс создания онтологий обычно является итеративным, т.е. сначала создается черновой набросок, а затем по мере необходимости происходит возврат для определения деталей, и так продолжается до тех пор, пока онтология не будет отражать концепцию предметной области с определенной степенью. Практически, создание онтологий включает: 1. Определение классов в онтологии, 2. Организация классов в некоторую иерархию (базовый класс → подкласс), 3. Определение слотов и их допустимых значений, 4. Заполнение значений слотов для экземпляров классов.

Как определить, правильно ли создана онтология?

Для любой предметной области может существовать бесчисленное количество онтологий; ведь каждая новая онтология – это всего лишь еще один из способов структурирования концепций и отношений между ними. Однако существуют несколько простых принципов, которые могут помочь при принятии решений о том, как создавать те или иные онтологии:

- Не может быть только одного способа описания модели предметной области – всегда есть жизнеспособная альтернатива. Лучшее решение почти всегда будет зависеть от того, какая система разрабатывается и от возможных будущих изменений в системе.
- Процесс разработки обязательно должен быть итеративным.
- Концепции в онтологии должны быть максимально близки к объектам (логическим или физическим) и отношениям между ними в интересующей области знаний. При правильном моделировании, онтология может быть представлена предложениями, где вероятней всего в качестве существительных будут объекты, а отношений – глаголы.

С чего начать?

Для начала необходимо понять, для чего должна быть использована онтология и как примерно выглядел бы ее детальный и

общий вариант. Затем среди многих альтернатив вы должны будете выбрать ту, которая будет лучше всего работать для намеченной задачи, а также будет наиболее интуитивна, расширяема, поддерживаема. При этом необходимо помнить, что онтология представляет предметную область в реальном окружающем мире, и потому понятия в онтологии также должны отражать эту реальность. После того как вы сделаете черновой вариант онтологии, вы можете проверить ее и откорректировать, используя Protégé, или путем обсуждения с экспертами в исследуемой области (как результат, почти наверняка придется пересмотреть черновой вариант). Такой процесс итеративной коррекции, вероятней всего, будет продолжаться на протяжении всего жизненного цикла онтологии.

Разработка простейшей системы Основные положения

Предположим, что мы хотим разработать систему, которая помогает управлять стоимостью и организацией печатного издания (для простоты можно взять некую газету). Система должна отвечать на следующие вопросы: • Кто ответственный за каждый раздел в газете? • Каково содержимое каждой статьи в разделе и кто автор? • Перед кем отчитывается каждый автор? • Каково расположение и расходы на каждую статью? После того как мы определились с идеей, мы можем расписать некоторые из важных положений системы. Сюда могут войти: основные концепции и их свойства, а также отношения между ними. Для начала мы можем просто определить термины, независимо от роли, которую они могут играть в онтологии. Итак, в любой газете есть разделы. Каждый раздел имеет содержимое, например, статьи, реклама и т.д. и ответственного редактора. У каждой статьи есть автор, который может быть как работником газеты, так и быть приглашенным со стороны. Для каждого автора, работающего в газете, мы хотим знать его имя и зарплату, а также перед кем он отчитывается. По мере определения понятий, мы неявно определяем рамки нашей онтологии, а именно, что мы должны будем включить в нашу модель, а что нет. К примеру,

при начальном рассмотрении термина «работник», мы, возможно, хотели бы включить в это понятие вахтера или водителя грузовика из службы доставки. Однако, подумав, мы могли осознать, что хотим чтобы наша онтология была сфокусирована на производственных затратах, связанных напрямую с тем что, как и где написано в газете. Таким образом, мы решаем не включать вахтера и т.п. в область рассмотрения. Получив достаточно полный список терминов, мы можем разделить эти понятия по категориям в зависимости от их функции в онтологии. Понятия (концепции/термины предметной области), являющиеся объектами, такие как статья или автор, будут представлены в виде классов. Свойства классов, такие как содержимое раздела или зарплата, могут быть представлены как слоты, а ограничения на свойства или отношения между классами как грани/аспекты (slotfacets). Определив основные понятия, теперь мы можем показать, как создавать и структурировать их, используя систему Protégé.

Создание проекта

Перед началом работы, вы должны создать новый проект в системе Protégé. Для этого:

1. Запустите Protégé. Если у вас уже открыт проект, просто сохранитесь и перезапустите программу. После того как программа запустилась, появляется диалог приветствия, предлагающий создать новый проект, открыть последний проект или посмотреть документацию.

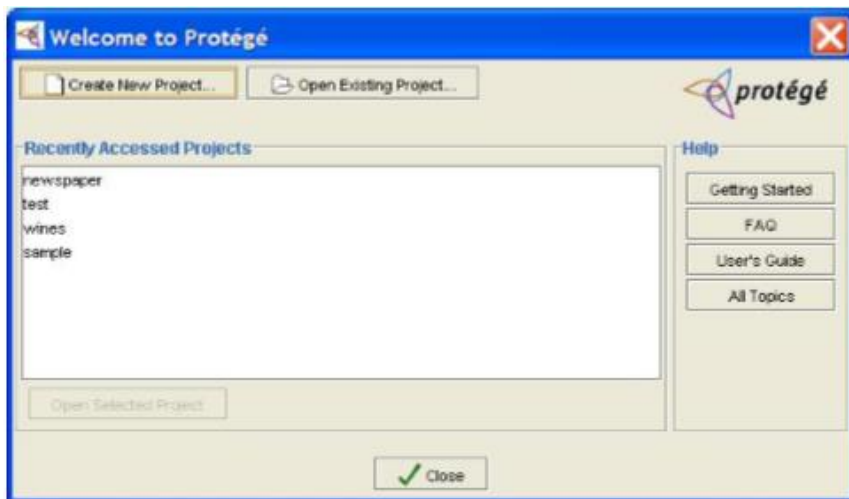


Рис.6. Окно приветствия

2. Щелкните мышкой по кнопке CreateNew

Project.... Появится диалоговое окно "CreateNewProject", позволяющее выбрать тип проекта. Если у вас нет необходимости в специальном формате для ваших файлов, просто нажмите кнопку Finish – будет выбран формат файла по умолчанию ProtegeFiles (.pontandpins).

3. Откроется окно проекта Protégé. Новый проект всегда открывается в области просмотра классов (Classesview). Видно, что в этой области на данный момент находятся только внутренние системные классы Protégé: THING и SYSTEM-CLASS. Никаких экземпляров классов создано к этому моменту не будет.

Сохранение проекта

Во время работы с программой вы можете захотеть сохранить промежуточные изменения, для этого:

1. Щелкните кнопку сохранить проект, вы также можете выбрать пункт Saveproject из меню файл File.

2. Для того чтобы указать место, куда будет сохранен Ваш проект, нажмите кнопку чуть выше самой верхней строчки (напротив надписи Project). В открывшемся диалоговом окне, перейдите по нужному пути в файловой системе (или создайте каталог, где будут храниться данные проекта и откройте созданную папку).

3. Введите имя файла проекта (например, "tutorial").

4. Нажмите кнопку Select.

5. Вы вернетесь в окно сохранения файлов проекта, нажмите ОК и проект будет сохранен.

Создание классов

Основное окно программы Protégé состоит из закладок (tabs) которые отображают различные аспекты модели знаний. Наиболее важной закладкой, когда вы только начинаете делать проект, является закладка классов (Classes). Обычно классы соответствуют объектам или типам объектов, в некой предметной области. В нашем примере с газетой, классы будут включать в себя людей, а именно, редакторов, репортеров, агентов по продаже, а также компоненты расположения информации газеты, такие как разделы, кроме того, содержимое газеты (реклама и статьи) будет также представлено в виде объектов. Классы в Protégé отображаются в виде иерархии наследования (inheritancehierarchy), которая располагается в области просмотра называемой ClassBrowser (или навигатор классов) в левой части закладки классов. Свойства классов выбранных в текущий момент в навигаторе, будут отображены в редакторе классов справа. Ниже вы узнаете, как создавать классы, подклассы, изменять иерархию классов, создавать абстрактные классы и добавлять дополнительные базовые классы к существующим классам.

Создание класса “Корреспондент”

Мы хотим знать происхождение каждой статьи, и потому мы начнем с задания типов сотрудников или служб, которые могут создавать статьи. Для начала создадим новый класс “корреспондент” (Columnist):

1. Выберите закладку классов.

2. Найдите область в навигаторе классов (ClassBrowser), где отображается иерархия классов (ClassHierarchy, в окне Protégé слева). Эта область отображает иерархию классов, с выделенным текущим выбранным классом.

3. Проверьте: по умолчанию класс :THING (вещь, нечто) должен быть выделен. Почти все классы в данном примере будут созданы на уровень ниже класса THING. Другой класс :SYSTEM_CLASS используется для определения структур различных форм Protégé.

4. Нажмите кнопку создать класс (CreateClass) в верхнем правом углу навигатора классов. Новый класс будет создан со стандартным именем (основанном на имени проекта). В нашем случае “tutorial_Class_0”. Вы можете увидеть, что имя нового класса в навигаторе классов после создания будет выделено, для указания того, что этот класс выбран в данный момент.

5. В активном поле редактора классов введите “Columnist”. В системе Protégé приняты правила наименования, когда первая буква в каждом слове в имени класса пишется в верхнем регистре, а остальные буквы в нижнем, при этом слова разделяются символом подчеркивания.

6. Нажмите ввод или щелкните мышью на отображаемый класс, чтобы подтвердить и отобразить свои изменения.

7. Если при изменении имени класса у вас возникли проблемы, посмотрите в панель редактора классов справа в главном окне Protégé. Стандартное имя нового класса должно быть отображено и

выделено в поле Name. Если правильное стандартное имя отображается, но не выделено, просто щелкните на поле Name мышкой, для того чтобы его отредактировать. Если имя неправильное, тогда скорей всего вы выбрали неверный класс в области отображения иерархии классов (ClassHierarchy), щелкните на нужном классе.

Создание класса “Автор”

Автором может быть любой возможный источник информации для статьи, такой как новостная служба или корреспондент. Для того чтобы создать класс автор:

1. Выделите класс :THING. Если вы этого не сделаете, то вы создадите класс, который будет подклассом класса Columnist.
2. Нажмите кнопку создать класс (Createclass) и наберите с клавиатуры имя класса: Author.
3. Нажмите ввод для завершения создания класса.

Создание подклассов класса “Автор”

Теперь мы хотим добавить больше источников статей (служба новостей и редактор), которые мы создадим как подклассы класса автор (Author).

1. Выберите класс “Author” в области отображения иерархии классов.
2. Нажмите кнопку “создать класс” (Createclass) и переименуйте новый класс в News_Service (служба новостей). Помните, что когда бы вы не создавали новый класс, он будет создан как подкласс текущего выбранного класса. Также заметьте, что когда вы создаете первый подкласс класса, то иконки или появляются слева от него. Вы можете использовать эти иконки для того чтобы показать или

скрыть подклассы класса. Продолжая разрабатывать онтологию, создадим еще один подкласс класса Автор (Author).

3. Выберите класс Автор (Author) в навигаторе классов (ClassBrowser).

4. Нажмите кнопку CreateClass и переименуйте созданный класс в Editor (редактор).

Изменение иерархии классов

На данной стадии разработки, можно заметить что “Автор” (Author) и “Корреспондент” (Columnist) находятся на одном уровне в создаваемой иерархии, в то время как “Служба новостей” (News_Service) и “Редактор” (Editor) являются подклассами класса Автор (Author). С точки зрения концепции и служба новостей, и редактор, и корреспондент, все могут являться авторами (источниками) статей, т.е. понятие “Автор” является общим для всех этих трех классов. Значит, текущее расположение в иерархии противоречит принципам хорошо спроектированных онтологий - все классы, имеющие одно и то же более общее понятие, должны находиться на одном уровне в иерархии. Таким образом, мы хотим модифицировать нашу иерархию, чтобы класс “Корреспондент” (Columnist) стал подклассом класса Автор (Author), правильно отражая структуру предметной области. Для этого необходимо сделать следующее:

1. Щелкните на классе Columnist (Корреспондент) и перетащите (drag-ndrop) его на класс Author (Автор).

2. Класс Columnist будет удален с предыдущего места в иерархии и создан в новом, но уже как подкласс Author. В данном случае, ошибка была введена явно, для примера. Однако, при создании собственных онтологий, в процессе разработки могут открываться отличия или сходства между классами, которые не были явно видны в начале и вам вероятней всего придется часто использовать


перемещение, создание, удаление классов, для того чтобы создавать иерархии, оптимально отражающие положение вещей в предметной области.

Создание абстрактных классов

В системе Protégé классы могут быть как конкретными (Concrete), на основе таких классов система может непосредственно создавать готовые экземпляры, так и абстрактными, у таких классов не может быть экземпляров. По умолчанию, при создании класса выбирается тип класса как “конкретный класс”. Так как в нашей системе понятие автора является скорее обобществляющим, нежели связанным с какой конкретной сущностью, мы делаем вывод что класс Автор (Author) не может сам по себе иметь экземпляров без более детального определения (к примеру, является ли автор службой новостей или корреспондентом). Поэтому мы решаем сделать класс Автор (Author) абстрактным.

1. Выберите класс “Author” в области иерархии классов (ClassHierarchy). Далее в редакторе классов (справа от навигатора классов), найдите меню Роль (Role), оно должно находиться прямо под именем класса.

2. Щелкните по списку в меню Role и выберите “Abstract”.

3. Заметьте, что когда вы меняете роль класса, иконка в перед именем класса меняется на . Такая иконка означает, что класс является абстрактным (соответственно означает конкретный класс).

Создание класса “Работник”

Теперь настало время создать класс “работник” (Employee). Этот класс подразумевает под собой любого работника газеты, независимо от того является ли он автором или нет. Заметим, однако, что нам интересны только те работники, которые как-то привлечены к созданию и управлению непосредственным содержанием газеты.

Здесь стоит еще раз отметить, что одним из выборов который делает разработчик при проектировании онтологий, является отсечение лишних сущностей. Хотя изначально может казаться, что многие классы являются важными, все же необходимо следить затем, чтобы создаваемая иерархия не становилась слишком сложной. Применительно к нашему случаю, это означает, что хотя технически вахтер и является служащим компании, мы не будем создавать такой подкласс.

1. Выберите класс :THING в навигаторе классов. Несмотря на то, что некоторые авторы являются служащими газеты, мы не хотим, чтобы класс “Employee” (работник) был подклассом класса Автор (Author).

2. Нажмите кнопку CreateClass и переименуйте вновь созданный класс в Employee.

Добавление дополнительного базового класса к существующему подклассу

Как было упомянуто выше, мы хотим чтобы “корреспондент” стал “работником”. Но поскольку мы уже создали такой класс, мы не хотим создавать его снова как подкласс класса “работник”. Вместо этого мы можем сделать так, чтобы существующий класс “корреспондент” (Columnist) стал подклассом “Employee” (работника). Для этого нужно сделать следующее:

1. Выберите класс Columnist (корреспондент) в навигаторе классов.

2. Найдите панель базовых классов (Superclasses) в нижней левой части окна системы Protégé (под навигатором классов). Заметьте, что когда выбран класс Columnist (корреспондент), его базовый класс (Автор) отображается в панели Superclasses.

3. Нажмите кнопку AddSuperclass (добавить базовый класс), в верхнем правом углу панели базовых классов (Superclasses). Появится диалоговое окно, отображающее все классы, созданные на тот момент времени, в виде иерархии.

4. Выберите класс “Employee” (работник) и нажмите ОК. Теперь класс “корреспондент” (Columnist) имеет два базовых класса (Автор и Работник). Оба класса показаны в панели базовых классов.

5. Заметьте, что рядом с именем класса “Employee” появилась иконка . Щелкните по ней, для того чтобы развернуть дерево подклассов и увидеть детей класса “работник” (Employee). Как видим, класс “корреспондент” теперь присутствует в двух местах в навигаторе классов: как подкласс класса Автор (Author) и еще раз как подкласс класса “Работник” (Employee).

Добавление базового класса с помощью перетаскивания (drag-androp)

Существует еще одна возможность задания базового класса – при помощи механизма перетаскивания (drag-n-drop):

1. Выберите класс Редактор (Editor) в навигаторе классов.

2. Удерживая нажатой левую кнопку мыши, перетащите класс Editor (редактор), так чтобы он находился над классом Employee (работник). Класс Employee автоматически будет выделен.

3. Перед тем как отпустить кнопку мыши, нажмите дополнительно клавишу Ctrl, затем отпустите кнопку мыши, для того чтобы перенести класс.

Для удаления базового класса от некоторого подкласса, выберите класс, который должен быть удален в панели отображения базовых классов (Superclasses), и нажмите кнопку удаления базового класса (RemoveSuperclass). Теперь вы готовы к тому чтобы добавить несколько атрибутов (свойств) к созданным классам. В следующей

секции будет показано, как это можно сделать с помощью механизма слотов (slots).

Создание слотов

Как вы могли убедиться выше, в системе Protégé под классами понимаются конкретные понятия (концепции) предметной области, такие как редактор или корреспондент. В то же время классы это больше чем объекты, объединенные в иерархию. Они также могут иметь атрибуты (свойства), к примеру, имя, номер телефона или уровень зарплаты и отношения между ними, такие как Автор Статьи. Атрибуты и отношения класса описываются конструкцией под названием слот. В данном разделе будет показано, как создавать слоты, привязывать слоты к классам, описывать отношения между классами, а также будет описан механизм наследования слотов.

Создание слота (используя закладку слоты (Slotstab))

Для создания слота есть несколько способов. Один из них – это создать слот используя закладку “Slots”, а затем связать его с одним или более классами. Вернемся к нашему примеру, для того, чтобы создать слот name, используя закладку Slots, необходимо:

1. Щелкнуть на закладку Slots. Заметьте, что расположение элементов управления на закладке слотов, схоже с закладкой классов, а именно, готовые слоты отображаются слева в области просмотра, а редактирование слотов возможно с помощью редактора (справа).

2. Нажмите кнопку создать слот (CreateSlot) в правом верхнем углу панели отображения иерархии слотов (SlotHierarchy). Будет создан новый слот. Также как и при создании класса, ему присваивается стандартное имя, в нашем случае “tutorial_Slot_0” (имя будет автоматически выделено, после создания слота).

3. Перед переименованием слота, убедитесь, что стандартное имя выделено в редакторе слота. Наберите новое имя слота (в нашем

случае name). Рекомендованные правила наименования, таковы, что имя слота должно быть написано в нижнем регистре, при этом разные слова разделяются подчеркиванием. Такое наименование (классы с большой буквы, слоты с маленькой) помогает отличить классы от слотов в созданной онтологии.

4. Заметьте, что слот имеет по умолчанию тип значения String (строка). Тип накладывает ограничения на то, какие значения может принимать слот. Строковый слот, к примеру, может принимать в качестве значений алфавитно-цифровые строки (включая пробелы). Для этого простого слота мы не будем менять никаких аспектов/граней (facets) в редакторе слотов.

Связывание слота с классом

Все, что мы пока сделали, это определили общий атрибут name (имя). Для того чтобы действительно задействовать его в нашей онтологии, мы должны привязать его к классу. К примеру, мы хотим, чтобы каждый из экземпляров подклассов класса Автор имел имя. Вернемся к закладке классов и откроем на редактирование класс Автор (Author). Любой из атрибутов, который вы создаете или связываете с классом, будет отображаться в редакторе классов, справа от навигатора классов. Мы уже использовали редактор классов для смены имени нового класса, а также для изменения роли класса Автор. Теперь мы будем использовать редактор классов для просмотра и именованя слотов. Для того чтобы связать слот name с классом:

1. Щелкните на закладке классов.

2. Выделите класс Автор в панели отображения иерархии классов (ClassHierarchy). Посмотрите на редактор классов (справа), в этой области отображается поле имя, роль класса, а также документация и ограничения (constraints). Под этими полями расположена, панель шаблонов слотов (Templateslots), которая занимает всю оставшуюся

нижнюю часть редактора классов. Эта область показывает слоты, связанные с классом. На текущий момент она пуста.

3. Для добавления слотов к классу, нажмите кнопку `AddSlot`. Кнопки управления слотами находятся в верхнем правом углу панели шаблонов слотов (`Templateslots`).

4. После того как вы нажмете кнопку, появится диалог выбора слота, в котором будет отображен список всех доступных слотов в вашем проекте (в алфавитном порядке, за исключением системных классов `Protégé`, которые будут видны в самом низу списка).

5. Выберите `name` и нажмите `ОК`.

Если вы посмотрите теперь на панель шаблонов слотов (`Templateslots`), то увидите, что слот `name` был добавлен в список, и вместе с ним отображаются его свойства, в нашем случае это мощность (количество элементов типа) и сам тип (строка, `String`).

Создание слота из закладки классов

Переключение между закладками классов и слотов может показаться утомительным. И так как слоты есть свойства класса, их можно создавать проще, непосредственно с закладки классов. Попробуем создать слот для класса `Employee`, для этого:

1. Выберите класс `Employee` в панели иерархии классов.

2. Нажмите кнопку `создать слот (CreateSlot)`, в правом углу панели шаблонов слотов, будет вызвано окно добавления слота:

3. Наберите `salary` (зарплата) в поле имя (`Name`), нажмите `ввод`. 4. Вернитесь в главное окно (при этом не обязательно закрывать окно редактирования, т.е. можно оставить его открытым и вернуться туда позже для редактирования свойств слота). Заметьте, что теперь новый слот показывается в панели шаблонов слотов, когда выбран класс “Работник” (`Employee`).

Слоты и наследование

Мы не должны добавлять слот `name` (имя) к любому классу, где мы хотим его видеть. В смысле того, что любой подкласс класса автоматически наследует все слоты базового класса. К примеру, если вы выберете класс Служба новостей (`News_Service`), то увидите, что:

- Слот `name` (имя) уже связан с этим классом через механизм наследования.

- При этом иконка для слота отличается от той, которая использовалась для класса Автор (`Author`), а именно, для наследованных слотов используется иконка.

Подклассы более чем с одним базовым классом наследуют слоты от всех базовых классов. К примеру, если Вы выберете класс `Editor` (редактор), то увидите, что он наследует слот `name` (имя) от Автора, и слот `salary` (зарплата) от Работника. Множественное наследование одна из основополагающих возможностей Protégé.

Создание аспектов/граней (facets) слота

Слоты, которые были созданы на предыдущем шаге, очень простые. Однако, слоты сами по себе, тоже могут иметь свойства. К примеру, зарплата всегда является числом. Вы также можете использовать слоты для задания отношений между классами. Свойства слота, называемые аспектами/гранями (facets), могут быть созданы, как на закладке классов (используя диалог спецификации слота), так и на закладке слотов (используя редактор слота).

Создание аспектов слота “зарплата”

Мы можем определить несколько аспектов для слота “зарплата”, который был создан ранее.

1. Выберите класс “работник” (`Employee`) в панели иерархии классов.

2. Щелкните два раза на слоте “зарплата” в панели шаблонов слотов (Templateslots), для того чтобы открыть форму выбора вида слота. Когда вы редактируете слот, Вы можете выбрать, будут ли изменения применяться к слоту и всем классам, связанным со слотом (вверх по иерархии до самого верхнего класса), или вы просто хотите чтобы изменения коснулись текущего класса и всех его детей.

3. В нашем случае, мы хотим просмотреть и отредактировать слот верхнего уровня. Потому убедитесь, что режим просмотра слотов верхнего уровня (Viewtop-levelslot) выбран и нажмите ОК. При этом изменение определения слота будет затрагивать всю онтологию.

4. В открывшейся форме редактирования слота, выберите Float из списка выбора типа значения (ValueType). Теперь при создании экземпляров, можно будет вводить для этого слота только правильные значения в формате с плавающей запятой.

5. Введите 0 (ноль) в поле Minimum (минимальное значение). Таким образом, мы можем быть уверены, что теперь любое значение для поля “зарплата” будет не отрицательным.

6. Закройте, диалог редактирования слота, и Вы сможете увидеть, что описание слота в панели шаблонов слотов изменилось. В колонке тип теперь указан Float а минимальное значение = 0 появилось в колонке Otherfacets (другие аспекты).

Создание отношения между классами

Система Protégé также позволяет вам создавать слоты, которые могут быть использованы для описания отношений между классами, которые не определены в иерархии классов. Для этого существуют слоты Instance (экземпляр) или Class (класс). К примеру, “Редактор” (Editor) может быть ответственным за одного или более работников. Мы можем создать новый слот, который бы описывал связь между “Редактором” и “Работником”:

1. Выберите класс “Редактор” (Editor) в навигаторе классов.
2. Нажмите кнопку CreateSlot для того чтобы создать и связать новый слот с классом “Редактор” (Editor).
3. В открывшейся форме редактирования, наберите в поле имя (Name) responsible_for (ответственный за).
4. Выберите Instance (экземпляр) из списка типов значений (ValueType).
5. Нажмите кнопку AddClass (справа сверху на панели Allowedclasses) . Появится окно выбора классов, где будут показаны все классы проекта. Выберите класс “Работник” (Employee) и нажмите ОК.
6. Чтобы разрешить редактору, быть ответственным более чем за одного сотрудника, поставьте галочку в пункте multiple, в панели мощности (cardinality).

После завершения шагов 1..6, слот форма для responsible_for будет выглядеть следующим образом:

Что же мы создали? Мы создали слот, который может содержать один или более экземпляров класса “работник” в качестве значения. Позднее, когда мы будем создавать экземпляры класса “Редактор” и захотим указать, за каких работников он несет ответственность, мы сможем выбрать один или более экземпляров класса “работник”, чтобы заполнить responsible_for слот.

Создание экземпляров классов

Экземпляры классов – это и есть собственно данные вашей базы знаний. Вообще, хорошим правилом, перед вводом конечных данных, является окончательная проверка структуры проекта, потому что когда данные будут введены, необходимость изменения структур проекта может повлечь за собой потерю уже введенной информации.

Кроме того, при добавлении новых слотов, необходимо заполнять их значения для старых экземпляров классов. В этой секции, мы создадим два экземпляра класса редактор:

1. Перейдите на закладку экземпляров (instances). Закладка имеет три панели. Первая, слева, отображает иерархию классов. Средняя панель, которая сейчас пуста, показывает список экземпляров, созданных для конкретного класса. Третья панель показывает редактор экземпляра класса, где вы можете ввести значения слотов текущего выбранного класса.

2. Раскройте список подклассов класса “работник” (Employee).

3. Выберите класс редактор (Editor). Кнопка CreateInstance станет активной, означая, что можно создать экземпляр класса.

4. Нажмите кнопку CreateInstance . Экземпляр создан и появилась форма редактора экземпляра. Видно, что на ней много полей, по одному полю для каждого созданного слота. Используйте эти поля, для того чтобы заполнить слоты значениями. Заметьте, что отображение для класса Редактор (Editor) в панели иерархии классов (ClassHierarchy) изменилось после того, как был создан новый экземпляр класса. Единица в скобках означает, что этот класс имеет один экземпляр.

5. Введите ChiefHoncho в поле Имя (Name).

6. Введите 15000 в поле зарплата (salary). Заметьте, что символы в этом поле будут подсвечены красным цветом, если что-то другое, нежели число в формате с плавающей запятой будет введено. В системе Protégé, при попытке ввода значений, которые не удовлетворяют ограничениям слота, значения подсвечиваются красным цветом.

Теперь закладка экземпляров выглядит следующим образом (заметим, что экземпляр в навигаторе экземпляров (InstanceBrowser)

все еще имеет стандартное имя, такое как “tutorial_instance_0”). Как изменить имя будет показано в следующем разделе.

Создадим еще один экземпляр класса Редактор (Editor):

1. Нажмите кнопку CreateInstance в навигаторе экземпляров (InstanceBrowser).
2. Наберите Mr. Science в поле имя (name).
3. Введите 60000 в поле зарплата (salary).

Теперь, так как вы создали более чем один экземпляр класса, вы можете определить отношения (связи) между ними, к примеру, вы могли бы сказать, что “ChiefHoncho” будет ответственным за работу “Mr. Science”. Перед тем, как это сделать, для того чтобы работа с экземплярами была легче, необходимо указать слот отображения для класса “Редактор” (Editor). Система Protégé будет показывать значение слота отображения, каждый раз при выводе на экран экземпляра класса. О том, как это сделать, будет рассказано в следующем разделе.

Установка слота отображения

Для каждого класса в вашей онтологии, вы можете указать, что один из его слотов будет слотом отображения. Система Protégé будет показывать значение этого слота, при каждом выводе экземпляра класса на экран. Если слот отображения не будет указан, то будет выведено стандартное имя, сгенерированное системой (например, “tutorial_instance_0”). Обычно очень полезно устанавливать слот отображения для классов, которые будут иметь экземпляры. На самом деле, вы можете выбрать слот отображения даже до того, как будут созданы экземпляры класса. Для того чтобы указать слот отображения для класса «Редактор» (Editor).

1. Выберите закладку экземпляров (Instances).

2. Выберите класс «Редактор» в панели иерархии классов.
3. Нажмите кнопку, меню экземпляров (стрелочка вниз), в верхней правой части навигатора экземпляров.
4. Выберите пункт задать слот отображения (setdisplayslot).
5. Выберите поле имя (name) из списка.
6. Вид списка экземпляров, в навигаторе экземпляров, поменяется, чтобы показать новые значения слота отображения. Экземпляры класса “Редактор” (Editor) теперь будут перечислены, как значения слота имя (name). Начиная с этого момента, вы можете перебирать экземпляры класса “редактор” по его имени везде, где будет появляться список экземпляров классов.

Создание отношений (связей) между экземплярами классов

В этом разделе, мы модифицируем экземпляр ChiefHoncho и сделаем так, чтобы он стал ответственным за экземпляр Mr. Science:

1. Перейдите на закладку экземпляров (instances), разверните класс “работник” (Employee) в панели иерархии классов (ClassHierarchy) и выберите класс “редактор” (Editor). Экземпляры редактора теперь показаны в навигаторе экземпляров (InstanceBrowser).

2. Выберите ChiefHoncho в навигаторе экземпляров. Слоты для ChiefHoncho будут показаны в редакторе экземпляров, включая слот responsible_for (ответственный за). Заметьте, что система Protégé использует имена слотов в форме редактора, но автоматически заменяет подчеркивания в пробелы и переводит в верхний регистр первую букву каждого слова.

3. Нажмите кнопку AddInstance , справа сверху рядом с полем ResponsibleFor.

4. Откроется окно диалога с двумя панелями. Слева будет показана иерархия доступных классов для слота `responsible_for`.

5. Выберите класс `Editor` (редактор). Справа будут показаны все экземпляры класса. Выберите `Mr. Science` и нажмите ОК.

6. Вы только что создали отношение (связь) в своей онтологии, которая гласит, что работник `ChiefHoncho` является ответственным за работника `Mr. Science`.

Настройка формы ввода

Для каждого класса в вашей онтологии, `Protégé` генерирует форму по умолчанию, которую вы можете использовать для ввода данных экземпляра. Формы содержат поля ввода данных, или “виджеты” для каждого слота связанного с классом. Для разных типов данных слотов существуют разные типы “виджетов”, например, `Protégé` использует текстовый “виджет” (`TextFieldWidget`) для слотов с типом данных строка, целочисленный “виджет” (`IntegerFieldWidget`) для полей, у которых значение представлено как целое число, “виджет” список экземпляров (`InstanceListWidget`) для слотов, у которых в качестве типа установлен экземпляр класса и при этом мощность (количество элементов) больше одного и т.д. Если вам не подходит стандартная форма, созданная `Protégé`, вы можете изменить ее с помощью закладки форм (`Forms`). Среди других возможностей, вы можете изменить размер “виджетов”, перемещать их по форме, скрывать и даже менять тип “виджета”. Для того чтобы понаблюдать, как изменения, которые были сделаны на закладке форм, отображаются в редакторе экземпляров, перейдите на закладку экземпляров, и два раза щелкните по `ChiefHoncho` в навигаторе экземпляров, чтобы появилось отдельное окно редактора. Заметьте, что если вы создавали слоты для класса `Editor` в другом порядке, чем было описано в данном руководстве, ваша форма может выглядеть отлично от картинок в следующих секциях.

Изменение размера “виджета”

Вы можете изменить размер выбранного “виджета”, растянув его за угол или границу, для этого:

1. Перейдите на закладку **Формы**.

2. Удостоверьтесь, что именно класс “редактор” (Editor) выбран в навигаторе форм (FormBrowser) слева. Затем, выберите FloatFieldWidget для слота salary (зарплата), щелкнув на нем в редакторе формы справа. В этом случае он будет подсвечен зеленым. Заметьте, что выбранный тип виджета в списке “SelectedWidgetType” справа сверху указывает на то, что это “виджет” используется для ввода элементов с плавающей запятой (FloatFieldWidget).

3. Щелкните по правой границе “виджета”, и удерживая кнопку мыши нажатой, перетащите ее, чтобы изменить размер “виджета”. Попытайтесь выровнять правую границу “виджета”, так чтобы она совпадала с правой границей “виджета” имя (name).

4. Заметьте, что иконка перед формой класса Editor (редактор) в навигаторе форм изменилась. Новая иконка указывает, форма этого класса была изменена, и больше не является стандартной.

Перемещение “виджета”

Вы можете изменить положение “виджета” на форме, также при помощи перетаскивания:

1. Выберите InstanceListWidget (“виджет” для редактора экземпляров) для слота responsible_for (ответственный за).

2. Перетащите его на правый верхний угол формы, так чтобы верхняя граница “виджета” совпадала с верхней границей “виджета” слота “имя” (name).

Изменение кнопок “виджета”

Вы можете настроить “виджет” так, чтобы он показывал другую надпись или другой список кнопок, нежели по умолчанию. К примеру, вы хотите удалять экземпляр из вашего проекта, просто нажав кнопку в InstanceListWidget (“виджете” экземпляров) для слота responsible_for. Для того чтобы показать кнопку удаления сделайте следующее:

1. Два раза щелкните на InstanceListWidget (“виджете” экземпляров) с именем “ResponsibleFor” в редакторе форм.
2. Перейдите на закладку кнопки.
3. Установите галочку для пункта ShowDeleteInstanceButton (показывать кнопку удаления).
4. Нажмите ОК. “Виджет” для слота responsible_for (ответственный за)

Скрытие “виджета”

Вы можете скрыть “виджет” так, чтобы он не был виден на форме и в редакторе экземпляров (при этом информация из онтологии не удаляется). Например, вы хотите скрыть “виджет” для слота salary (зарплата). Для этого:

1. Выберите FloatFieldWidget (виджет для поля с плавающей запятой) под названием Salary в редакторе формы.
2. Выберите "<none>" из списка выбора типа “виджета” (SelectedWidgetType).
3. “Виджет” для слота salary (зарплата) больше не виден в редакторе формы.

Отображение скрытого “виджета”

Для того чтобы показать виджет, который был скрыт на предыдущем шаге, сделайте следующее:

1. В навигаторе форм, нажмите кнопку ViewFormCustomizations (посмотреть изменения формы) , сверху справа.

2. В диалоге конфигурации формы, вы сможете увидеть список слотов и соответствующих им “виджетов”.

3. Щелкните по "<none>" справа от поля salary (зарплата), а затем выберите FloatFieldWidget (“виджет” для полей с плавающей точкой).

4. Нажмите ОК.

5. “Виджет” будет снова виден в редакторе формы.

Использование расположения по умолчанию

Если вы хотите отменить все свои изменения, то система Protégé позволяет вернуть все «виджеты» на форме в стандартное положение, в соответствии с их текущим размером. Для того чтобы расположить “виджеты” стандартным образом:

1. Выберите форму класса редактор (Editor) в навигаторе форм.

2. Нажмите кнопку RemoveFormCustomizations (удалить изменения формы) в верхнем правом углу навигатора форм.

3. Форма будет выровнена по стандартам Protégé. Все скрытые виджеты будут показаны снова. Иконка, информирующая о том, что форма была изменена, будет заменена на .

Создание и сохранение запросов

Закладка запросов позволяет вам писать получать сведения из вашего проекта по всем экземплярам классов, которые удовлетворяют интересующим критериям. Для того чтобы создать запрос, вы должны выбрать один или более классов и один или более слотов в классе. Вы можете также сохранить запросы в библиотеку для последующего использования.

Создание запроса

Предположим, что вам интересно найти всех работников, которые имеют зарплату больше чем 75000\$ в год. Для этого создадим запрос:

1. Щелкните на закладке запросов.
2. Щелкните на кнопке SelectClasses (выбрать классы) чуть повыше текстового поля Class (класс) в панели запросов.
3. Выберите класс “работник” (Employee) из панели выбора классов, затем нажмите ОК.
4. Теперь класс Employee (работник) отображается в текстовом окне Class (класс).
5. Нажмите кнопку выбора слота (SelectSlot) чуть выше текстового поля Slot (слот).
6. Выберите salary (зарплата) в диалоговом окне выбора слота и нажмите ОК.
7. Меню справа от поля Slot (слот) теперь активно, и текстовое поле в правом дальнем углу напоминает, что тип значение выбранного слота число с плавающей запятой.
8. Выберите “isgreaterthan” (больше чем) в выпадающем списке. Затем, введите “75000” в поля ввода под надписью Float.

Запуск запроса

Теперь, когда вы создали запрос, вы можете запустить его и посмотреть результаты.

1. Для запуска, нажмите кнопку поиска (Find) в нижнем правом углу панели запросов.
2. Результаты будут показаны в панели отображения результатов поиска справа. Если вы не можете видеть результаты полностью, то можно расширить окно или подвинуть разделитель полей.

Сохранение запроса

Вы можете сохранить запрос перед или после того как он будет запущен. Для того чтобы сохранить запрос в библиотеку запросов, сделайте следующее:

1. Нажмите кнопку добавить запрос в библиотеку (AddtoQueryLibrary) справа от поля queryname (имя запроса).
2. Наберите "sample_query" в окне ввода имени запроса.
3. Нажмите ОК. Имя будет показано в поле queryname (имя запроса) и запрос будет отображен в панели библиотеки запросов (QueryLibrary).

Загрузка запроса

Для того чтобы загрузить сохраненный запрос, вы должны выбрать его в библиотеке запросов.

1. Для начала, так как мы запускаем тот же запрос снова, нажмите кнопку очистки панели результатов Clear. Иначе не будет заметно изменений.
2. Выберите sample_query в библиотеке запросов внизу экрана.
3. Нажмите кнопку retrievequery .
4. Сохраненный запрос теперь отображается в верхней части окна (при желании можно изменить его). Вы также можете объединить запросы, нажав кнопку More.
5. Для запуска запроса нажмите кнопку Find (поиск).

Контрольные вопросы

1. Как представить онтологию в Protégé?

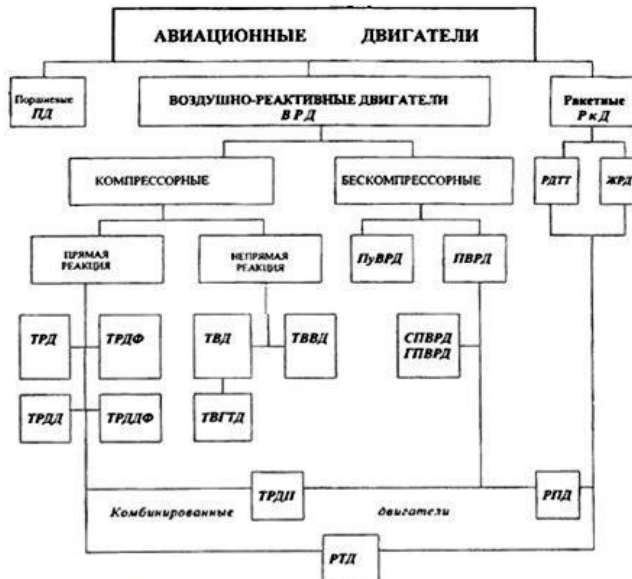
2. Объясните назначение редактора классов для представления знаний.
3. В какие форматы можно передать онтологию, созданную в Protégé?
4. Какие основные задачи можно решать с помощью Protégé?
5. Назовите и охарактеризуйте основные функциональные возможности Protégé.
6. Какие действия необходимо последовательно выполнить при построении онтологии в системе Protégé?
7. Что такое слоты, какие атрибуты слотов имеются в Protégé?
8. Какие типы классов реализованы в Protégé?
9. Есть ли возможность изменения ранее введенных исходных данных, добавления в онтологию значений новых, ранее не известных параметров?

Лабораторная работа 3

Онтологический инжиниринг знаний в системе FluentEditor

FluentEditor – онтологический редактор польской компании Cognitum для редактирования сложных онтологий, при создании которых используется контролируемый язык (от англ. ControlledNaturalLanguage – CNL).

Рассмотрим работу в системе FluentEditor на примере создания базы данных авиационных двигателей. Под базой данных (БД) будем понимать коллекцию согласованных взаимосвязанных данных. БД используются для описания некоторой предметной области с целью хранения, обработки и доступа к необходимой информации о ней. Базы данных содержат (и способны обрабатывать) большие массивы относительно простой информации (при этом доступ возможен



ТОЛЬКО К ЭТИМ ЯВНО ВВЕДЕННЫМ ДАННЫМ).

Рис. 7. Классификация авиационных двигателей

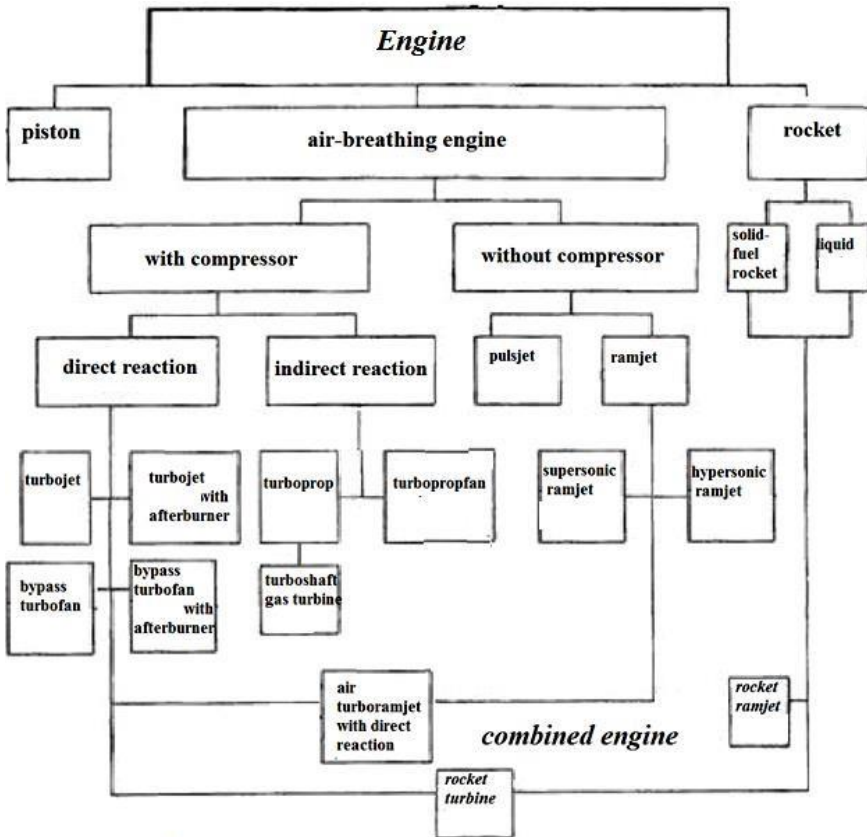


Рис. 8. Classification of aircraft engine

Подбор двигателя является важным этапом при проектировании самолета, так как инженеру-проектировщику необходимо знать основные параметры двигателя (массу, мощность, расход топлива) и на их основе осуществлять выбор силовой установки.

Цель работы: Освоить работу с онтологическим редактором FluentEditor, получить навыки построения онтологий.

Задачи:

1. Создать онтологию классификации авиационных двигателей.
2. Наполнить онтологию экземплярами авиационных двигателей.
3. Осуществить выбор наиболее подходящего двигателя по параметрам массы, мощности, тяги и удельного расхода топлива.

Создание проекта

Рассматриваемый онтологический редактор находится в свободном для скачивания доступе в сети Интернет. Пройдя по ссылке [11] необходимо нажать на «Downloadnow» («Скачать сейчас»), ввести свой e-mail адрес, имя, организацию (Samara University) в выплывшем окне регистрации и нажать на «Start». После чего начнется скачивание дистрибутива FluentEditor на компьютер. Когда дистрибутив будет полностью скачен на компьютер, запустите его и следуйте инструкциям, выдаваемым программой во время установки.

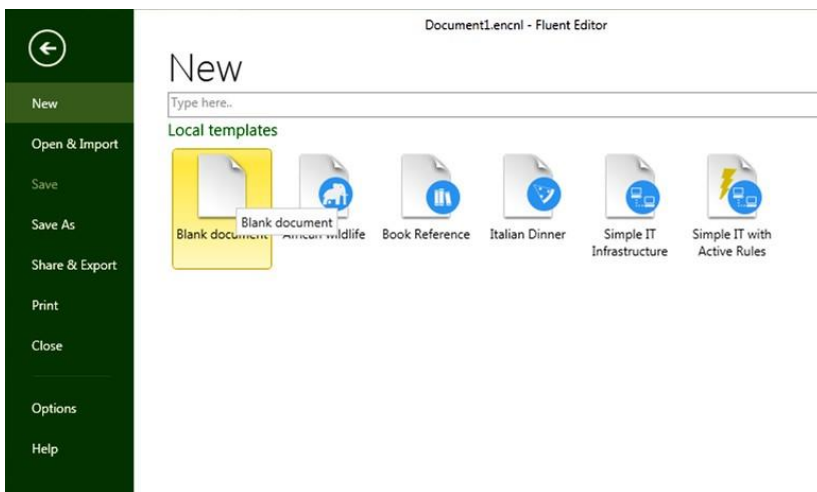


Рис. 9. Создание проекта

Завершив установку онтологического редактора, для запуска программы нажмите «Пуск» -> «Все программы» -> «OntorionFluentEditor».

Чтобы создать новый проект, выберите «File» на панели быстрого доступа, а затем нажмите кнопку «New». Тогда представится возможность либо создать новый проект «Blankdocument», либо открыть уже существующий, который послужит примером структуры проекта. Для создания собственной онтологии выберите «Blankdocument» (рис. 9).

Основные части рабочего окнаредатора

В правой части рабочего окна редактора находится вкладка отображения дерева таксономии «TaxonomyTree» (рис. 10). Таксономия (от др. – греч. τάξις – строй, порядок и νόμος – закон) – учение о принципах и практике классификации и систематизации [9]. Дерево таксономии отображается для каждого файла OWL, а строится и редактируется на основе данных из этого файла [1].

В центре рабочего окна программы находится поле для создания онтологии – окно редактора CNL. Окно CNL редактора – основная часть FluentEditor, в которой формируются, просматриваются и редактируются файлы онтологий.

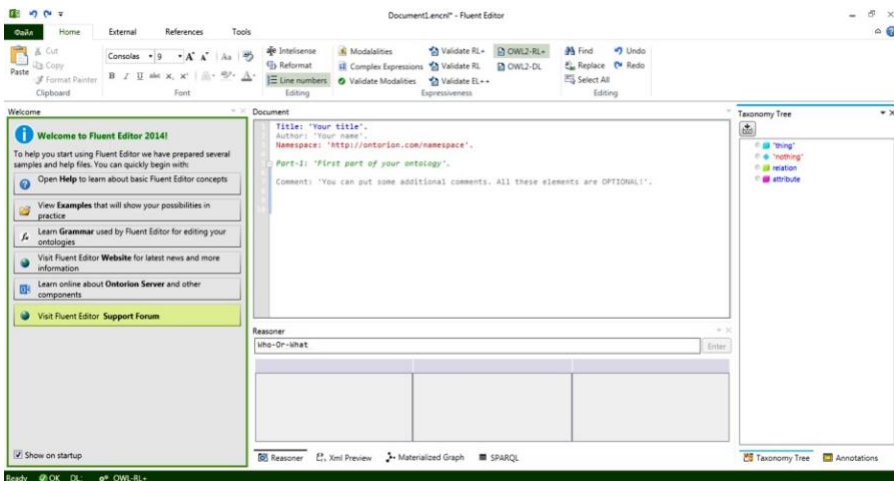


Рис. 10. Рабочее окно редактора

В левой части рабочего окна редактора разработчики FluentEditor предлагают пользователю окно помощи, где можно изучить грамматику редактора, его функциональные составляющие, а также посмотреть примеры уже построенных онтологий или посетить интернет-форум редактора.

Сохранение проекта

Для сохранения проекта откройте в панели быстрого доступа вкладку «File» -> «Saveas» или наберите сочетание клавиш Ctrl + S.

Окно Document

Document находится в окне редактора CNL. После создания проекта перед пользователем открывается рабочее окно программы, посередине которого и находится окно Document, уже содержащее в себе несколько полей: Title (Заголовок), Author (Автор), Part-1: 'Firstpartofyourontology' (Часть-1: 'Первая часть вашей онтологии'), Comment (Комментарий).

После слова «Title» вводится название/заголовок онтологии, в строке «Author» записываются авторы, создавшие онтологию. Название и авторов онтологии необходимо указывать в кавычках, а в конце ставить точку. Например, «Author: 'Ivanov I.I.'»

Написание онтологии на языке CNL начинается после слов «Part-1: 'Firstpartofyourontology'».

Создание классов и подклассов

Когда вы начинаете вводить текст онтологии в Document, FluentEditor автоматически помогает вам в нескольких направлениях:

Подсказка «Вставка» дает список последующих слов, которые могут быть написаны после введенной фразы. Можно либо ввести слово вручную, либо выбрать его из всплывающего списка. Ключевые слова редактора помечены синим шрифтом, а

пользовательские слова отмечены черным. Чтобы открыть окно подсказки CNL, нажмите кнопку Intelisense или поставьте курсор в Part-1 и воспользуйтесь сочетанием клавиш Ctrl + Spacebar.

- Синтаксические ошибки будут отмечены **красным** подчеркиванием. Если ввести фразу, которая неверна в соответствии с грамматикой FluentEditor, то эта фраза выделится красным подчеркиванием.

Рассмотрим построение онтологии в редакторе FluentEditor на примере пассажирского самолета Ан-28 с газотурбинным двигателем ТВД10-Б. Параметры двигателя: масса (кг) – 300, мощность (кВт) – 754.

Создадим 3 класса:

- двигатель(engine),
- самолет(plane),
- характеристики двигателя (enginecharacteristics).

Класс создается при задании класса как вещи (от англ. thing) или, когда классу присваивается подкласс или экземпляр. В первом случае необходимо записать «Everyengineis a thing.», во втором – для присвоения классу подкласса необходимо поставить название класса и подкласса в предложении на CNL в качестве дополнения и подлежащего соответственно. Например, присвоим классу «engine» подкласс «turboprop». Запишем «Everyturbopropisanengine.» («Каждый турбовинтовой двигатель – это двигатель»). В дереве таксономии появились данные объекты в описанной иерархии (рис. 11). По аналогии присвойте классу «engine» подклассы «plane» и «enginecharacteristic».



Рис.11. Отражение присвоения классу «engine» подкласса «turboprop» в дереве таксономии

Следует отметить, что редактор не воспринимает пробелы между словосочетаниями в имени класса, поэтому если вы хотите ввести имя класса или подкласса, состоящего из нескольких слов, разделите их дефисом. Также не следует использовать заглавные буквы. Например, класс «характеристики двигателя» на CNL запишется как «engine-characteristics».

Подобным образом к созданным классам «plane» («самолет»), которому присваиваем тип самолета, например, подкласс «passenger» («пассажирский»), к классу «enginecharacteristics» («характеристики двигателя») подклассы «weight» («вес»), «power» («мощность»),

«thrust» («тяга»), «specific fuel consumption» («удельный расход топлива»).

Классы и подклассы в дереве таксономии обозначаются голубыми ромбами (◆), как показано на рис. 12.

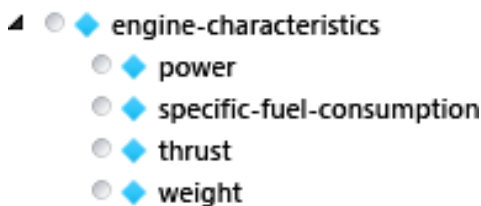


Рис. 12. Обозначение классов и подклассов в дереве таксономии

Создание экземпляров

Для создания экземпляра, также следует описать его зависимость от класса/подкласса, подобно тому, как создавали зависимость подкласса от класса. В отличие от класса, экземпляры описываются

всегда с заглавной буквы. Таким образом, на примере, приведенном ранее, присваиваем подклассу «turboprop» экземпляр «Tvd 10 B». Описываем зависимость в окне Document так «Tvd 10 B is a turboprop.». Таким же образом присваиваем подклассу «passenger» («пассажирский») экземпляр «An 28». При задании имени экземпляра из нескольких слов следует разделять их дефисом, а также каждое слово в имени экземпляра необходимо начинать с заглавной буквы. Например, «Tvd-10-B».

В дереве таксономии экземпляры, присвоенные классам, обозначаются зеленым кругом (●), как показано на рис. 13.

В результате проделанных операций получаем предложения на языке CNL и построенную иерархию в дереве таксономии (рис. 13).

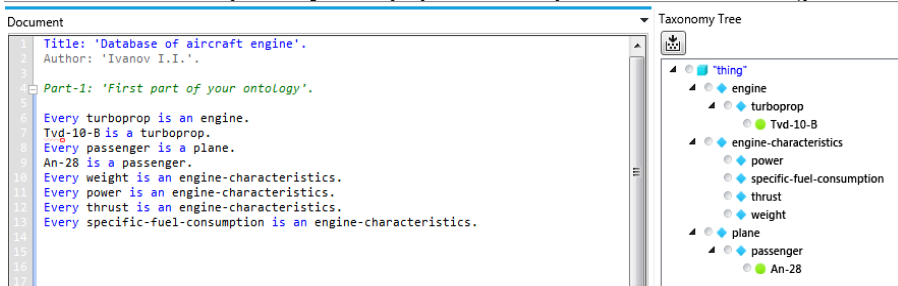


Рис. 13. Создание классов, подклассов, экземпляров

Создание отношений

В дереве таксономии в пункте «relation», обозначенном зеленым кубом (■) (рис. 14) отражаются все созданные отношения между объектами. Для создания отношений между объектами следует описать их связь на языке CNL, заменяя пробелы между словами в словосочетании на дефис. Например, если нужно указать, что двигатель ТВД10-Б установлен на самолете Ан-28, следует написать «Tvd 10 B used on An 28.». В дереве таксономии в пункте «relation» появилось заданное отношение между объектами (рис. 14).

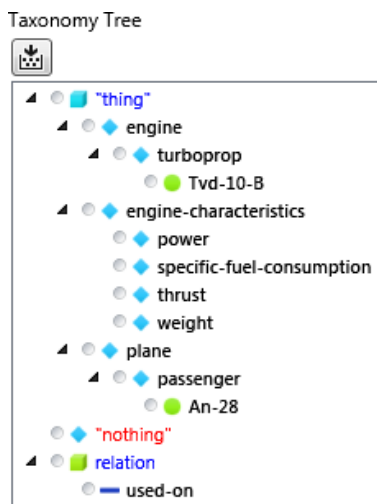


Рис. 14. Отражение созданных отношений в дереве таксономии

Присвоение свойств

Созданные свойства отражаются в дереве таксономии подпунктом «attribute», обозначенном розовым кубом « attribute » (см. рис.1.9). Чтобы присвоить объекту какое-либо свойство следует сделать тоже самое, что и при создании отношений. Программа сама определит, что относится к отношениям, а что к свойствам объектов (все свойства объектов начинаются с глагола иметь «have»). При создании свойств следует учитывать, что свойства, состоящие из нескольких слов, разделяются дефисом и пишутся с маленькой буквы. Введя предложение «Tvd 10B hasweightequalto 300.» двигателю добавилось такое свойство как масса. Логическое выражение «equalto» присваивает весу двигателя определенное значение. В нашем случае, фраза означает, что масса ТВД10-Б равна 300 кг. Присваиваем двигателю ТВД-10-Б мощность 754 кВт: «Tvd 10 B haspowerequalto 754». Описанные свойства двигателя, такие как «haveweight» и «havepower» появились справа, в окне дереве таксономии в пункте «attribute» (рис. 15).

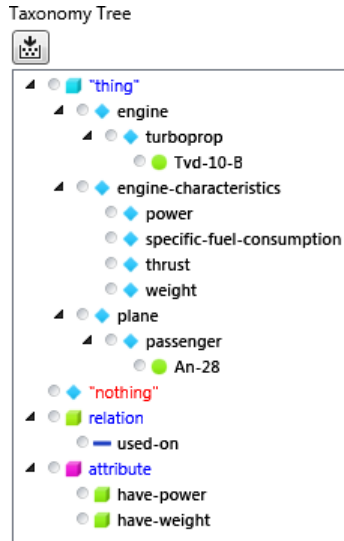



Рис. 15. Создание отношений и свойств, отражение их в дереве таксономии

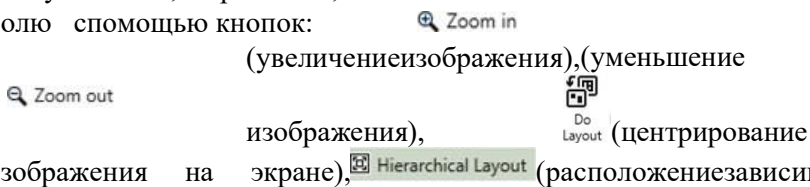
Проверка грамматики CNL

После написания онтологии, можно переформатировать выделенные строки, нажав на кнопку «Reformat» ( Reformat), находящуюся во вкладке «Home» на панели быстрого доступа, или Ctrl + U после выбора предложения. Переформатирование выполняет синтаксический анализ и преобразования CNL-предложений, что приводит к исправлению редактором предложений в соответствии с английской грамматикой. Например, если вместо «Tvd 10 B is a gasturbine.» Вы напишете «Tvd 10 B isgasturbine.», то с точки зрения английской грамматики данное предложение неправильно. Для исправления этой ошибки нажмите кнопку «Reformat» или Ctrl + U и программа проанализирует написанную онтологию и исправит ошибки. В результате получаем «Tvd 10 B is a gasturbine.»

CNL-диаграмма

FluentEditor 2015 предоставляет пользователю несколько возможностей представить созданную онтологию в графическом виде. Один из них построение диаграммы CNL, которая отражает зависимости классов и подклассов, экземпляров, а также отношения между ними в виде диаграммы. Для использования данной функции редактора на панели быстрого доступа выберете вкладку «Tools», далее нажмите «CNL Diagram». Рядом с вкладкой «Document» в окне редактора CNL появилась еще одна вкладка «DocumentDiagram», где сформировалась иерархическая схема описанной онтологии (рис. 16).

После открытия вкладки «Tools» и выбора «CNL Diagram», рядом с вкладкой «Tools» образовалась вкладка «Diagram», содержащая в себе инструменты редактирования диаграммы, в ней можно отправить схему на печать, сохранить ее, а также изменить положение объектов по полю с помощью кнопок:



Изменять расположение объектов можно и вручную, передвигая объекты с помощью нажатия и удержания их левой кнопкой мыши. Диаграмму можно сохранять в форматах *.jpg и *.svg, нажав на кнопки «Save as image» и «Save as vector» соответственно.

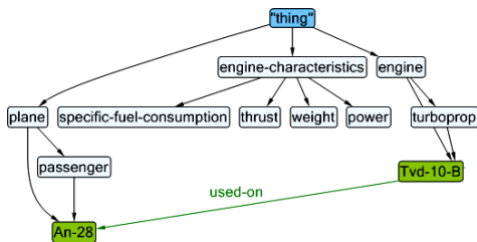


Рис.16. Результат построения онтологии в виде «CNL Diagram»

Reasoner

В FluentEditor встроен механизм, с помощью которого задаются вопросы к онтологии – Reasoner (от англ. «мыслитель»). Интерфейс Reasoner представлен на рис. 17. В верхней строке задаются вопросы, обязательно начинающиеся со слов «WhoOrWhat». Далее следует написать вопрос, учитывая грамматику CNL. Для начинающего пользователя редактор предлагает помощь в написании вопроса к онтологии. Необходимо нажать сочетание клавиш Ctrl+Spacebar, выплывет окно подсказок, т.е. слов, которые вы можете использовать на данном месте, учитывая написанную Вами онтологию и синтаксис CNL. В соответствии с правилами большинства языков после вопросительного предложения следует поставить знак «?»». Ответы на вопрос выводятся в трех колонках под строкой, в которой задаются вопросы, в зависимости от того, какие отношения существуют между вопросом и ответом на него. Три колонки отражают соотношение вопроса и ответа как класса с классом или как экземпляра с экземпляром, или как класса с экземпляром.

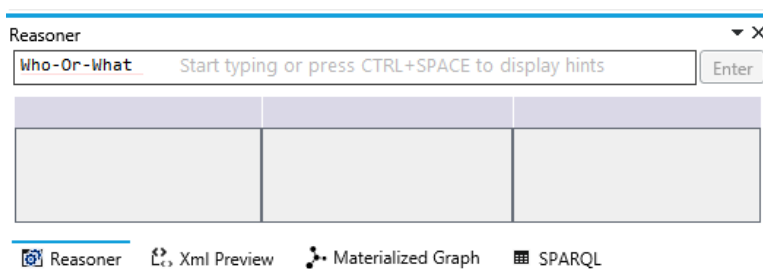



Рис. 17.ОкноReasoner

Например, зададим вопрос к онтологии «Какой двигатель используется на самолете Ан-28?», на языке CNL этот вопрос представляется так: «WhoOrWhatusedonAn 28?». Для продолжения

процесса необходимо либо нажать кнопку  в окне Reasoner, либо клавишу «Enter» на клавиатуре. После обработки редактором данных и нахождения решения был получен ответ «Tvd 10 B». Решение соответствует верному (рис.18).

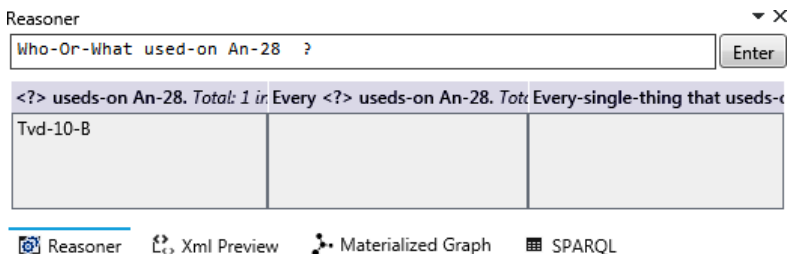


Рис. 18. Окно Reasoner с ответом на вопрос «WhoOrWhatusedonAn 28?», заданный к написанной онтологии

Следует обратить внимание на набор логических выражений, встроенных в редактор. В редакторе они всегда выделены **синим** цветом. Одно из них, использовавшееся ранее для задания массы и мощности двигателя, выражение «equalto». При задании вопроса о массе и мощности, которым соответствуют конкретные числовые значения, также можно использовать логические выражения «lowerthan» («менее чем»), «morethan» («более чем»), «lowerorequalto» («равен или менее чем») и т.д.

К описываемому примеру применим следующий набор вопросов:

«Who Or What has weight lower than 400?», «Who Or What is Tvd 10 B?», «Who Or What has-power equal to 754?», «Who Or What has power greater or equal to 700 and has weight lower than 350?» ит.д.

Создание интернет-ссылок

Если в онтологии используются данные, взятые с сети Интернет, необходимо указать ссылки на них. В редакторе FluentEditor есть специальная функция Reference (от англ. «ссылка») формирующая список использованных источников.

Ссылки создаются в окне Document и прикрепляются только к классамиэкземплярам. Для создания Reference необходимо поставить курсор в окно написания онтологии и нажать сочетание клавиш Ctrl+Spacebar, выбрать в выплывающем окне строку

References: [id] ('http://uri.net

или ввести в окне Document слово «References» после чего написать «[id]» («ссылка на источник») в конце написания списка всех ссылок обязательно ставится точка «.».

Вместо «[id]» следует написать присвоенное классу-экземпляру слово, кратко описывающее ссылку, к примеру, «[engine]», в кавычках («») указывается интернет-адрес («https://en.wikipedia.org/wiki/Engine»). Необходимо добавить выбранный «id» ко всем упоминаниям класса/экземпляра в онтологии. Цифры в имени «id» не используются, а также пробелы, дефисы и заглавные буквы. Чтобы завершить написание Reference ставится точка. На рис. 19 представлен пример создания ссылки на источник в редакторе FluentEditor.

```
Part-1: 'First part of your ontology'.  
  
Every turboprop is an engine.  
Tvd-10-B[engine] is a turboprop.  
Every passenger is a plane.  
An-28 is a passenger.  
Every weight is an engine-characteristics.  
Every power is an engine-characteristics.  
Every thrust is an engine-characteristics.  
Every specific-fuel-consumption is an engine-characteristics.  
  
Tvd-10-B[engine] used-on An-28.  
Tvd-10-B[engine] has-weight equal-to 300.  
Tvd-10-B[engine] has-power equal-to 754.  
  
References:  
[engine] ('https://en.wikipedia.org/wiki/Engine').|
```

Рис. 19. Пример создания интернет-ссылки в онтологии

FluentEditor предоставляет возможность ознакомиться со всеми ссылками в онтологии, используя функцию ReferenceExplorer (от англ. «проводник ссылок»). Данная функция находится в окне быстрого доступа во вкладке «Tools». При этом справа рядом с вкладкой «TaxonomyTree» появляется вкладка

Reference Explorer, где отражаются ссылки на использованные интернет-источники (рис. 20). Для отображения источников, необходимо нажать на стрелочку слева от слов «ReferenceElements».

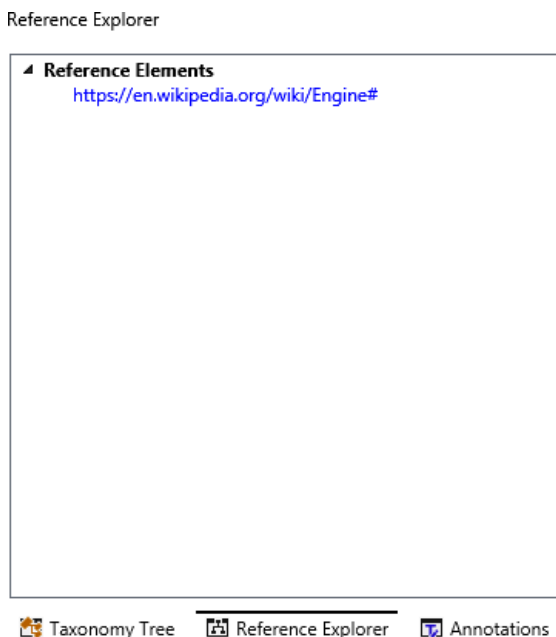


Рис. 20.

Окно Reference Explorer. Создайте проект и сохраните его под именем «Data_base_of_ aircraft_engine»

Задание

1. Озаглавьте проект. Для этого введите «Data_base_of_ aircraft_engine» в строку «Title», а в следующей строке «Author» опишите всех авторов, выполняющих лабораторную работу.

Таблица 1. Примеры самолетных двигателей для создания
ОНТОЛОГИИ

№ п/п	Самолет	Типсамолета	Двигатель	Типдвигателя	Мощностьдвигателя, кВт	Тягадвигателя, кгс	Масса, кг	Удельныйрасходтопливадвигателя, /)
1	Ан-28	пассажирский (passenger)	ТВД-10Б	ТВД (turboprop)	754		300	
2	Ан-38-200	пассажирский (passenger)	ТВД-20-03Б	ТВД (turboprop)	1011		420	0,175
3	Ан-225	транспортный (transport)	Д-18Т	ТРДД (bypass turbofan)		23430	5615	0,34
4	Ту-95МС	бомбардировщик- ракетоносец (bomber-missile)	НК-12МП	ТВД (turboprop)	11029		3170	0,152
5	Ту-4	стратегическийб омбардировщик (strategicbomber)	АШ-73ТК	поршнево й (piston)	1765,2		1355	0,257
6	Ан-10	пассажирский (passenger)	АИ-20	ТВД (turboprop)	2942		1040	0,19
7	Ан-24	пассажирский (passenger)	АИ-24	ТВД (turboprop)	1875,5		600	0,254
8	А-29	многоцелевой (multipurpose)	ВАЗ-426	поршнево й (piston)	154,4		145	0,15
9	МиГ-23Б/БН	истребитель- бомбардировщик (fighter-bomber)	Р-29Б-300	ТРДФ (turbojet with afterburner)		8250	1777	0,78
10	Ту-22М3	бомбардировщик- ракетоносец (bomber-missile)	НК-25	ТРДДФ (bypass turbofan with afterburner)		19000	4275	2,08

Продолжение табл. 1

11	Су-35	истребитель (fighter)	АЛ-31ФП	ТРДДФ (bypass turbofan with afterburner)		7670	1520	0,75
12	Ил-76	военно-транспортный (military transport)	ПС-90А	ТРДД (bypass turbofan)		16000	4160	0,595
13	Ту-154М	пассажирский (passenger)	Д-30КУ-154	ТРДД (bypass turbofan)		10500	2318	0,482
14	Ан-148	пассажирский (passenger)	Д-436-148	ТВД (turbo-prop)		6400	1400	0,36
15	Су-25	штурмовик (attack plane)	Р-95Ш	ТРД (turbojet)		4100	990	0,86
16	МиГ-29	истребитель (fighter)	РД-33	ТРДДФ (bypass turbofan with afterburner)		5040	1055	0,77
17	МиГ-31	перехватчик (interceptor)	Д-30Ф6	ТРДДФ (bypass turbofan with afterburner)		9500	2416	0,72
18	Як-44	ДРЛОиУ и РЭБ (AEW and EW)	Д-27	ТВВД (turbo-propfan)		14000	1650	0,17
19	Ан-72	военно-транспортный (military transport)	Д-36	ТРДД (bypass turbofan)		6500	1124	0,63
20	Ил-86	пассажирский (passenger)	НК-86	ТРДД (bypass turbofan)		13300	2540	0,74
21	МиГ-АТР	учебно-тренировочный (trainer)	РД-1700	ТРДД (bypass turbofan)		1700	297,5	0,5
22	Як-40	пассажирский (passenger)	АИ-25	ТРДД (bypass turbofan)		1500	312	0,56
23	Як-38	штурмовик (attack plane)	Р27В-300	ТРД (turbojet)		6100	1350	0,883
24	И-15	истребитель (fighter)	М-25	поршнево-й (piston)	552		434	0,21
25	Ту-16	многоцелевой (multipurpose)	АМ-3 (РД-3)	ТРД (turbojet)		7000	3100	1

26	3М	бомбардировщик (bomber)	РД-7 (ВД-7)	ТРД (turbojet)		10300	2765	0,8
27	МиГ-15	истребитель (fighter)	ВК-1 (РД-45)	ТРД (turbojet)		2650	872	1,091
28	М-17	перехватчик (interceptor)	РД-36-51В	ТРД (turbojet)		20000	3900	0,882
29	Як-28И	бомбардировщик (bomber)	Р-11АФ2-300	ТРДФ (turbojet with afterburner)		4200	1040	1,96
30	МиГ-19	истребитель (fighter)	РД-9Б	ТРДФ (turbojet with afterburner)		2600	695	0,88

2. Из табл. 1 выберите 5 самолетов. Каждому из них соответствует тип самолета, двигатель, тип двигателя, а также значения массы, мощности, тяги и удельного расхода топлива двигателя.

3. Создайте 3 класса: двигатель (engine), самолет (plane), характеристики двигателя (enginecharacteristics).

4. Классу «enginecharacteristics» присвойте 4 подкласса: «weight» («вес»), «power» («мощность»), «thrust» («тяга»), «specificfuelconsumption» («удельный расходтоплива»).

5. Классу двигатель (engine) присвойте типы двигателей, приведенных в классификации авиационных двигателей. Иерархия классов двигателей указана на рис.7.

6. При создании зависимостей класс-подкласс в онтологии классификации авиационных двигателей используйте отношение «is a».

7. Классу самолет (plane) присвойте типы выбранных Вами самолетов.

8. Добавьте подклассам типов двигателей и типов самолетов экземпляры, выбранные Вами из таблицы 1.

9. Создайте отношение «used on» между экземплярами двигателей и экземплярами самолетов.

10. Присвойте значения характеристик каждому экземпляру двигателей, используя логическое выражение «equalto».

11. Проверьте грамматику написанной онтологии.

12. Постройте CNL диаграмму онтологии, отражающую отношения между экземплярами двигателей и экземплярами самолетов.

13. Осуществите выбор двигателя самолета по параметрам мощности, массы, типа двигателя, самолета, на котором двигатель использовался. Задайте не менее 5 вопросов к онтологии, используя не повторяющиеся логические выражения.

14. Создайте Reference, присвоив каждому двигателю-экземпляру свой [id]. Ссылки на источники возьмите из сети Интернет.

15. О на контрольные вопросы.

Контрольные вопросы

1. Что такое онтология?
2. Что является результатом онтологического анализа?
3. Цель создания онтологии?
4. Что является главным отличием FluentEditor от других редакторов онтологий?
5. Что такое контролируемые языки и для чего они необходимы?
6. Какой встроенный механизм использует FluentEditor для отслеживания грамматики онтологии на CNL?
7. Каким образом осуществляется создание классов, подклассов, экземпляров в FluentEditor?

8. Какие существуют правила написания имени класса, подкласса, экземпляра?
9. Для чего используется функция Reasoner?
10. Какие логические выражения используются для создания онтологий в FluentEditor?

Литература

1. Башмаков А. И., Башмаков И. А. Интеллектуальные [информационные технологии](#): Учеб. пособие. – М.: Изд-во МГТУ им. Н. Э. Баумана, 2005. – 304 с.
2. Варшавский П. Р., Куриленко И. Е., Михайлов И. С. Программное обеспечение интеллектуальных систем: [учебное пособие](#) / – М.: Издательский дом МЭИ, 2011. – 64 с.
3. Куриленко И. Е. Современные методологии разработки программных средств: учебное пособие / – М.: Издательский дом МЭИ, 2011. – 112 с.

Лабораторная работа 4

Поддержка принятия решений на основе построения моделей в системе WiMi

В данной лабораторной работе представлено детальное описание практического применения программного продукта WiMi в предметной области «Информационная система грузового порта». Так же представлено краткое описание самого WiMi, расписана область его применения, функционирование программы в целом, а также приведена краткая выжимка из инструкции по работе с данным ПО.

Описание функционирования программы и область применения

WiMi предназначена для создания специализированных информационных систем, способных в режиме реального времени решать сложные логические задачи, связанные с обработкой больших массивов информации. WiMi позволяет вычислять искомые параметры из исходных параметров путём нахождения алгоритмов вычислений. При этом производится расчёт всех необходимых промежуточных параметров.

WiMi в процессе вычисления рассчитывает только необходимые параметры, связанные в алгоритмические цепочки вычислений, а не все возможные параметры. Благодаря такому подходу достигается высокая производительность программы при небольших затрачиваемых ресурсах.

Технология runtime, добавленная во второй версии WiMi, при построении алгоритма решения задачи также учитывает значения рассчитанных промежуточных параметров. Так, при наличии нескольких альтернативных цепочек система основывает свой выбор на значении рассчитанного параметра (см. Рис. 21). Таким образом, если правила имеют несколько выходных параметров, а рассчитаны будут только некоторые из них, то дальнейший алгоритм может быть перестроен таким образом, чтобы в расчётах участвовали только рассчитанные параметры.

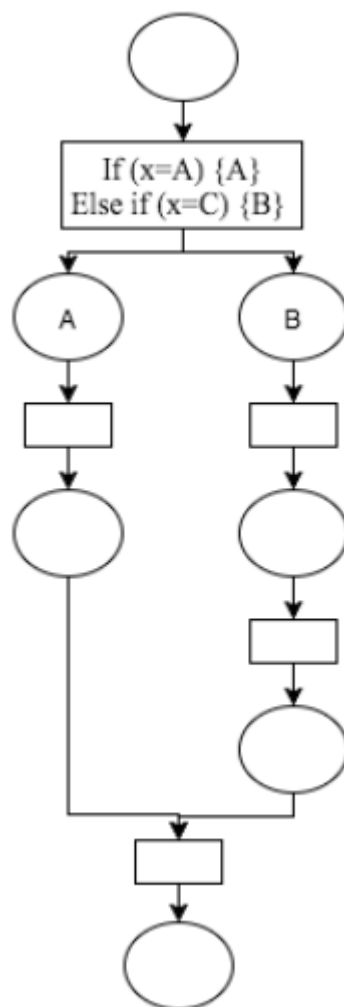


Рис. 21 – Иллюстрация технологии runtime

Область применения:

Программа обеспечивает выполнение следующих функций:

- 1) Создание и редактирование описания моделей предметных областей:
 - создание и редактирование объектов предметной области (параметров и классов);
 - создание и редактирование отношений и правил, связывающих эти объекты.
- 2) Структурный анализ моделей, получение логического вывода решения и объяснение его в виде последовательности выполненных действий;
- 3) Создание и вывод полученного алгоритма логического вывода разрешения ситуации, расчет необходимых значений.

Краткая выжимка из руководства пользователя по WiMi

В процессе создания модели все объекты будут располагаться в древовидном списке в левой части приложения, а все отношения — в правой части. Редактирование соответствующих элементов модели производится во вкладках, расположенных в центральной части приложения. Если в процессе создания модели были обнаружены ошибки, соответствующие записи появятся в списке в нижней части приложения (рис. 22).

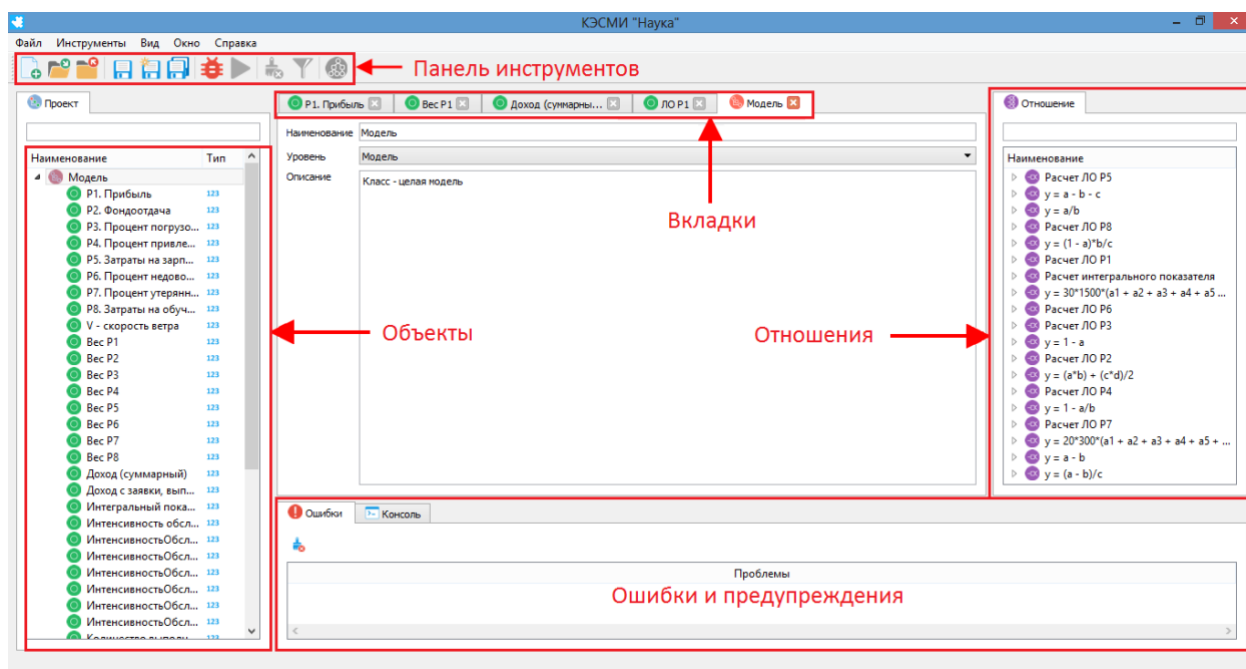


Рис. 22 – Основное окно приложения

Существуют 2 типа объектов: классы и параметры.

Классы – это некие абстрактные сущности, в которых хранятся другие классы и параметры. Для лучшего понимания классы можно представлять как некие папки, внутри которых находятся параметры и другие папки (классы).

Параметры – это все переменные в модели (и те, которые необходимо будет найти, и те, что являются входной информацией). Параметры, как и классы, имеют наименование, уровень иерархии и описание. Так же параметры имеют один из двух типов значений (число; текст). В параметрах можно задавать значения по умолчанию.

У каждого параметра есть 2 таблицы: таблица «Правил» и таблица «Ограничений». Эти таблицы заполняются автоматически, когда мы создаем правила или ограничения. Они служат для отображения связей между параметром и связанными с ним правилами и ограничениями. У одного параметра может быть несколько правил и ограничений.

Отношение – это некая сущность, которая хранит в себе математические выражения (формулы, условия, ограничения, сложные выражения), которые используются в модели для преобразования входных данных (параметров) в выходные.

У каждого отношения есть свой «Тип отношения» и «Тело отношения».

Тип отношения – это тип математического выражения. Существует четыре типа отношения:

- 1) формула – математическая формула, состоящая из переменных и операторов;
- 2) условное отношение – отношение вида "если..., то... иначе...";
- 3) ограничение – отношение, накладывающее ограничение на значения параметров;
- 4) сложное отношение – программируемое отношение для опытных пользователей.

Каждому типу отношения соответствует свое тело отношения.

Тело отношения – строка выражения данного отношения. В зависимости от типа отношения его тело будет отличаться по оформлению. Если тип отношения «формула», то в теле отношения будет расписана формула, если тип отношения «сложное отношение», то в теле отношения будет расписан код на JavaScript.

Правило – это некая сущность, которая связывает отношение с конкретными параметрами.

Например, в отношении мы задаем формулу: $y = (a+b) / c$. После записи данной формулы мы ждем на кнопку, чтобы проверить правильность заведения формулы. Если формулу мы записали правильно, то программа автоматически расставляет входные и выходные параметры. Входными параметрами она считает все переменные, записанные после знака «=», а выходными переменными – те, что указаны перед знаком «=».

Теперь нам необходимо все эти переменные «у», «а», «b», «с» соотнести с реальными параметрами, которые занесены у нас в системе. Как раз для этого и создаются правила. В правилах мы указываем соответствие параметров из отношения (переменные «у», «а», «b», «с») и параметров, имеющих в системе.

Ограничение – это вид правила, проверяющий входные данные на корректность. Например, сторона треугольника не может быть отрицательной. Если при запуске модели мы зададим для параметра, являющегося стороной треугольника отрицательное значение, то система выдаст сообщение о прекращении работы модели и укажет, что сработало ограничение.

Описание задачи по предметной области

На работу грузового порта, как в отдельности, так и в целом, влияют внутренние и внешние факторы (например, интенсивность работы того или иного отдела/сотрудника, скорость прихода разрешения на работу с заявкой, и т.д.).

Так же не маловажную роль на работу грузового порта оказывают и показатели. К ним можно отнести затраты на обучение сотрудников, процент недовольных клиентов, процент утерянного груза при погрузке, и т.д.

Для того чтобы правильно рассчитывать коэффициенты показателей необходимо внести новые понятия: «Лингвистическая оценка» и «Общий интегральный показатель».

Лингвистическая оценка – это определенный метод, способный каждому значению показателя установить оценку для будущего определения общего интегрального показателя.

Общий интегральный показатель рассчитывается по формуле, описанной ниже:

$$\varphi = \sum_{i=1}^n w_i * l_i - \sum_{j=1}^m w_j * l_j, \text{ где}$$

φ - общий интегральный показатель,

w_i - вес целевого показателя,

l_i - лингвистическая оценка целевого показателя,

w_j - вес контрольного показателя,

l_j - лингвистическая оценка контрольного показателя.

Для расчета интегрального показателя необходимо расставить степень важности каждому показателю, т.е. его вес и определить лингвистическую оценку для каждого показателя.

Определение общего интегрального показателя будет являться основанием для принятия управленческих решений с учетом всех проведенных экспериментов (комбинаций изменения факторов влияния).

Из всех полученных результатов выбор делается в пользу варианта, имеющего наибольшее значение общего интегрального показателя.

Главной задачей является нахождение баланса между внутренними и внешними факторами, значением показателей, их лингвистической оценкой и весами, при которых управление грузовым портом будет наиболее эффективным.

Решение поставленной задачи с помощью WiMi

После прочтения руководства пользователя по WiMi приходит понимание, что в самом начале работы с системой необходимо определиться с конечным количеством параметров, которые будут принимать участие в работе. Параметрами являются все переменные в модели (и те, которые необходимо будет найти, и те, что являются входной информацией). Параметрами в системе будут являться показатели, их вес,

лингвистическая оценка, внутренние и внешние факторы, а так же второстепенные параметры, которые необходимы для нахождения первостепенных параметров.

Для начала выпишем в виде таблицы параметры, которые мы явно имеем (первостепенные параметры). Каждое поле в таблице является параметром.

Таблица 2. Явные (первостепенные) параметры в системе

Показатели	Лингв. оценка	Вес показателя	Внутренние и внешние факторы
P1. Прибыль	ЛО P1	Вес P1	Интенсивность обслуживания (далее – ИО) ведущего инженера
P2. Фондоотдача	ЛО P2	Вес P2	ИО тальманского отдела
P3. Процент погрузочных работ, выполненных в срок	ЛО P3	Вес P3	ИО директора
P4. Процент привлеченных клиентов	ЛО P4	Вес P4	ИО отдела бухгалтерии
P5. Затраты на зарплату сотрудникам	ЛО P5	Вес P5	ИО начальника порта
P.6 Процент недовольных клиентов	ЛО P6	Вес P6	ИО ведущего тальмана
P7. Процент утерянного груза при погрузке	ЛО P7	Вес P7	ИО инженерного отдела
P8. Затраты на обучение сотрудников	ЛО P8	Вес P8	-
Интегральный показатель	-	-	-

Вид таблицы был выбран не случайно, с помощью нее лучше воспринимается взаимосвязь первоначальных параметров между собой.

Далее необходимо понять, как в системе будет задаваться/рассчитываться каждый из вышеприведенных параметров. Если эти параметры будут рассчитываться по какой-либо формуле, то каждый элемент формулы так же будет являться параметром (второстепенным).

Начнем с параметров-показателей:

1) **P1. Прибыль** = Доход (суммарный) – P5. Затраты на зарплату сотрудникам – P8. Затраты на обучение;

Доход (суммарный) = (Доход с заявки, выполненной в срок * Количество заявок, выполненных в срок) + (Доход с заявки, выполненной в срок * Количество заявок, выполненных не в срок)/2;

Доход с заявки, выполненной в срок = 65000;

Количество заявок, выполненных в срок = Количество выполненных заявок - Количество заявок, выполненных не в срок;

Количество выполненных заявок – значение задается вручную;

Количество заявок, выполненных не в срок - значение задается вручную;

2) **P2. Фондоотдача** = Доход (суммарный)/Основные средства;

Основные средства = 2500000;

3) **P3. Процент погрузочных работ, выполненных в срок** = (Количество выполненных заявок - Количество заявок, выполненных не в срок)/Количество поступивших заявок;

Количество поступивших заявок – значение задается вручную;

4) **P4. Процент привлеченных клиентов** = (1 – P7. Процент утерянного груза при погрузке) * Количество выполненных заявок / Количество поступивших заявок;

5) **P5. Затраты на зарплату сотрудникам** = 30 * 150 * (отдел1 + отдел2 + отдел3 + отдел4 + отдел5 + отдел6 + отдел7);

Отделы – это факторы «Интенсивность обслуживания ...»;

6) **P6. Процент недовольных клиентов** = 1 – P4. Процент привлеченных клиентов;

7) **P7. Процент утерянного груза при погрузке** = 1 - Количество выполненных заявок / Количество поступивших заявок;

8) **P8. Затраты на обучение сотрудников** = 2,5 * 300 * (отдел1 + отдел2 + отдел3 + отдел4 + отдел5 + отдел6 + отдел7);

Выше были выявлены новые параметры (второстепенные), которые участвуют в расчете параметров-показателей. Выпишем эти параметры:

- Доход (суммарный);
- Доход с заявки, выполненной в срок;
- Количество заявок, выполненных в срок;
- Количество заявок, выполненных не в срок;
- Количество выполненных заявок;
- Основные средства;

- Количество поступивших заявок;

Перейдем к параметрам, которые являются весами для показателей. Все следующие параметры-веса имеют predetermined значения, которые задаются на вкладках данных параметров: Вес P1, Вес P2, Вес P3, Вес P4, Вес P5, Вес P6, Вес P7, Вес P8.

Для параметров, которые являются внутренними и внешними факторами, значение задается вручную при тестировании модели.

Все параметры лингвистических оценок рассчитываются сложным отношением. Далее эти отношения будут рассмотрены в специальной таблице.

И последний, самый главный параметр «Интегральный показатель» рассчитывается по формуле: $y = (w1 \cdot I1 + w2 \cdot I2 + w3 \cdot I3 + w4 \cdot I4) - (w5 \cdot I5 + w6 \cdot I6 + w7 \cdot I7 + w8 \cdot I8)$,

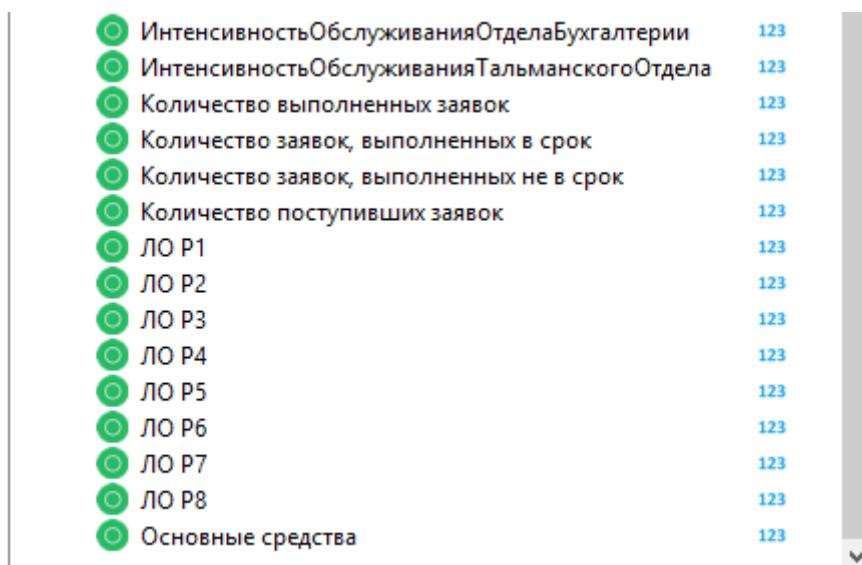
где $w1, w2, \dots, w8$ – веса соответствующих параметров;

$I1, I2, \dots, I8$ – лингвистические оценки соответствующих параметров.

После того как были выявлены новые (второстепенные) параметры, и определены способы расчета для всех параметров, можно создавать в системе данные параметры (рис. 23, 24).

Наименование	Тип
Модель	
Р1. Прибыль	123
Р2. Фондоотдача	123
Р3. Процент погрузочных работ, выполненных в срок	123
Р4. Процент привлеченных клиентов	123
Р5. Затраты на зарплату сотрудникам	123
Р6. Процент недовольных клиентов	123
Р7. Процент утерянного груза при погрузке	123
Р8. Затраты на обучение сотрудников	123
Вес P1	123
Вес P2	123
Вес P3	123
Вес P4	123
Вес P5	123
Вес P6	123
Вес P7	123
Вес P8	123
Доход (суммарный)	123
Доход с заявки, выполненной в срок	123
Интегральный показатель	123
Интенсивность обслуживания начальника порта	123
ИнтенсивностьОбслуживанияВедущегоИнженера	123
ИнтенсивностьОбслуживанияВедущегоТальмана	123
ИнтенсивностьОбслуживанияДиректора	123
ИнтенсивностьОбслуживанияИнженерногоОтдела	123
ИнтенсивностьОбслуживанияОтделаБухгалтерии	123
ИнтенсивностьОбслуживанияТальманскогоОтдела	123
Количество выполненных заявок	123
Количество заявок, выполненных в срок	123

Рис. 23 – Занесенные параметры в WiMi



○ Интенсивность Обслуживания Отдела Бухгалтерии	123
○ Интенсивность Обслуживания Тальманского Отдела	123
○ Количество выполненных заявок	123
○ Количество заявок, выполненных в срок	123
○ Количество заявок, выполненных не в срок	123
○ Количество поступивших заявок	123
○ ЛО Р1	123
○ ЛО Р2	123
○ ЛО Р3	123
○ ЛО Р4	123
○ ЛО Р5	123
○ ЛО Р6	123
○ ЛО Р7	123
○ ЛО Р8	123
○ Основные средства	123

Рис. 24 – Занесенные параметры в WiMi

После заведения всех параметров в системе, создаются отношения, хранящие в себе математические выражения, которые будут использоваться в модели для преобразования входных данных (параметров) в выходные.

После, к каждому отношению создается правило, которое связывает параметры из данного отношения с параметрами системы.

Для удобства восприятия все отношения и правила будут расписаны в 2-х таблицах.

Таблица 3. Отношения

№	Наименование	Тип отношения	Тело отношения	Пояснение
1	Расчет ЛО P5	Сложное	<pre>var p5, y; if (p5 < 245000) {y = 1;} else if (p5 < 265000) {y = 2;} else if (p5 < 285000) {y = 3;} else if (p5 < 300000) {y = 4;} elseif (p5 > 300000) {y = 5;}</pre>	<p>Данное отношение сложное. Поэтому сначала создаются 2 переменные: p5, y. P5 –это параметр «Затраты на зарплату сотрудникам», а Y–это значение ЛОР5. Далее идут простые условия типа:</p> <p>ЕСЛИ p5 <x1, ТО y = z1, ИНАЧЕ ЕСЛИ p5 <x2, ТО y =z2, ИНАЧЕ ЕСЛИ ...</p>
2	$y = a - b - c$	Формула	$y = a - b - c$	Выражение, когда из одного числа вычитаются два других
3	$y = a/b$	Формула	$y = a/b$	Выражение, показывающее отношение двух величин
4	Расчет ЛО P8	Сложное	<pre>var p8, y; if (p8 < 10000) {y = 1;} else if (p8 < 20000) {y = 2;} else if (p8 < 30000) {y = 3;} else if (p8 < 35000) {y = 4;} elseif (p8 > 35000) {y = 5;}</pre>	<p>Данное отношение сложное. Поэтому сначала создаются 2 переменные: p8, y. P8 –это параметр «Затраты на обучение сотрудников», а Y–это значение ЛО P8. Далее идут простые условия типа:</p> <p>ЕСЛИ p8<x1, ТО y = z1, ИНАЧЕ ЕСЛИ p8<x2, ТО y =z2, ИНАЧЕ ЕСЛИ ...</p>
5	$y = (1 - a)*b/c$	Формула	$y = (1 - a)*b/c$	Выражение, предназначенное для расчета процента привлеченных клиентов
6	Расчет ЛО P1	Сложное	<pre>var p1, y; if (p1 < 50000) {y = 1;} else if (p1 < 100000) {y = 2;}</pre>	<p>Данное отношение сложное. Поэтому сначала создаются 2 переменные: p1, y. P1–это параметр «Прибыль», а Y–это значение ЛО P1. Далее идут простые условия типа:</p>

№	Наименование	Тип отношения	Тело отношения	Пояснение
			<pre>else if (p1 < 200000) {y = 3;} else if (p1 < 300000) {y = 4;} elseif (p1 > 300000) {y = 5;}</pre>	ЕСЛИ $p1 < x1$, ТО $y = z1$, ИНАЧЕ ЕСЛИ $p1 < x2$, ТО $y = z2$, ИНАЧЕ ЕСЛИ ...
7	Расчет интегрального показателя	Формула	$y = (w1 * I1 + w2 * I2 + w3 * I3 + w4 * I4) - (w5 * I5 + w6 * I6 + w7 * I7 + w8 * I8)$	Формула расчета интегрального показателя
8	$y = 30 * 1500 * (a1 + a2 + a3 + a4 + a5 + a6 + a7)$	Формула	$y = 30 * 1500 * (a1 + a2 + a3 + a4 + a5 + a6 + a7)$	Выражение, предназначенное для расчета затрат на выдачу заработной платы сотрудникам
9	Расчет ЛО Р6	Сложное	<pre>var p6, y; if (p6 < 0.01) {y = 1;} else if (p6 < 0.02) {y = 2;} else if (p6 < 0.03) {y = 3;} else if (p6 < 0.06) {y = 4;} elseif (p6 > 0.06) {y = 5;}</pre>	<p>Данное отношение сложное. Поэтому сначала создаются 2 переменные: p6, y. Р6—это параметр «Процент недовольных клиентов», а Y—это значение ЛО Р6. Далее идут простые условия типа:</p> <p>ЕСЛИ $p6 < x1$, ТО $y = z1$, ИНАЧЕ ЕСЛИ $p6 < x2$, ТО $y = z2$, ИНАЧЕ ЕСЛИ ...</p>
10	Расчет ЛО Р3	Сложное	<pre>var p3, y; if (p3 < 0.2) {y = 1;} else if (p3 < 0.4) {y = 2;} else if (p3 < 0.6) {y = 3;} else if (p3 < 0.8) {y = 4;} elseif (p3 > 0.8) {y = 5;}</pre>	<p>Данное отношение сложное. Поэтому сначала создаются 2 переменные: p3, y. Р3—это параметр «Процент погрузочных работ, выполненных в срок», а Y—это значение ЛО Р3.</p> <p>Далее идут простые условия типа:</p> <p>ЕСЛИ $p3 < x1$, ТО $y = z1$, ИНАЧЕ ЕСЛИ $p3 < x2$, ТО $y = z2$, ИНАЧЕ ЕСЛИ ...</p>

№	Наименование	Тип отношения	Тело отношения	Пояснение
11	$y = 1 - a$	Формула	$y = 1 - a$	Выражение, предназначенное для расчета процента недовольных клиентов
12	Расчет ЛО P2	Сложное	<pre> var p2, y; if (p2 < 0.2) {y = 1;} else if (p2 < 0.4) {y = 2;} else if (p2 < 0.6) {y = 3;} else if (p2 < 0.8) {y = 4;} elseif (p2 > 0.8) {y = 5;} </pre>	Данное отношение сложное. Поэтому сначала создаются 2 переменные: p2, y. P2—это параметр «Фондоотдача», а Y—это значение ЛО P2. Далее идут простые условия типа: ЕСЛИ p2<x1, ТО y = z1, ИНАЧЕ ЕСЛИ p2<x2, ТО y =z2, ИНАЧЕ ЕСЛИ ...
13	$y = (a*b) + (c*d)/2$	Формула	$y = (a*b) + (c*d)/2$	Выражение для расчета суммарного дохода (по всем выполненным заявкам)
14	Расчет ЛО P4	Сложное	<pre> var p4, y; if (p4 < 0.01) {y = 1;} else if (p4 < 0.02) {y = 2;} else if (p4 < 0.03) {y = 3;} else if (p4 < 0.06) {y = 4;} elseif (p4 > 0.06) {y = 5;} </pre>	Данное отношение сложное. Поэтому сначала создаются 2 переменные: p4, y. P4—это параметр «Процент привлеченных клиентов», а Y—это значение ЛО P4. Далее идут простые условия типа: ЕСЛИ p4<x1, ТО y = z1, ИНАЧЕ ЕСЛИ p4<x2, ТО y =z2, ИНАЧЕ ЕСЛИ ...
15	$y = 1 - a/b$	Формула	$y = 1 - a/b$	Выражение, предназначенное для расчета процента утерянного груза при погрузке
16	Расчет ЛО P7	Сложное	<pre> var p7, y; if (p7 < 0.1) {y = 1;} </pre>	Данное отношение сложное. Поэтому сначала создаются 2 переменные: p7, y. P7—это параметр «Процент утерянного

№	Наименование	Тип отношения	Тело отношения	Пояснение
			else if (p7 < 0.3) {y = 2;} else if (p7 < 0.4) {y = 3;} else if (p7 < 0.6) {y = 4;} elseif (p7 > 0.6) {y = 5;}	груза при погрузке», а Y–это значение ЛО Р7. Далее идут простые условия типа: ЕСЛИ p7<x1, ТО y = z1, ИНАЧЕ ЕСЛИ p7<x2, ТО y =z2, ИНАЧЕ ЕСЛИ ...
17	$y = 20*300*(a1 + a2 + a3 + a4 + a5 + a6 + a7)$	Формула	$y = 20*300*(a1 + a2 + a3 + a4 + a5 + a6 + a7)$	Выражение для расчета затрат на обучение сотрудников
18	$y = a - b$	Формула	$y = a - b$	Выражение, когда из одного числа вычитается другое
19	$y = (a - b)/c$	Формула	$y = (a - b)/c$	Выражение, показывающее отношение между разностью двух величин и третьей величиной

Таблица 4. Правила

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметрв системе	Наименование	Параметр в системе
1	1	ЛО Р5	p5	Р5. Затраты на зарплату сотрудникам	y	ЛО Р5
2	2	Расчет прибыли	a	Доход (суммарный)	y	Р1. Прибыль
			b	Р5. Затраты на зарплату сотрудникам		

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметр в системе	Наименование	Параметр в системе
			с	Р8. Затраты на обучение сотрудников		
3	3	Расчет фондоотдачи	а	Доход (суммарный)	у	Р2. Фондоотдача
			б	Основные средства		
4	4	ЛО Р8	р8	Р8. Затраты на обучение сотрудников	у	ЛО Р8
5	5	Расчет процента привлеченных клиентов	а	Р7. Процент утерянного груза при погрузке	у	Р4. Процент привлеченных клиентов
			б	Количество выполненных заявок		
			с	Количество поступивших заявок		
6	6	ЛО Р1	р1	Р1. Прибыль	у	ЛО Р1
7	7	Интегральный показатель	l1	ЛО Р1	у	Интегральный показатель
			w1	Вес Р1		
			w4	Вес Р4		
			w6	Вес Р6		
			l7	ЛО Р7		
			w3	Вес Р3		

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметр в системе	Наименование	Параметр в системе
			14	ЛО P4		
			12	ЛО P2		
			18	ЛО P8		
			w5	Вес P5		
			15	ЛО P5		
			16	ЛО P6		
			13	ЛО P3		
			w7	Вес P7		
			w2	Вес P2		
			w8	Вес P8		
8	8	Расчет затрат на заработную плату сотрудникам	a2	Интенсивность Обслуживания Тальманского Отдела	у	P5. Затраты на зарплату сотрудникам
			a7	Интенсивность Обслуживания Инженерного Отдела		
			a4	Интенсивность Обслуживания Отдела Бухгалтерии		
			a3	Интенсивность Обслуживания Директора		
			a6	Интенсивность Обслуживания		

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметр в системе	Наименование	Параметр в системе
				Ведущего Тальмана		
			a1	Интенсивность обслуживания Ведущего Инженера		
			a5	Интенсивность обслуживания начальника порта		
9	9	ЛО Р6	p6	Р6. Процент недовольных клиентов	у	ЛО Р6
10	10	ЛО Р3	p3	Р3. Процент погрузочных работ, выполненных в срок	у	ЛО Р3
11	11	Расчет процента недовольных клиентов	a	Р4. Процент привлеченных клиентов	у	Р6. Процент недовольных клиентов
12	12	ЛО Р2	p2	Р2. Фондоотдача	у	ЛО Р2
13	13	Расчет дохода суммарного	a	Доход с заявки, выполненной в срок	у	Доход (суммарный)
			d	Количество заявок, выполненных не в срок		
			b	Количество заявок, выполненных в срок		

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметр в системе	Наименование	Параметр в системе
			с	Доход с заявки, выполненной в срок		
14	14	ЛО Р4	р4	Р4. Процент привлеченных клиентов	у	ЛО Р4
15	15	Расчет процента утерянного груза при погрузке	а	Количество выполненных заявок	у	Р7. Процент утерянного груза при погрузке
			б	Количество поступивших заявок		
16	16	ЛО Р7	р7	Р7. Процент утерянного груза при погрузке	у	ЛО Р7
17	17	Расчет затрат на обучение сотрудников	а2	Интенсивность Обслуживания Тальманского Отдела	у	Р8. Затраты на обучение сотрудников
			а7	Интенсивность Обслуживания Инженерного Отдела		
			а4	Интенсивность Обслуживания Отдела Бухгалтерии		
			а3	Интенсивность Обслуживания Директора		
			а6	Интенсивность Обслуживания		

№	№ отношения	Наименование	Соответствие параметров из отношения с параметрами системы			
			Входные параметры		Выходные параметры	
			Наименование	Параметр в системе	Наименование	Параметр в системе
				Ведущего Тальмана		
			a1	Интенсивность обслуживания Ведущего Инженера		
			a5	Интенсивность обслуживания начальника порта		
18	18	Расчет количества выполненных заявок в срок	a	Количество выполненных заявок	y	Количество заявок, выполненных в срок
			b	Количество заявок, выполненных не в срок		
19	19	Расчет процента погрузочных работ, выполненных в срок	a	Количество выполненных заявок	y	P3. Процент погрузочных работ, выполненных в срок
			b	Количество заявок, выполненных не в срок		
			c	Количество поступивших заявок		

Созданные отношения и правила на их основе изображены на рисунках 25, 26, 27. У отношений рядом с наименованием значок фиолетового цвета, а у правил оранжевого цвета.

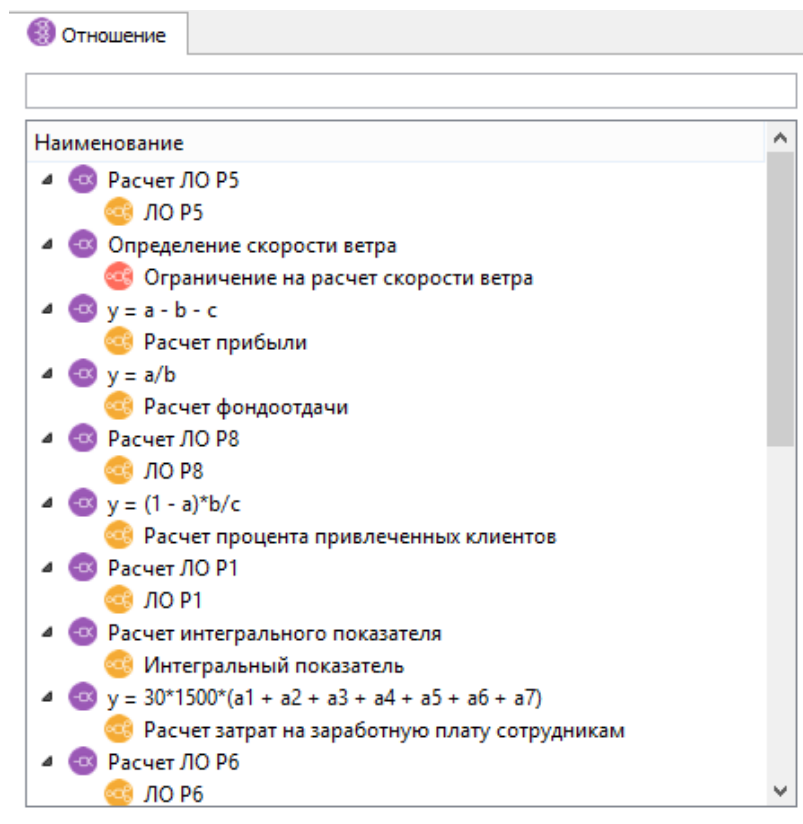


Рис. 25 – Занесенные отношения и правила в WiMi

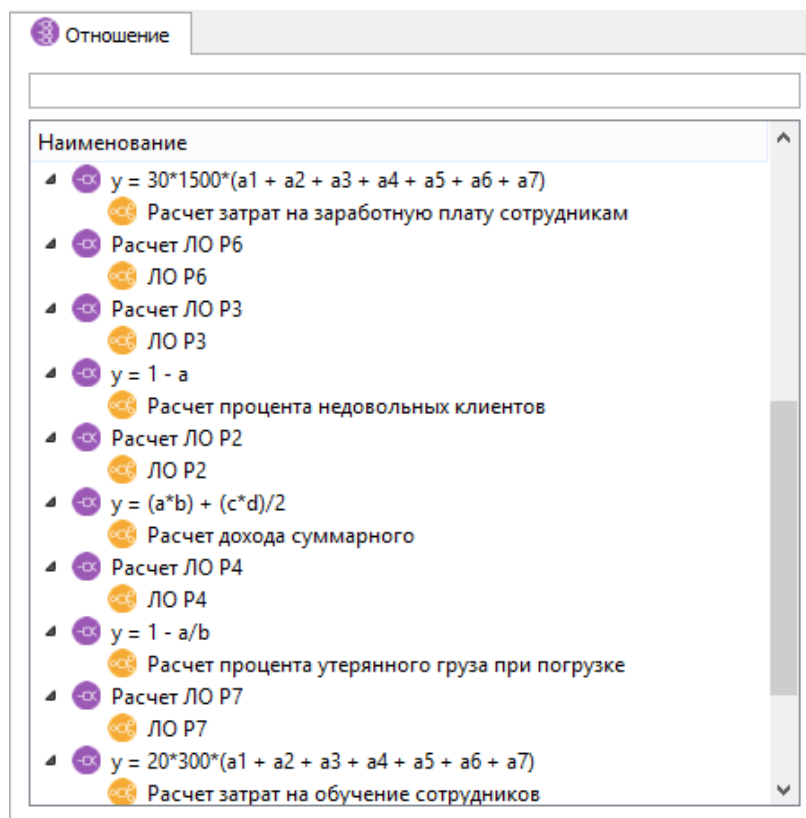


Рис. 26 – Занесенные отношения и правила в WiMi

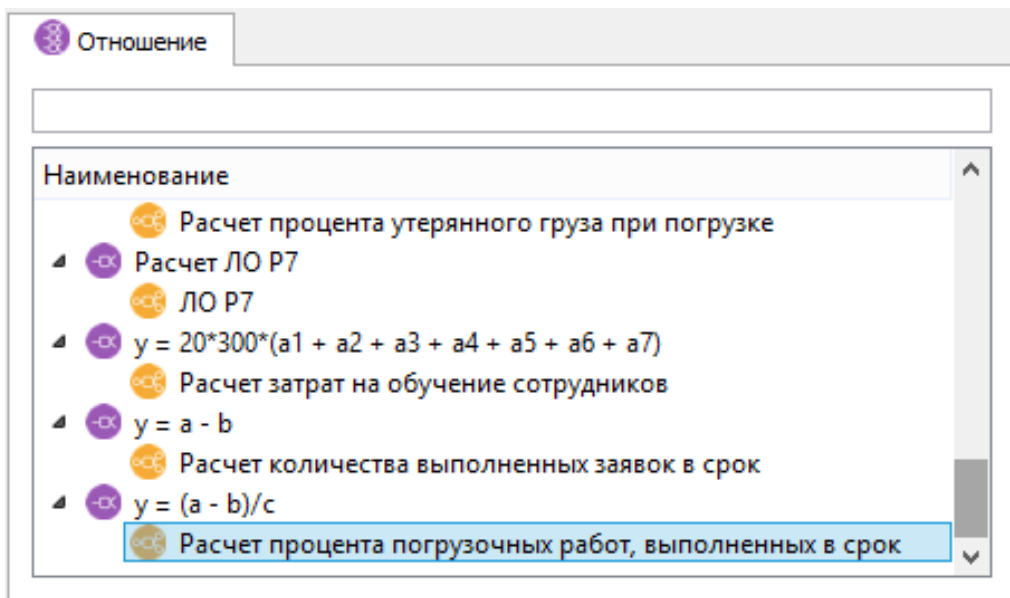


Рис. 27 - Занесенные отношения и правила в WiMi

Таким образом, в систему были занесены все параметры, отношения и правила. После всех этих действий можно приступать к тестированию модели.

Вкладка тестирования модели выглядит следующим образом.

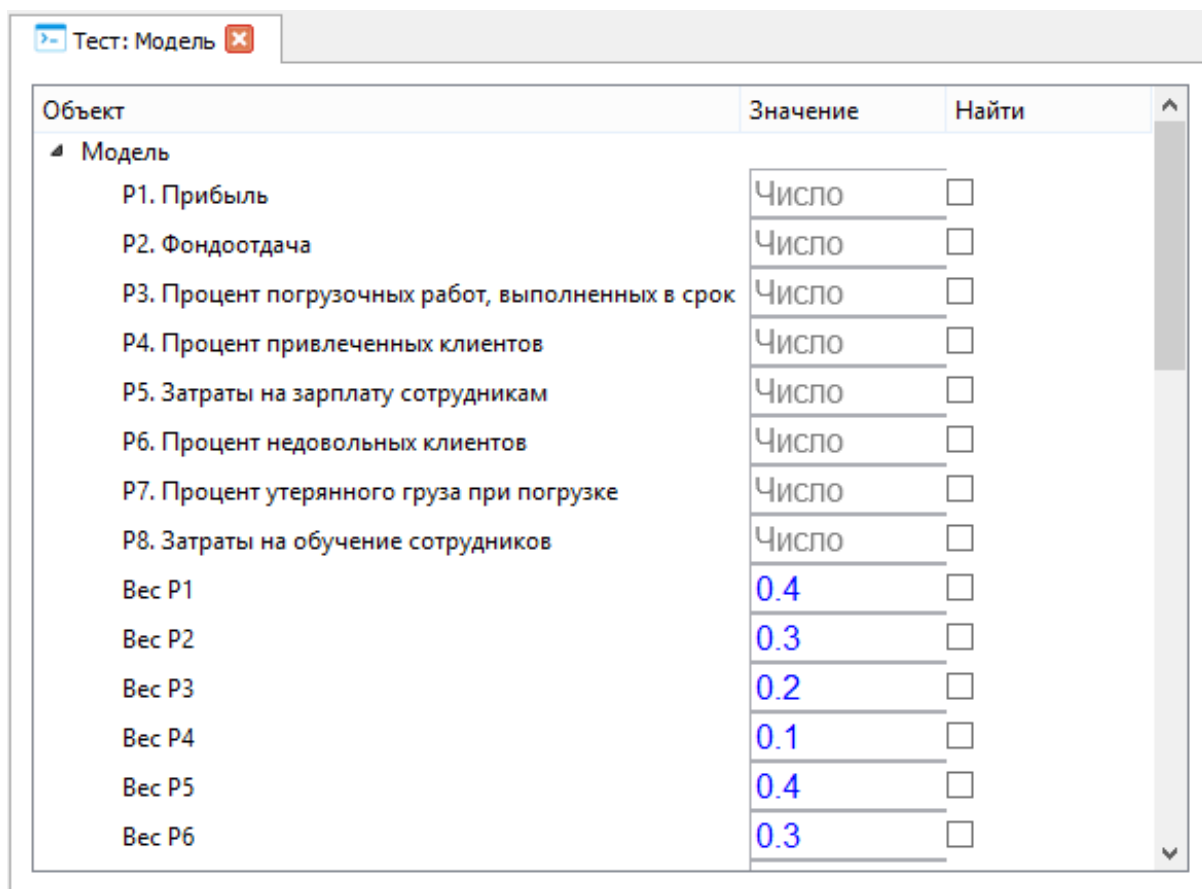


Рис. 28 – Вкладка «Тестирование модели»

Объект	Значение	Найти
Вес Р6	0.3	<input type="checkbox"/>
Вес Р7	0.2	<input type="checkbox"/>
Вес Р8	0.1	<input type="checkbox"/>
Доход (суммарный)	Число	<input type="checkbox"/>
Доход с заявки, выполненной в срок	65000	<input type="checkbox"/>
Интегральный показатель	Число	<input type="checkbox"/>
Интенсивность обслуживания начальника порта	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоИнженера	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоТальмана	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияДиректора	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияИнженерногоОтдела	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияОтделаБухгалтерии	Число	<input type="checkbox"/>
ИнтенсивностьОбслуживанияТальманскогоОтдела	Число	<input type="checkbox"/>
Количество выполненных заявок	Число	<input type="checkbox"/>
Количество заявок, выполненных в срок	Число	<input type="checkbox"/>

Рис. 29 – Вкладка «Тестирование модели»

Объект	Значение	Найти
ИнтенсивностьОбслуживанияТальманскогоОтдела	Число	<input type="checkbox"/>
Количество выполненных заявок	Число	<input type="checkbox"/>
Количество заявок, выполненных в срок	Число	<input type="checkbox"/>
Количество заявок, выполненных не в срок	Число	<input type="checkbox"/>
Количество поступивших заявок	Число	<input type="checkbox"/>
ЛО Р1	Число	<input type="checkbox"/>
ЛО Р2	Число	<input type="checkbox"/>
ЛО Р3	Число	<input type="checkbox"/>
ЛО Р4	Число	<input type="checkbox"/>
ЛО Р5	Число	<input type="checkbox"/>
ЛО Р6	Число	<input type="checkbox"/>
ЛО Р7	Число	<input type="checkbox"/>
ЛО Р8	Число	<input type="checkbox"/>
Основные средства	2500000	<input type="checkbox"/>

Рис. 30 – Вкладка «Тестирование модели»

Данная вкладка визуально поделена на 3 колонки. Первая колонка – название параметра, вторая колонка – значение данного параметра, третья колонка – реквизит типа «Булево», в нем ставится флажок, который показывает, что данный параметр необходимо найти в итоге.

Для некоторых параметров значения уже заполнены – это predetermined значения, которые задаются на вкладке параметра.

В нашей задаче необходимо найти интегральный показатель. Для этого необходимо заполнить значениями лишь часть параметров (начальные), остальные параметры будут рассчитаны в процессе нахождения интегрального показателя, и поставить флажок рядом с параметром «Интегральный показатель». На рисунке 31 представлены заполненные значения начальных параметров.

Объект	Значение	Найти
Доход (суммарный)	Число	<input type="checkbox"/>
Доход с заявки, выполненной в срок	65000	<input type="checkbox"/>
Интегральный показатель	Число	<input checked="" type="checkbox"/>
Интенсивность обслуживания начальника порта	0,9	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоИнженера	0,85	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоТальмана	0,79	<input type="checkbox"/>
ИнтенсивностьОбслуживанияДиректора	0,87	<input type="checkbox"/>
ИнтенсивностьОбслуживанияИнженерногоОтдела	0,89	<input type="checkbox"/>
ИнтенсивностьОбслуживанияОтделаБухгалтерии	0,8	<input type="checkbox"/>
ИнтенсивностьОбслуживанияТальманскогоОтдела	0,71	<input type="checkbox"/>
Количество выполненных заявок	50	<input type="checkbox"/>
Количество заявок, выполненных в срок	Число	<input type="checkbox"/>
Количество заявок, выполненных не в срок	4	<input type="checkbox"/>
Количество поступивших заявок	50	<input type="checkbox"/>
ЛО Р1	Число	<input type="checkbox"/>

Рис. 31 – Заполненные необходимые для расчета значения начальных параметров

После заполнения необходимых для расчета значений начальных параметров и проставления флага рядом с параметром «Интегральный показатель», необходимо нажать на кнопку «Запустить», расположенную на панели инструментов. В случае успеха, рассчитанные параметры будут выделены красным цветом, а в консоли будет выводиться вся текущая информация: найденные и не найденные значения и логический вывод с указанием использованного правила, его описания, входных, выходных и найденных значений (рис. 32, 33, 34, 35, 36, 37).

Объект	Значение	Найти
Модель		
Р1. Прибыль	2823690	<input type="checkbox"/>
Р2. Фондоотдача	1.248	<input type="checkbox"/>
Р3. Процент погрузочных работ, выполненных в срок	0.92	<input type="checkbox"/>
Р4. Процент привлеченных клиентов	1	<input type="checkbox"/>
Р5. Затраты на зарплату сотрудникам	261450	<input type="checkbox"/>
Р6. Процент недовольных клиентов	0	<input type="checkbox"/>
Р7. Процент утерянного груза при погрузке	0	<input type="checkbox"/>
Р8. Затраты на обучение сотрудников	34860	<input type="checkbox"/>
Вес Р1	0.4	<input type="checkbox"/>
Вес Р2	0.3	<input type="checkbox"/>
Вес Р3	0.2	<input type="checkbox"/>
Вес Р4	0.1	<input type="checkbox"/>
Вес Р5	0.4	<input type="checkbox"/>
Вес Р6	0.3	<input type="checkbox"/>

Рис. 32 – Рассчитанные параметры после запуска модели

Объект	Значение	Найти
Вес Р6	0.3	<input type="checkbox"/>
Вес Р7	0.2	<input type="checkbox"/>
Вес Р8	0.1	<input type="checkbox"/>
Доход (суммарный)	3120000	<input type="checkbox"/>
Доход с заявки, выполненной в срок	65000	<input type="checkbox"/>
Интегральный показатель	3.3	<input checked="" type="checkbox"/>
Интенсивность обслуживания начальника порта	0,9	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоИнженера	0,85	<input type="checkbox"/>
ИнтенсивностьОбслуживанияВедущегоТальмана	0,79	<input type="checkbox"/>
ИнтенсивностьОбслуживанияДиректора	0,87	<input type="checkbox"/>
ИнтенсивностьОбслуживанияИнженерногоОтдела	0,89	<input type="checkbox"/>
ИнтенсивностьОбслуживанияОтделаБухгалтерии	0,8	<input type="checkbox"/>
ИнтенсивностьОбслуживанияТальманскогоОтдела	0,71	<input type="checkbox"/>
Количество выполненных заявок	50	<input type="checkbox"/>
Количество заявок, выполненных в срок	46	<input type="checkbox"/>

Рис. 33 – Рассчитанные параметры после запуска модели

Объект	Значение	Найти
Интенсивность Обслуживания Тальманского Отдела	0,71	<input type="checkbox"/>
Количество выполненных заявок	50	<input type="checkbox"/>
Количество заявок, выполненных в срок	46	<input type="checkbox"/>
Количество заявок, выполненных не в срок	4	<input type="checkbox"/>
Количество поступивших заявок	50	<input type="checkbox"/>
ЛО Р1	5	<input type="checkbox"/>
ЛО Р2	5	<input type="checkbox"/>
ЛО Р3	5	<input type="checkbox"/>
ЛО Р4	5	<input type="checkbox"/>
ЛО Р5	2	<input type="checkbox"/>
ЛО Р6	1	<input type="checkbox"/>
ЛО Р7	1	<input type="checkbox"/>
ЛО Р8	4	<input type="checkbox"/>
Основные средства	2500000	<input type="checkbox"/>

Рис. 34 – Рассчитанные параметры после запуска модели

Ошибки | **Консоль**

Шаг № 0
 Отношение: $y = a - b$ (Выражение, когда из одного числа вычитается другое.)
 Правило: Расчет количества выполненных заявок в срок (Расчет количества выполненных заявок в срок)
 Входные параметры:
 Количество выполненных заявок=50;
 Количество заявок, выполненных не в срок=4;
 Формула:
 $y = a - b$
 Результат: Количество заявок, выполненных в срок=46;

Шаг № 1
 Отношение: $y = (a*b) + (c*d)/2$ (Выражение для расчета суммарного дохода (по всем выполненным заявкам))
 Правило: Расчет дохода суммарного (Расчет суммарного дохода (по всем выполненным заявкам))
 Входные параметры:
 Доход с заявки, выполненной в срок=65000;
 Количество заявок, выполненных не в срок=4;
 Количество заявок, выполненных в срок=46;
 Доход с заявки, выполненной в срок=65000;
 Формула:
 $y = (a*b) + (c*d)/2$
 Результат: Доход (суммарный)=3120000;

Шаг № 2
 Отношение: $y = a/b$ (Выражение, показывающее отношение двух величин)
 Правило: Расчет фондоотдачи (Расчет фондоотдачи)

Рис. 35 – Консоль с логическим последовательным решением

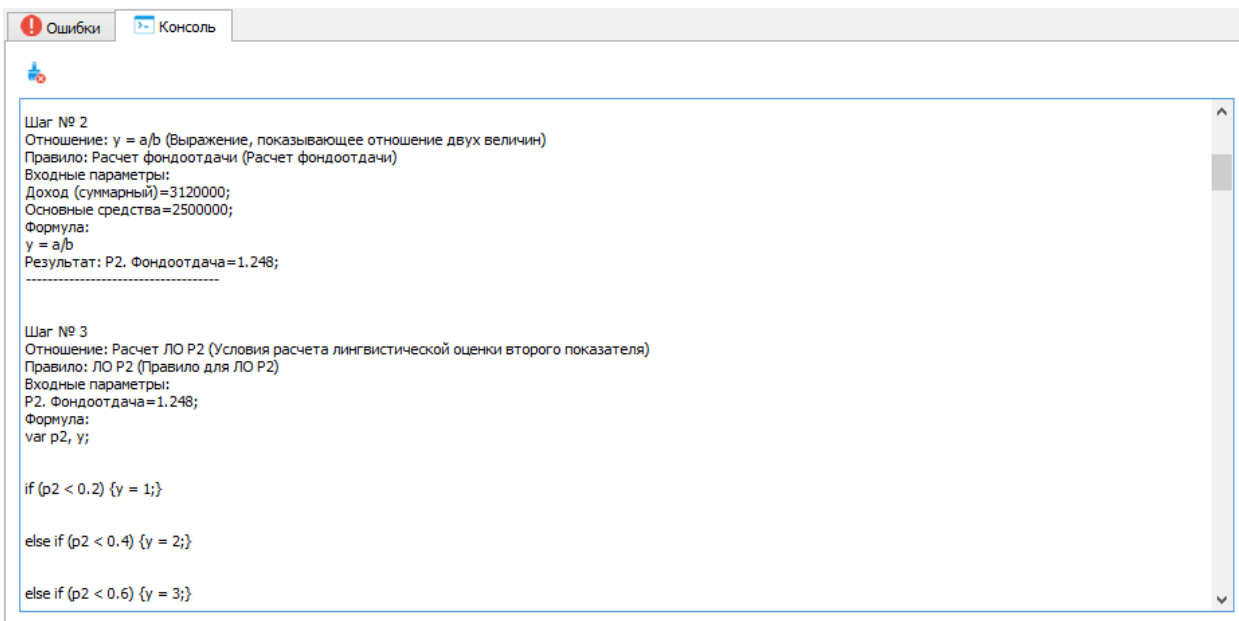


Рис. 36 – Консоль с логическим последовательным решением

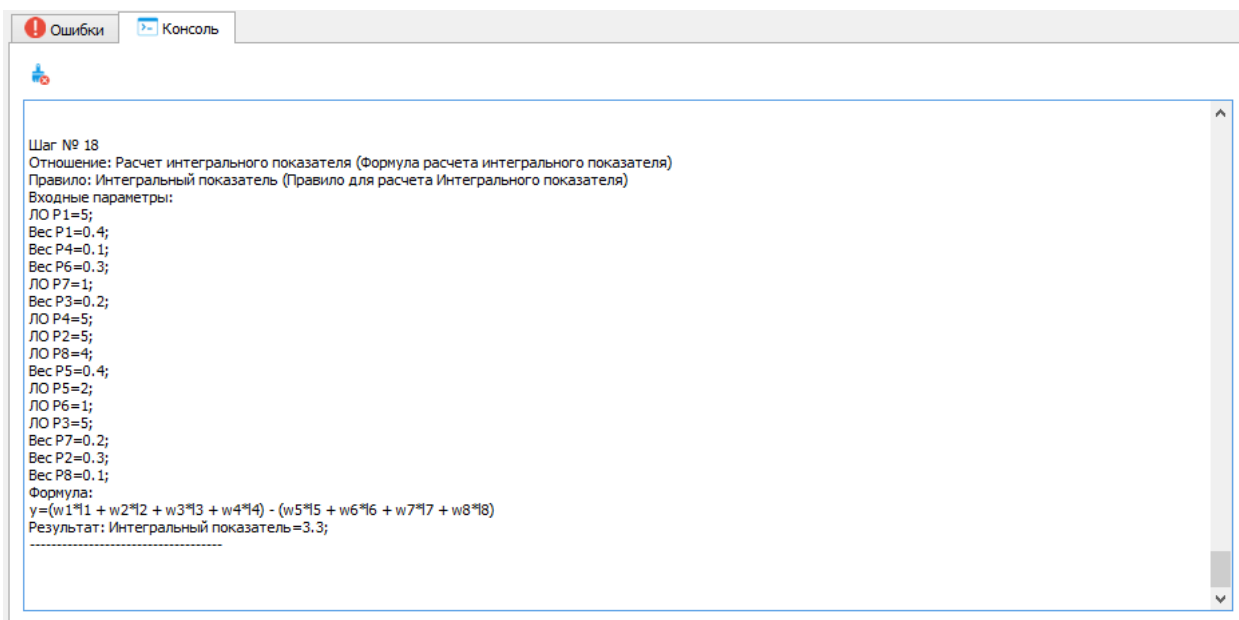


Рис. 37– Консоль с логическим последовательным решением

По рис. 37 видно, что для нахождения интегрального показателя системе потребовалось пройти 19 шагов (нумерация начинается с 0).

Визуальное представление решения

Визуально миварная модель приводится в графовом представлении. Чтобы отобразить граф решения, необходимо выполнить тестирование модели. В случае успешного тестирования необходимо открыть меню «Инструменты» и выбрать пункт «Отобразить граф». Граф решения представлен на рис. 38.

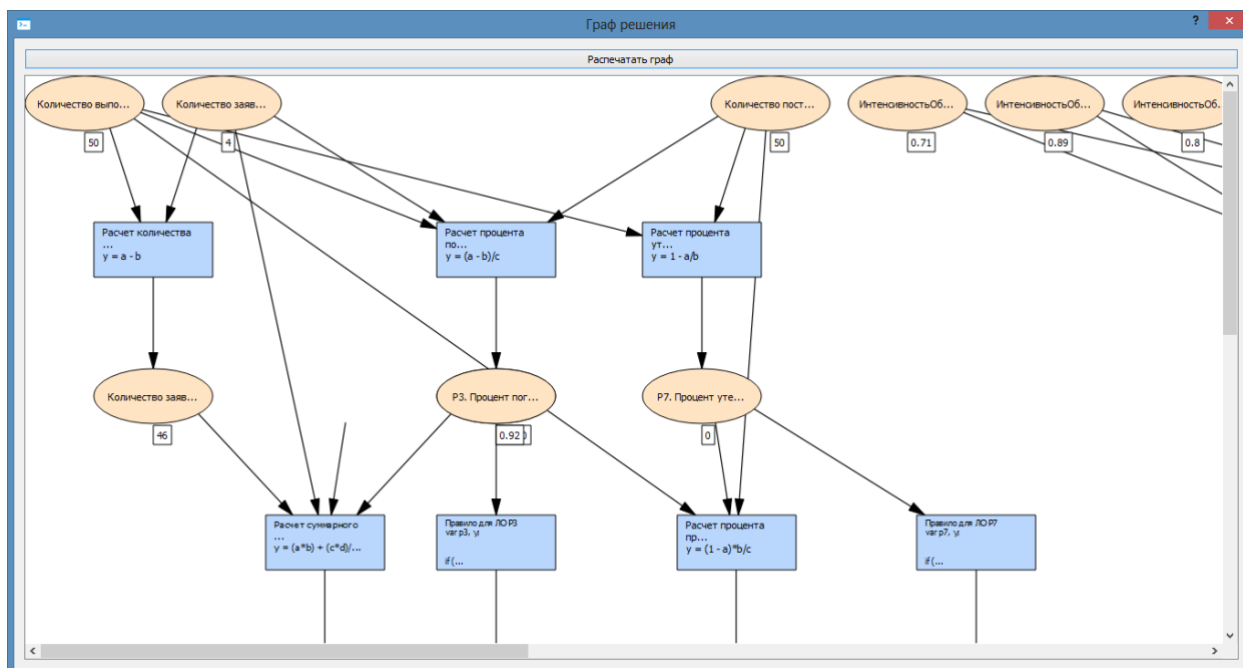


Рис. 38 – Граф решения

Правила обозначаются прямоугольниками, параметры – эллипсами. При щелчке по правилу отображаются все входящие и выходящие параметры, которые соединяются с правилом зелеными и красными стрелками соответственно. Щелчок по параметру показывает все связанные с ним правила и параметры. Около каждого параметра обозначается его текущее значение. Перемещаться по графу можно используя полосы прокрутки и мышь. Для перемещения с помощью мыши, необходимо навести курсор на пустое место на графе, нажать левую кнопку мыши, и, не отпуская кнопку, перемещать курсор. Вместе с движением курсора будет перемещаться и граф.

Если щелкнуть на какое-либо правило двойным щелчком, то будет отображена лишь часть графа, состоящая из данного правила и связанных с ним входных и выходных параметров. Для возврата к обзору полного графа необходимо снова осуществить двойной щелчок по правилу.

Существует возможность вывода графа на печать. Для этого необходимо открыть окно графа и щелкнуть по кнопке «Распечатать граф». В открывшемся диалоговом окне, выбрать нужные параметры. По завершению настройки нажать на кнопку «Печать».

Контрольные вопросы

- 1) Каковы основные функции программного продукта WiMi?
- 2) В чем заключается Технология runtime?
- 3) Области применения программного продукта WiMi.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ
МЕТОДЫ ОДНОМЕРНОГО И МНОГОМЕРНОГО БЕЗУСЛОВНОГО ПОИСКА

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

МЕТОДЫ ОДНОМЕРНОГО И МНОГОМЕРНОГО БЕЗУСЛОВНОГО ПОИСКА:
методические указания для выполнения лабораторных работ по дисциплине
«МЕТОДЫ ОПТИМИЗАЦИИ В БИЗНЕС-АНАЛИТИКЕ» для студентов
направления 09.04.03 «Прикладная информатика» направленность «Искусственный
интеллект и бизнес-аналитика» / Составитель: Т. В. Хоменко. – Ульяновск: УлГТУ,
2021. – 24с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и бизнес-
аналитика». Цель данного практикума – ориентировать студентов на содержание
и порядок выполнения лабораторных работ во время прохождения студентами
курсов «Искусственный интеллект и бизнес-аналитика». Даются задания для
лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание

ВВЕДЕНИЕ	4
Практическая работа №1	5
Тема: Методы одномерного поиска	5
1. Краткие теоретические сведения	5
2. Задание на практическую работу	15
3. Вопросы к практической работе	15
Практическая работа № 2	17
Тема: Методы многомерного безусловного поиска: методы прямого поиска	17
1. Краткие теоретические сведения	17
2. Задание на практическую работу	22
3. Вопросы к практической работе	22
Список литературы	23

ВВЕДЕНИЕ

В учебно-методическом пособии описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.).

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Практическая работа №1

Тема: Методы одномерного поиска

Цель работы: знакомство с оптимизационными задачами, изучение различных методов одномерной оптимизации и сравнение эффективности их применения для конкретных целевых функций.

1. Краткие теоретические сведения

1.1 Постановка задачи одномерной безусловной оптимизации

Задача поиска минимума целевой функции формулируется в виде:

$$x^* = \arg \min f(x), x \in X,$$

где X – множество допустимых точек, среди которых ищется точка x^* , доставляющая минимум $f(x)$ целевой функции.

Другая распространенная запись задачи минимизации

$$\min_{x \in X} f(x)$$

Когда $X=R$, мы имеем дело с одномерной безусловной задачей минимизации, т.е. когда целевая функция $f(x)$ имеет только один простой аргумент и область X есть вся вещественная ось чисел.

В методах одномерной оптимизации вместо $X=R$ рассматривается отрезок $X=[a,b]$, содержащий искомое решение x^* . Такой отрезок называется отрезком неопределенности, или отрезком локализации. Относительно целевой функции $f(x)$ часто предполагается, что она унимодальная.

Определение: Функция $f(x)$ называется унимодальной на $X=[a,b]$, если существует такая точка $x^* \in X$, что:

$$\begin{aligned} f(x_1) > f(x_2), & \text{ если } x_1 < x_2 < x^*, x_1, x_2 \in X, \\ f(x_1) < f(x_2), & \text{ если } x^* < x_1, x_2, x_1, x_2 \in X. \end{aligned}$$

Если ограничиваться рассмотрением лишь непрерывных функций $f(x)$, то свойство унимодальности функции попросту означает наличие у нее единственного локального минимума и этот минимум достигается в точке $x=x^*$.

В ряде методов относительно целевой функции $f(x)$ предполагается, что она выпуклая на X .

Определение: Функция $f(x)$ называется выпуклой на $x=[a,b]$, если $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$ при любых $x_1, x_2 \in X$ и всех $\alpha, 0 \leq \alpha \leq 1$.

Если при любых $x_1, x_2 \in X$ неравенство будет строгим, то функция $f(x)$ называется строго выпуклой.

1.2 Алгоритм пассивного поиска минимума

Пусть N - число точек, в которых необходимо провести вычисления целевой функции $f(x)$, т.е. N экспериментов. Точки, в которых необходимо провести эксперименты, определяются следующим образом:

$$\left\{ \begin{array}{l} \text{Если } N = 2k - 1 - \text{нечётное, то} \\ x_i = a + \frac{b-a}{N+1} \cdot i \\ i = 1, 2, 3, \dots, N \\ \text{Если } N = 2k - \text{чётное, то} \\ x_{2i} = a + \frac{b-a}{k+1} \cdot i \\ x_{2i-1} = x_{2i} - \delta \\ i = 1, 2, \dots, k \end{array} \right.$$

Среди вычисленных значений $\{f(x_i)\}$ ($i=1, N$), ищется точка x_j , в которой достигается минимум:

$$f(x_j) = \min_{1 \leq i \leq N} f(x_i), \text{ где } 1 \leq i \leq N$$

Найденная точка принимается за приближенное решение задачи $\tilde{x} = x_j$. Исходный отрезок неопределенности $[a, b]$ после экспериментов в N точках сужается до $[x_{j-1}, x_{j+1}]$, длина которого равна:

$$L_N = L_N(x_1, x_2, \dots, x_N) = \max_{1 \leq i \leq N} (x_{i+1} - x_{i-1}) = x_{j+1} - x_{j-1} = \begin{cases} 2 \frac{b-a}{N+1}, & \text{если } N = 2k - 1 \\ \frac{b-a}{N/2+1} + \delta, & \text{если } N = 2k \end{cases}$$

Точность найденного решения \tilde{x} равна половине отрезка неопределенности, т.е. $|x^* - \tilde{x}| \leq \varepsilon$, где $\varepsilon = \frac{1}{2} L_N$ и x^* - точное решение.

1.3 Алгоритмы активного поиска минимума

В алгоритмах активного поиска очередная точка, в которой производится эксперимент, выбирается с учетом информации, полученной в предыдущих опытах. Рассмотрение этих алгоритмов начнем с методов блочного поиска, которые сочетают в себе пассивные и последовательные стратегии поиска. При этом вычисления в точках объединены в блоки, в каждом из которых проводится одновременно n_i экспериментов), общее число экспериментов будет $\sum_{i=1}^m n_i = N$, т.е. блок – это совокупность из нескольких экспериментов, которые проводятся одновременно (пассивный поиск). Результаты, полученные в $(i-1)$ -м блоке, становятся известны перед началом экспериментов в i -м блоке $\{x_{ij} (j=1, 2, \dots, n_i)\}$ (последовательный поиск). Если размеры блоков равны единице, т.е. $n_i=1$, то мы имеем обычный последовательный алгоритм поиска.

1.3.1 Алгоритм равномерного блочного поиска

Схема алгоритма

Шаг 1. Задаются исходный отрезок неопределенности $[a, b]$, ε - точность приближенного решения \tilde{x} , число экспериментов в блоке – n (нечетное, $n=2k-1$). Проводим эксперимент в середине отрезка $[a, b]$, т.е. вычисляем $u_k=f(x_k)$, где $x_k=(a+b)/2$.

Шаг 2. Проводим эксперименты в остальных точках блока: $y_i=f(x_i)$, где $x_i=a+i*(b-a)/(n+1)$, $i=1, 2, \dots, n$, $i \neq k$. Находим точку x_j , в которой достигается минимум среди вычисленных значений: $f(x_j)=\min f(x_i)$, следовательно, точное значение минимума x^*

содержится на отрезке $[x_{j-1}, x_{j+1}]$.

Шаг 3. Полагаем $a=x_{j-1}$, $b=x_{j+1}$, $x_k=x_j$, $y_k=y_j$. Если $b-a \leq 2\varepsilon$, то $\tilde{x} = x_k$, $\tilde{y} = y_k$ и поиск заканчивается. Иначе перейти к шагу 2. Если заданная точность ε достигнута после m итераций, т.е. после экспериментов в m блоках, то длина отрезка неопределенности после всех N вычислений ($N=n+(m-1)(n-1)=(n-1)m+1$) будет:

$$L_N = \left(2 \frac{b-a}{n+1}\right)^m \text{ и } |x^* - \tilde{x}| \leq \frac{1}{2} L_N.$$

1.3.2 Алгоритм деления интервала пополам

Это вариант предыдущего алгоритма при $n=3$.

Схема алгоритма.

Шаг 1. Задаются a, b, ε . Производим эксперимент в точке $x_2=(a+b)/2$, т.е. вычисляем $y_2=f(x_2)$.

Шаг 2. Проводим эксперименты в остальных точках блока: $x_1=(a+x_2)/2$, $y_1=f(x_1)$, $x_3=(x_2+b)/2$, $y_3=f(x_3)$.

Находим x_j такую, что $f(x_j)=\min \{f(x_i)\}$, где $1 \leq i \leq 3$

Тогда точное решение x^* содержится на отрезке $[x_{j-1}, x_{j+1}]$. Предполагается $x_0 = a, x_4 = b$.

Шаг 3. Полагаем $a=x_{j-1}$, $b=x_{j+1}$, $x_2=x_j$, $y_2=y_j$. Если $b-a \leq 2\varepsilon$, то $\tilde{x} = x_2$, $\tilde{y} = y_2$ и поиск заканчивается. Иначе перейти к шагу 2.

После k итераций общее число проведенных экспериментов равно $N=2k+1$, а длина получившегося отрезка неопределенности будет $L_N = \frac{b-a}{2^k} = \frac{b-a}{2^{\lfloor N/2 \rfloor}}$, где $\lfloor z \rfloor$ – целая часть числа z . Следовательно, достигнутая точность будет $|x^* - \tilde{x}| \leq \varepsilon$, $\varepsilon = 1/2L_N$.

1.3.3 Метод дихотомии

Это алгоритм блочного поиска для $n_i=n=2$, т.е. когда в блоке два эксперимента. Так как пассивная составляющая алгоритма, т.е. блок, содержит четное число экспериментов, то оптимальный выбор точек x_{ij} , в которых необходимо провести эксперименты, будет неравномерным, в отличие от предыдущих алгоритмов, где число экспериментов в блоке было нечетным и, соответственно, расположение точек равномерным. Если блок содержит два эксперимента, то оптимальное (дельта оптимальное) расположение точек, в которых будут проводиться эксперименты, это как можно ближе к середине отрезка. Такое расположение точек позволяет получить наименьший отрезок неопределенностей после экспериментов в блоке.

Схема алгоритма.

Шаг 1. Задаются a, b, ε и δ – малое положительное число, значительно меньшее чем ε .

Шаг 2. Определяется середина отрезка $x=(a+b)/2$. Производятся эксперименты в двух точках близких середине: $y_1=f(x-\delta)$, $y_2=f(x+\delta)$.

Шаг 3. Определяется следующий отрезок локализации, т.е. определяется какой из отрезков $[x-\delta, b]$ или $[a, x+\delta]$ содержит точное решение x^* . Если $y_1 \leq y_2$, то это отрезок $[a, x+\delta]$ и $b=x+\delta$, иначе это отрезок $[x-\delta, b]$ и $a=x-\delta$, т.е. выбранный отрезок локализации мы снова обозначили как $[a, b]$.

Шаг 4. Если $b-a \leq 2\varepsilon$, то $x=(a+b)/2$, $\tilde{y} = f(\tilde{x})$ и поиск заканчивается. Иначе перейти к

шагу 2.

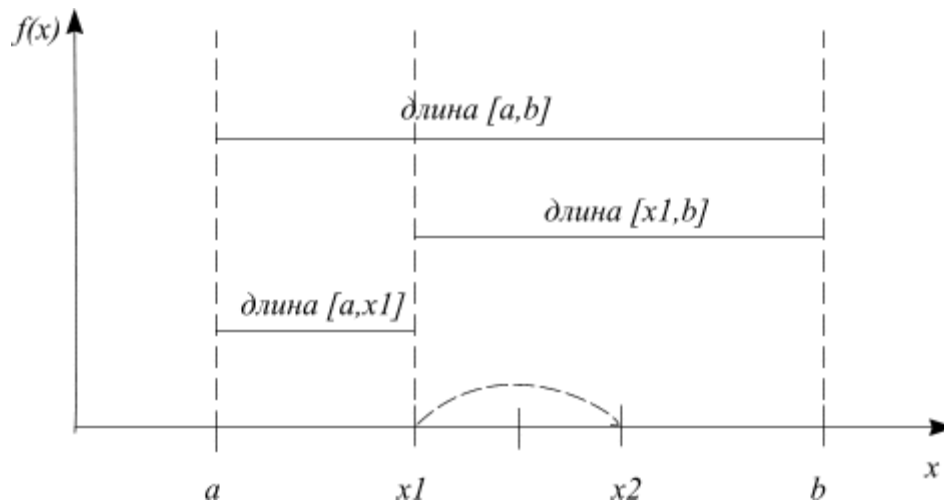
После k итераций общее число экспериментов будет $N=2k$, а длина получившегося отрезка неопределённости $L_N < \frac{b-a}{2^{N/2}} + \delta$. Следовательно, $|x^* - \tilde{x}| < \frac{1}{2} L_N$.

А теперь перейдём к рассмотрению чисто последовательных методов поиска.

1.3.4 Метод золотого сечения

Для того чтобы уменьшить отрезок неопределённости $[a, b]$, нам необходимо вычислить значение целевой функции $f(x)$, по крайней мере, в двух точках на отрезке $[a, b]$.

В результате этих двух экспериментов отрезок неопределённости сузится до отрезка $[a, x_2]$ или $[x_1, b]$.



Так как у нас нет никаких оснований предпочесть один из этих вариантов, то точки x_1 и x_2 должны быть симметричны относительно середины отрезка $[a, b]$. В этом случае длины отрезков $[a, x_2]$ и $[x_1, b]$ будут равны. Таким образом, остаётся вопрос, как выбрать точку x_1 .

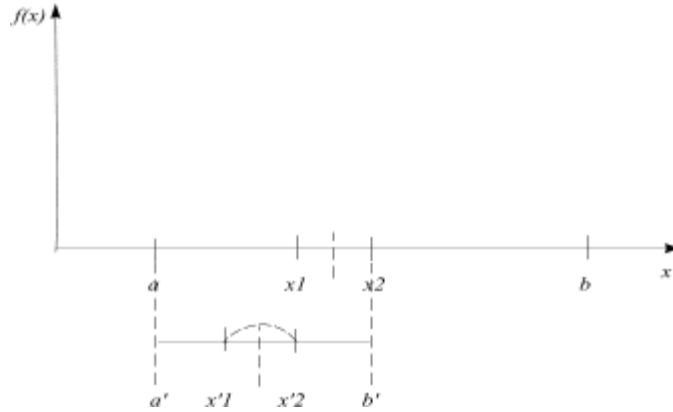
В методе золотого сечения точка x_1 выбирается из соображения, что должно выполняться соотношение:

$$\frac{\text{длина}[a, b]}{\text{длина}[x_1, b]} = \frac{\text{длина}[x_1, b]}{\text{длина}[a, x_1]} = \lambda = 1.618033989\dots$$

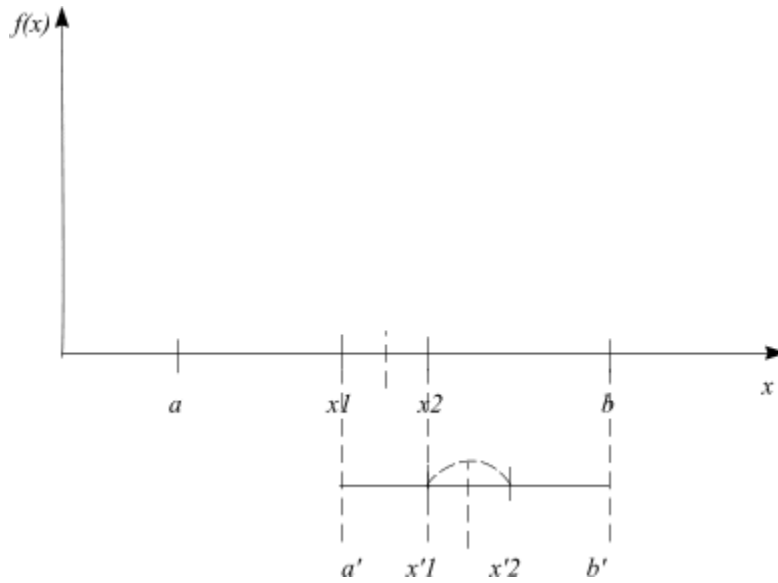
т.е. точка x_1 делит отрезок $[a, b]$ по правилу «золотого сечения», где λ - есть «золотое отношение». Точка x_2 определяется как точка симметричная к x_1 относительно середины отрезка.

В результате экспериментов у нас получается отрезок неопределённости $[a, x_2]$, содержащий точку x_1 , или отрезок неопределённости $[x_1, b]$, содержащий точку x_2 . Оказывается, что остающаяся точка на суженном отрезке неопределённости делит его вновь по правилу «золотого сечения». Следовательно, чтобы, в свою очередь, уменьшить новый отрезок неопределённости, нам не достаёт одного эксперимента, а именно, вычисления целевой функции в точке, симметричной к оставшейся точке относительно середины этого нового отрезка. Всё продемонстрировано на рисунке,

а)



б)



где буквы со штрихами обозначают новый отрезок неопределённости. Вариант а) соответствует случаю, если новым отрезком неопределённости будет $[a, x_2]$, а вариант б) – отрезку $[x_1, b]$.

В приводимой ниже схеме алгоритма остающиеся отрезки неопределённости переименовываются каждый раз как $[a, b]$, а точки, в которых проводятся эксперименты на этом отрезке, обозначается через x_1 и x_2 , причём $x_1 < x_2$. Кроме того, y_1 и y_2 имеют следующие значения: $y_1 = f(x_1)$ и $y_2 = f(x_2)$.

Схема алгоритма

Шаг1. Задаются a, b, ε и $\lambda = 1.618\dots$. Вычисляют

$$x_1 = b - \frac{b-a}{\lambda}, x_2 = a + \frac{b-a}{\lambda}, y_1 = f(x_1), y_2 = f(x_2).$$

Шаг2. а) Если $y_1 \leq y_2$, то полагают $b = x_2, x_2 = x_1, y_2 = y_1$ и вычисляют $x_1 = a + b - x_2, y_1 = f(x_1)$.

б) Если $y_1 > y_2$, то полагают $a = x_1, x_1 = x_2, y_1 = y_2$ и вычисляют $x_2 = a + b - x_1, y_2 = f(x_2)$.

Шаг3. Если $b - a > \varepsilon$, то переходят к шагу 2. Иначе если $y_1 < y_2$, то полагают $\tilde{x} = x_1$

и $\tilde{y} = y_1$, если $y_1 \geq y_2$, то полагают $\tilde{x} = x_2$ и $\tilde{y} = y_2$.

Закончить поиск.

После каждой итерации длина отрезка неопределённости уменьшается в λ раз. Так как первая итерация начинается после двух экспериментов, то после N экспериментов длина отрезка неопределённости будет

$$L_N = \frac{b-a}{\lambda^{N-1}}.$$

1.3.4 Метод чисел Фибоначчи

Этот метод применяется, когда число экспериментов N заранее задано. Метод чисел Фибоначчи, также как и метод золотого сечения относится к симметричным методам, т.е. точки, в которых выполняются два эксперимента, на основе которых происходит уменьшение отрезка неопределённости, расположены симметрично относительно середины отрезка. Вот только выбор точки x_1 происходит на основе других соотношений. Для этого используются числа Фибоначчи: $F_0, F_1, F_2, F_3, \dots$, где $F_i = F_{i-2} + F_{i-1}$ ($i = 2, 3, \dots$) и $F_0 = F_1 = 1$.

Точка x_1 определяется из соотношения:

$$\frac{\text{длина}[a, x_1]}{\text{длина}[a, b]} = \frac{F_{N-2}}{F_N},$$

т.е. $x_1 = a + (b-a) \frac{F_{N-2}}{F_N}$. Точка x_1 делит $[a, b]$ на две неравные части. Отношение малого отрезка к большему равно F_{N-2}/F_{N-1} . Точка x_2 определяется как точка, симметричная к x_1 относительно середины отрезка $[a, b]$. Поэтому $x_2 = b - (b-a) \frac{F_{N-2}}{F_N} = a + (b-a) \frac{F_{N-1}}{F_N}$. При этом будет выполняться условие $x_1 < x_2$.

В результате экспериментов в точках x_1 и x_2 у нас получится отрезок неопределённости $[a, x_2]$, содержащий точку x_1 , или отрезок неопределённости $[x_1, b]$, содержащий точку x_2 . Остающаяся точка делит новый отрезок неопределённости на две неравные части в отношении:

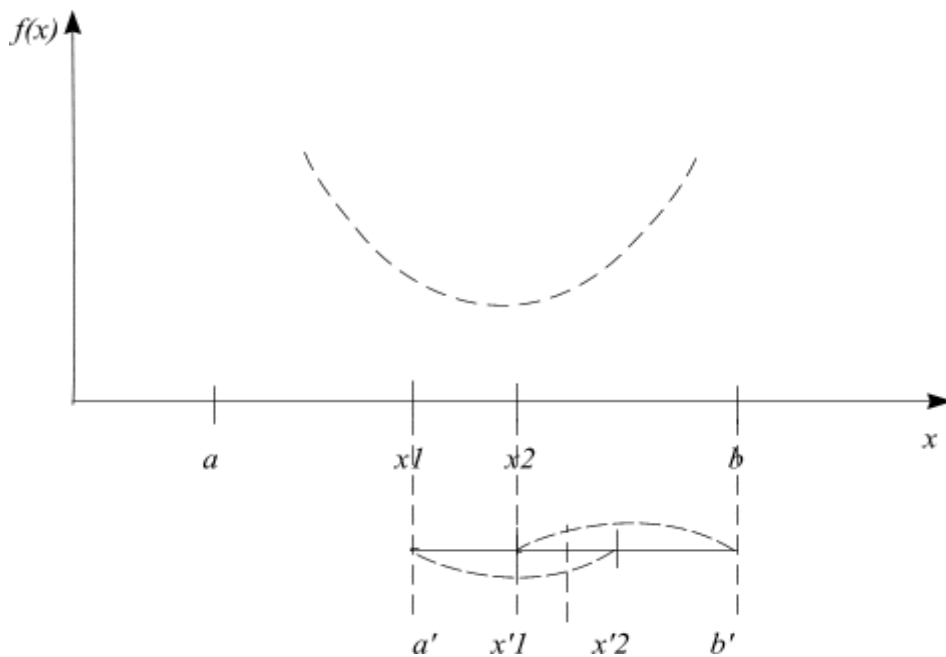
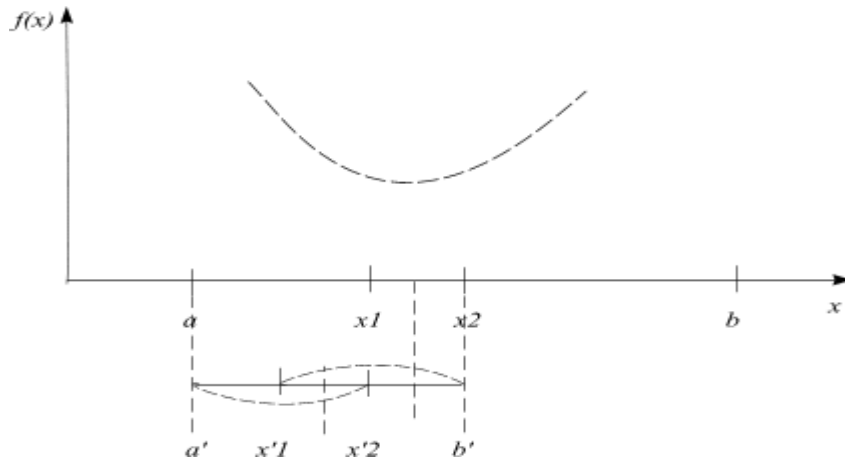
$$\frac{\text{меньшая_часть}}{\text{большая_часть}} = F_{N-3} / F_{N-2}.$$

То есть в методе Фибоначчи остающаяся точка делит отрезок на две неравные части в пропорциях определяемых числами Фибоначчи. Так на k -ом шаге это отношение равно

$$\frac{\text{меньшая_часть}}{\text{большая_часть}} = \frac{F_{N-k-2}}{F_{N-k}}, \text{ а длины отрезков равны: } \text{меньшая_часть} = \frac{F_{N-k-1}}{F_N} (b-a) \text{ и}$$

$$\text{большая_часть} = \frac{F_{N-k}}{F_N} (b-a). \text{ Всё это показано на рисунке:}$$

а)



б)

$$\frac{\text{длина}[a, x_1]}{\text{длина}[x_1, b]} = \frac{F_{N-2}}{F_{N-1}}$$

$$\frac{\text{длина}[x'_2, b']}{\text{длина}[a', x'_2]} = \frac{F_{N-3}}{F_{N-2}}$$

$$\frac{\text{длина}[a', x'_1]}{\text{длина}[x'_1, b']} = \frac{F_{N-3}}{F_{N-2}}$$

Для того чтобы в свою очередь уменьшить получившийся отрезок неопределённости, надо определить симметричную точку относительно середины отрезка и произвести эксперимент в ней. Этот процесс продолжается, пока не будет проведено N экспериментов.

Схема алгоритма

Шаг 1. Задаются a, b, N . Вычисляются числа Фибоначчи F_0, F_1, \dots, F_N . Определяется:

$$x_1 = a + (b-a)F_{N-2}/F_N,$$

$$x_2 = a + (b-a)F_{N-1}/F_N,$$

$$y_1 = f(x_1), y_2 = f(x_2).$$

Шаг 2. а) Если $y_1 \leq y_2$, то полагают $b = x_2, x_2 = x_1, y_2 = y_1$ и вычисляют $x_1 = a + b - x_2, y_1 = f(x_1)$.

б) Если $y_1 > y_2$, то полагают $a = x_1, x_1 = x_2, y_1 = y_2$ и вычисляют $x_2 = a + b - x_1, y_2 = f(x_2)$.

Повторить шаг 2 $N-2$ раза.

Шаг 3. Если $y_1 < y_2$, то полагают $\tilde{x} = x_1$ и $\tilde{y} = y_1$. Если $y_1 \geq y_2$, то полагают $\tilde{x} = x_2$ и $\tilde{y} = y_2$.

Закончить поиск.

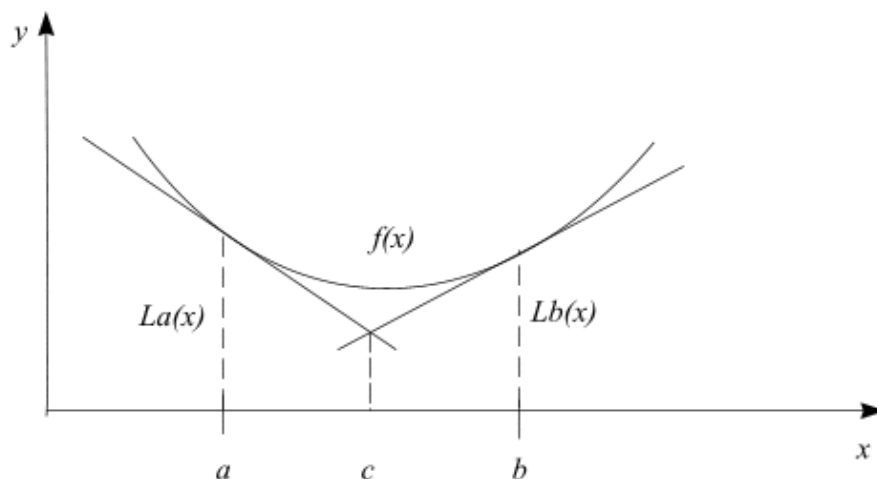
Длина отрезка неопределённости в методе Фибоначчи $L_N \approx (b-a)/F_N$.

1.4 Методы поиска, основанные на аппроксимации целевой функции

Суть этих методов заключается в том, что по полученной в ходе вычислений информации строится аппроксимирующая функция и её минимум принимается за точку очередного вычисления. Такие методы дают хорошие результаты при минимизации достаточно гладких унимодальных функций.

1.4.1 Метод касательных

Пусть функция $f(x)$ выпукла и дифференцируема на $[a, b]$. Идея метода состоит в следующем. Пусть $[a, b]$ - отрезок неопределённости и $f(a), f'(a), f(b), f'(b)$ - результаты вычислений в точках a и b . По этой информации строится аппроксимирующая функция, представляющую из себя кусочно-линейную функцию, состоящую из касательной $L_a(x) = f(a) + f'(a)(x-a)$ к $f(x)$ в точке a и касательной $L_b(x) = f(b) + f'(b)(x-b)$ к $f(x)$ в точке b .



Полученная аппроксимирующая функция есть ломанная состоящая из прямой $L_a(x)$ на $[a, c]$ и $L_b(x)$ на $[c, b]$, где c – точка пересечения касательных. Легко заметить, что при $f(a) < 0$ и $f(b) > 0$ минимум аппроксимирующей функции достигается в точке c . Значение точки пересечения c можно определить по формуле

$$c = \frac{(bf'(b) - af'(a)) - (f(b) - f(a))}{f'(b) - f'(a)}$$

В точке c производятся вычисления $f(c)$ и $f'(c)$. Если $f'(c) = 0$, то решением задачи будет $x^* = c$. Если же $f'(c) > 0$, то в качестве следующего отрезка неопределённости будет $[a, c]$. Если же $f'(c) < 0$, то – отрезок $[c, b]$. Процесс повторяется до тех пор, пока $f'(c) = 0$ или отрезок неопределённости не достигнет заданной точности.

Схема алгоритма

Шаг 1. Заданы a, b, ε . Вычислить $y_1 = f(a), y_2 = f(b), z_1 = f'(a), z_2 = f'(b)$.

Шаг 2. Если $b - a \leq 2\varepsilon$, то полагаем $\tilde{x} = (a + b)/2, \tilde{y} = f(\tilde{x})$. Поиск окончен. Если $b - a > 2\varepsilon$, то вычислить $c = \frac{(bz_2 - az_1) - (y_2 - y_1)}{z_2 - z_1}, y = f(c), z = f'(c)$. Если $z = 0$, то полагаем $\tilde{x} = c, \tilde{y} = y$ и поиск окончен. Если $z < 0$, то $a = c, y_1 = y, z_1 = z$. Если $z > 0$, то $b = c, y_2 = y, z_2 = z$. Повторить шаг 2.

1.4.2 Метод парабол

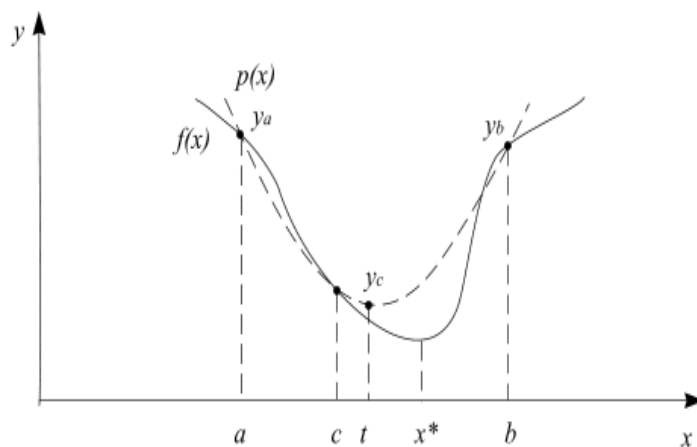
Рассмотрим алгоритм квадратичной интерполяции или метод парабол, т.е. в качестве аппроксимирующей функции используется парабола. Для однозначного задания параболы необходимы три точки. Пусть имеются три точки, для которых выполняется $a < c < b, f(c) \leq f(a), f(c) \leq f(b)$. Так как $[a, b]$ – отрезок неопределённости и $f(x)$ – унимодальная функция, то найти такую точку c нетрудно. Парабола, проходящая через три точки $(a, f(a)), (c, f(c)), (b, f(b))$, имеет вид

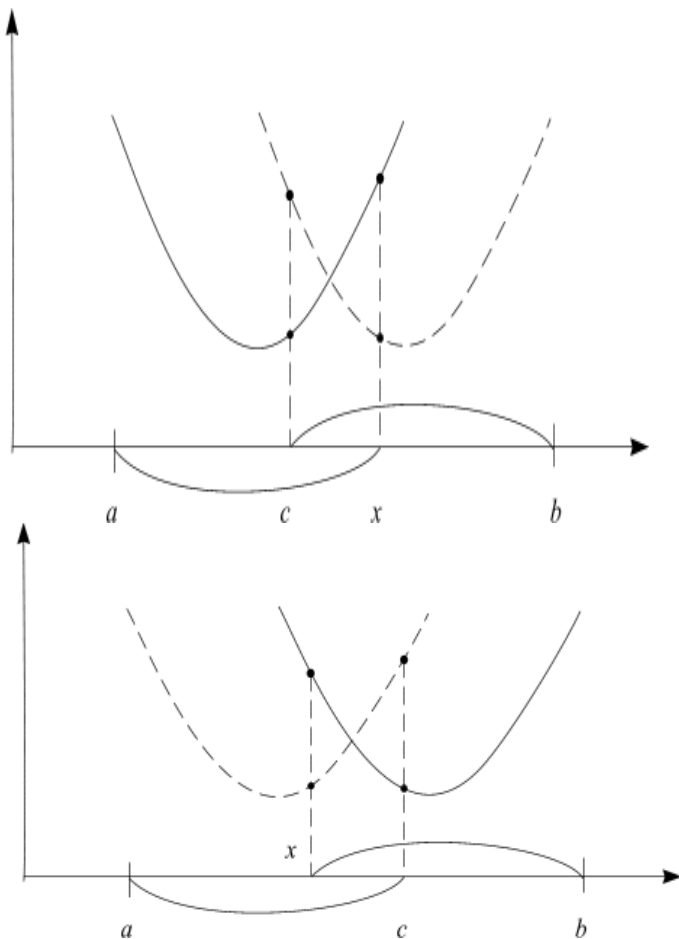
$$P(x) = \left(\frac{f(b) - f(c)}{b - c} + \frac{f(a) - f(c)}{a - c} \right) \frac{(x - c)(x - b)}{b - a} + \frac{f(b) - f(c)}{b - c} (x - c) + f(c)$$

Поскольку f – унимодальная функция, то хотя бы одно из неравенств $f(c) \leq f(a), f(c) \leq f(b)$ строгое и, следовательно, коэффициент при старшем члене $P(x)$ положителен. Тогда $P(x)$ достигает минимума в точке

$$t = c + \frac{1}{2} \frac{(b - c)^2 (f(a) - f(c)) - (c - a)^2 (f(b) - f(c))}{(b - c)(f(a) - f(c)) + (c - a)(f(b) - f(c))},$$

причем $(a + c)/2 \leq t \leq (c + b)/2$. Эта точка и выбирается в качестве точки очередного вычисления значения функции.





Если оказалось, что $t=c$, условимся в качестве точки очередного вычисления выбирать точку $(a+c)/2$. Итак, следующее вычисление проводится в точке

$$x = \begin{cases} t, & \text{если } t \neq c \\ (a+c)/2, & \text{если } t = c. \end{cases}$$

Определим новый отрезок неопределенности с лежащей внутри него точкой, для которой выполняются условия аналогичные условиям, которым удовлетворяла точка c . В силу унимодальности функции f и в зависимости от выполнения или невыполнения условий $x < c$, $f(x) < f(c)$, $f(x) = f(c)$ это будут отрезки с точкой внутри: $[a, c]$ и x ; $[x, b]$ и c ; $[x, c]$ и $(x+c)/2$; $[a, x]$ и c ; $[c, b]$ и x ; $[c, x]$ и $(x+c)/2$. Смотри рисунок. Далее строится парабола, определяется ее минимум, и т.д., до тех пор, пока длина отрезка неопределенности не удовлетворяет заданной точности.

Схема алгоритма

Шаг 1. Задаются a, c, b и ε . Вычислить $y_a = f(a)$, $y_c = f(c)$, $y_b = f(b)$.

Шаг 2. Вычислить $x = \begin{cases} t, & \text{если } t \neq c \\ (a+c)/2, & \text{если } t = c. \end{cases}$, $y = f(x)$, где

$$t = c + \frac{1}{2} \cdot \frac{(b-c)^2(y_a - y_c) - (c-a)^2(y_b - y_c)}{(b-c)(y_a - y_c) + (c-a)(y_b - y_c)}.$$

Шаг 3. А) При $x < c$.

Если $y < y_c$, то $b=c$, $c=x$, $y_b=y_c$, $y_c=y$.

Если $y > y_c$, то $a = x$, $y_a = y$.

Если $y = y_c$, то $a = x$, $b = c$, $c = (x+c)/2$, $y_a = y$, $y_b = y_c$, $y_c = f(c)$.

Б) При $x < c$.

Если $y < y_c$, то $a = c$, $c = x$, $y_a = y_c$, $y_c = y$.

Если $y > y_c$, то $b = x$, $y_b = y$.

Если $y = y_c$, то $a = c$, $b = x$, $c = (x+c)/2$, $y_a = y_c$, $y_b = y$, $y_c = f(c)$.

Шаг 4. Если $b - a \leq \varepsilon$, то закончить поиск, положив $\tilde{x} = x$, $\tilde{y} = y$, иначе перейти к шагу 2.

2. Задание на практическую работу

2.1. Изучить предлагаемые методы одномерной безусловной оптимизации.

2.2. В соответствии с вариантом задания, определенным преподавателем, найти точку минимума функции $f(x)$ на отрезке $[a, b]$ одним из рассмотренных методов (составить программу, реализующую метод поиска).

№	Целевая функция	Отрезок $[a, b]$	Точность ε или число экспериментов N
1	$x^2 + 6 \cdot e^{0,15x}$	$[-1, 0]$	$N=22$
2	$x^2 + 4 \cdot e^{-0,25x}$	$[0, 1]$	$N=23$
3	$x^4 + 0,4 \cdot \arctg 5x$	$[-1, 0]$	$N=20$
4	$x^4 - 1,5 \arctg x$	$[0, 1]$	$N=21$
5	$x^2 + 8 \cdot e^{0,55x}$	$[-2, 0]$	$\varepsilon=10^{-3}$
6	$-4x + e^{ x-0,2 }$	$[0, 2]$	$\varepsilon=1,5 \cdot 10^{-3}$
7	$1,4x + e^{ x-2 }$	$[0, 2]$	$\varepsilon=5 \cdot 10^{-3}$
8	$x^2 + e^x$	$[-1, 0]$	$N=18$
9	$ x + e^{10x}$	$[-1, 0]$	$\varepsilon=10^{-3}$
10	$10 \cos x + e^x$	$[0, 3]$	$\varepsilon=5 \cdot 10^{-4}$

а) Если в варианте задания указано число экспериментов N , то сравнить заданные в варианте методы по получаемой длине отрезка неопределенности.

б) Если же указана точность искомого решения, то сравнить методы по количеству экспериментов, понадобившихся для достижения заданной точности.

2.3. Оформить отчет о выполнении задания с приведением условия задачи (алгоритмов и программ указанных методов поиска), таблицы результатов сравнения рассмотренных методов, заключения по результатам сравнения методов.

3. Вопросы к практической работе

1) Приведите математическую модель задачи безусловной оптимизации функции одной переменной.

2) Сформулируйте необходимые условия оптимальности в задачах безусловной оптимизации функции одной переменной.

3) Представьте алгоритм классического метода в задачах безусловной оптимизации функции одной переменной.

4) Охарактеризуйте сущность прямых методов минимизации функции одной переменной.

5) Представьте характеристику методов исключения отрезков.

6) Дайте характеристику метода деления отрезка пополам (дихотомии).

7) Представьте алгоритм метода деления отрезка пополам (дихотомии).

- 8) Дайте характеристику метода золотого сечения.
- 9) Представьте алгоритм метода золотого сечения.
- 10) Перечислите методы безусловной минимизации функции одной переменной с использованием производных.
- 11) Дайте характеристику метода касательных.
- 12) Приведите алгоритм метода касательных.
- 13) Какие существуют достоинства и недостатки метода касательных?

Практическая работа № 2

Тема: Методы многомерного безусловного поиска: методы прямого поиска

Цель работы: знакомство с методами многомерной безусловной оптимизации нулевого порядка и их освоение, сравнение эффективности применения этих методов конкретных целевых функций.

1. Краткие теоретические сведения.

1.1 О численных методах многомерной оптимизации

Задача многомерной безусловной оптимизации формулируется в виде:

$$\min_{x \in X} f(x),$$

где $x = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ – точка в n -мерном пространстве $X = \mathbb{R}^n$, то есть целевая функция $f(x) = f(x^{(1)}, \dots, x^{(n)})$ – функция n аргументов.

Так же как и в первой лабораторной работе мы рассматриваем задачу минимизации. Численные методы отыскания минимума, как правило, состоят в построении последовательности точек $\{x_k\}$, удовлетворяющих условию $f(x_0) > f(x_1) > \dots > f(x_n)$. Методы построения таких последовательностей называются методами спуска. В этих методах точки последовательности $\{x_k\}$ вычисляются по формуле:

$$x_{k+1} = x_k + \alpha_k r_k, \quad k=0, 1, 2, \dots,$$

где r_k – направление спуска, α_k – длина шага в этом направлении.

Различные методы спуска отличаются друг от друга способами выбора направления спуска r_k и длины шага α_k вдоль этого направления.

1.2 Общая характеристика

Методы прямого поиска – это методы, в которых используются только значения целевой функции (методы нулевого порядка). Рассмотрим следующие методы, основанные на эвристических соображениях. Эти методы довольно часто применяются на практике, позволяя в ряде случаев получить удовлетворительные решения.

Основное достоинство методов нулевого порядка состоит в том, что они не требуют непрерывности целевой функции и существования производных.

1.3 Метод конфигураций (метод Хука и Дживса)

Алгоритм включает в себя два основных этапа поиска.

а) В начале обследуется окрестность выбранной точки (базисной точки), в результате находится приемлемое направление спуска;

б) Затем в этом направлении находится точка с наименьшим значением целевой функции. Таким образом находится новая базисная точка.

Эта процедура продолжается пока в окрестностях базисных точек удается находить приемлемые направления спуска.

Схема алгоритма

Шаг 1.

Задаются начальное приближение (первая базисная точка):

$$x_0 = \{x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)}\}$$

начальный шаг h для поиска направления спуска, точность решения δ (предельное значение для шага h). Присваивается $k=0$.

Шаг 2. (Первый этап).

Определяется направление минимизации целевой функции $f(x)=f(x^{(1)},x^{(2)},\dots,x^{(n)})$ в базисной точке

$$x_0 = \{x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)}\}$$

Для этого последовательно дают приращение переменным $x^{(j)}$ в точке x_k . Присвоим $z=x_k$. Циклически даем приращение переменным $x^{(j)}$ и формируем $z^{(j)}=x_k^{(j)}+h$, если $f(z)<f(x_k)$, если же нет, то $z^{(j)}=x_k^{(j)}-h$, если $f(z)<f(x_k)$, иначе $z^{(j)}=x_k^{(j)}$. Так для всех $j(j=1,2,\dots,n)$.

Шаг 3.

Если $z=x_k$, то есть не определилось подходящее направление, то обследование окрестности базисной точки x_k повторяется, но с меньшим шагом h (например, $h=h/2$).

Если $h>\delta$, то перейти к шагу 2, то есть повторить обследование точки x_k .

Если $h\leq\delta$, то поиск заканчивается, то есть достигнуто предельное значение для шага h и найти приемлемое направление спуска не удастся. В этом случае полагаются:

$$\tilde{x} = x_k, \tilde{y} = f(x_k)$$

Шаг 4. (Второй этап).

Если $z\neq x_k$, то требуется найти новую базисную точку в направлении вектора:

$$z-x_k: x_{k+1}=x_k + \lambda(z-x_k), \text{ где } \lambda - \text{коэффициент «ускорения поиска»}.$$

Определяется такое значение $\lambda=\lambda_k$, при котором достигается наименьшее значение целевой функции в выбранном направлении, то есть функции:

$$f(x_k + \lambda(z-x_k)) = \varphi(\lambda).$$

В зависимости от способа выбора λ_k возможны варианты метода:

- а) $\lambda_k=\lambda=\text{const}$ постоянная для всех итераций;
- б) задается начальное $\lambda_0=\lambda$, а далее $\lambda_k=\lambda_{k-1}$, если $f(x_{k+1})<f(x_k)$, иначе дробим λ_k , пока не выполнится это условие;
- в) λ_k определяется решением задачи одномерной минимизации функции $\varphi(\lambda)$.

Таким образом, определяется новая базисная точка $x_{k+1}=x_k + \lambda(z-x_k)$. Полагаем $k=k+1$ и поиск оптимального решения повторяется с шага 2.

1.4 Метод симплекса

Под симплексом понимается n -мерный выпуклый многогранник n -мерного пространства, имеющий $n+1$ вершину. Для $n=2$ это треугольник, а при $n=3$ это тетраэдр.

Идея метода состоит в сравнении значений функции в $n+1$ вершинах симплекса и перемещении симплекса в направлении лучшей точки. В рассматриваемом методе симплекс перемещается с помощью операций отражения. Далее принято следующее: $x_0(k), x_1(k), \dots, x_n(k)$ – вершины симплекса, где k – номер итерации.

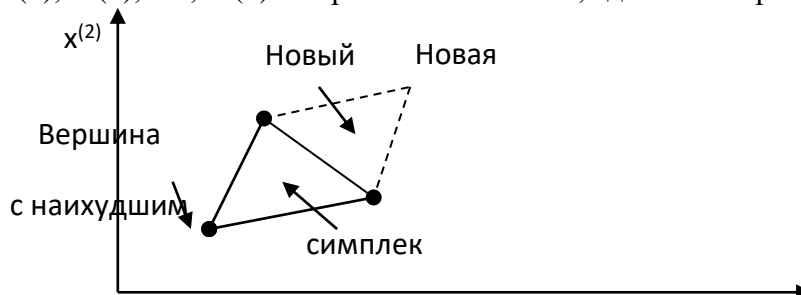


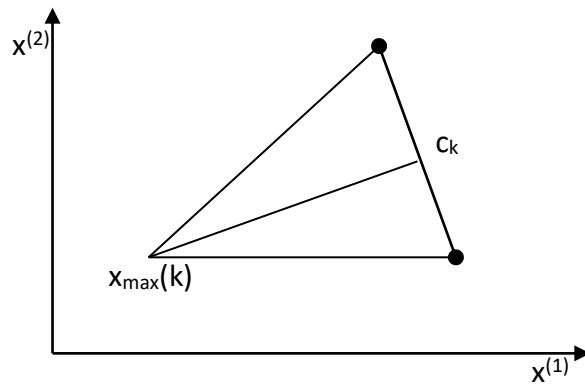
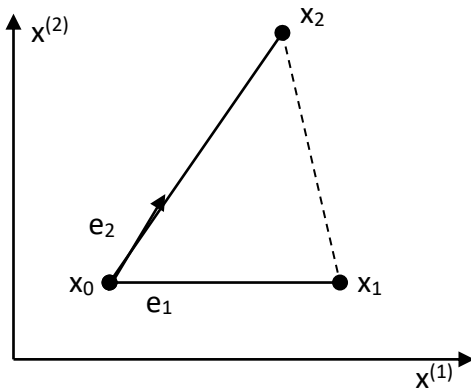
Схема алгоритма

Шаг 1.

Построение начального симплекса.

Для этого задаются начальная точка $x_0(0)$ и длина ребра симплекса l . Формируются остальные вершины симплекса:

$$x_i(0) = x_0(0) + l \cdot e_i \quad (i=1,2,\dots,n), \text{ где } e_i - \text{единичные векторы}.$$



Шаг 2.

Определение направления улучшения решения.

Для этого на k -й итерации вычисляются значения целевой функции в каждой точке симплекса. Пусть для всех i :

$$f(x_{\min}(k)) \leq f(x_i(k)) \leq f(x_{\max}(k)),$$

где \min , \max , i – номера соответствующих вершин симплекса. Определим центр тяжести всех точек, исключая точку $x_{\max}(k)$,

$$C_k = (\sum x_i(k)) / n.$$

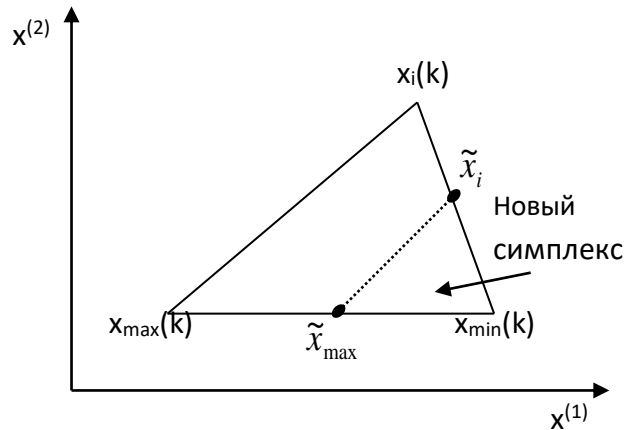
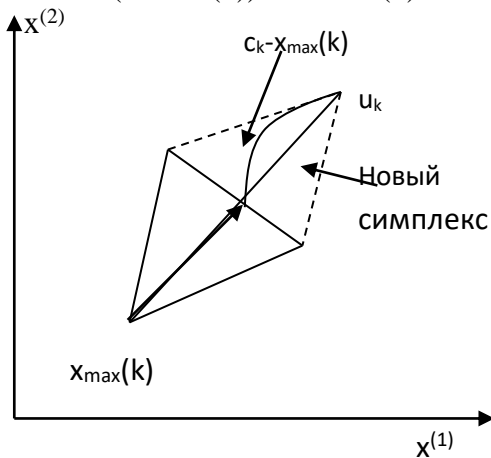
Тогда направление улучшения решения определяется вектором $C_k - x_{\max}(k)$.

Шаг 3.

Построение отраженной точки.

Замена вершины $x_{\max}(k)$ с максимальным значением целевой функции на новую точку с помощью операции отражения, результатом которой является новая точка:

$$u_k = C_k + (C_k - x_{\max}(k)) = 2C_k - x_{\max}(k)$$



Шаг 4.

Построение нового симплекса.

Вычисляем $f(u_k)$. При этом возможен один из двух случаев:

а) $f(u_k) < f(x_{\max}(k))$;

б) $f(u_k) \geq f(x_{\max}(k))$.

Случай а): вершина x_{\max} заменяется на u_k , чем определяется набор вершин $k+1$ -й итерации и k -я итерация заканчивается.

Случай б): в результате отражения получается новая точка u_k , значение функции в которой еще хуже, чем в точке x_{\max} , то есть отражать симплекс некуда. Поэтому в этом случае производится пропорциональное уменьшение симплекса (например, в 2 раза) в сторону вершины $x_{\min}(k)$:

$$x_i(k+1) = \hat{x}_i = (x_i(k) + x_{\min}(k)) / 2, \text{ где } i=0, 1, \dots, n.$$

На этом k -я итерация заканчивается.

Шаг 5.

Проверка сходимости.

Если

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i(k+1)) - f(x_0(k+1)))^2} \leq \varepsilon,$$

то поиск минимума заканчивается и полагается

$$\tilde{x} = x_0(k+1), \tilde{y} = f(x_0(k+1))$$

В противном случае $k=k+1$ и происходит переход к шагу 2.

1.5 Метод деформируемого симплекса (метод Нелдера – Мида).

В рассматриваемом методе симплекс перемещается с помощью трех основных операций над симплексом: отражение, растяжение и сжатие.

Схема алгоритма.

Шаг 1.

Построение начального симплекса.

Задаются начальная точка $x_0(0)$ и длина ребра l . Формируются остальные вершины симплекса: $x_i(0) = x_0(0) + l e_i$ ($i=1, 2, \dots, n$), где e_i – единичные векторы.

Шаг 2.

Определение направления улучшения решения.

Для этого на каждой итерации вычисляются значения целевой функции в каждой вершине симплекса. Пусть для всех i

$$f(x_{\min}(k)) \leq f(x_i(k)) \leq f(x_m(k)) \leq f(x_{\max}(k)),$$

где \min, m, \max, i -номера соответствующих вершин симплекса. Определим центр тяжести всех точек, исключая точку $x_{\max}(k)$,

$$C_k = \frac{1}{n} \sum_{i \neq \max} X_i(k)$$

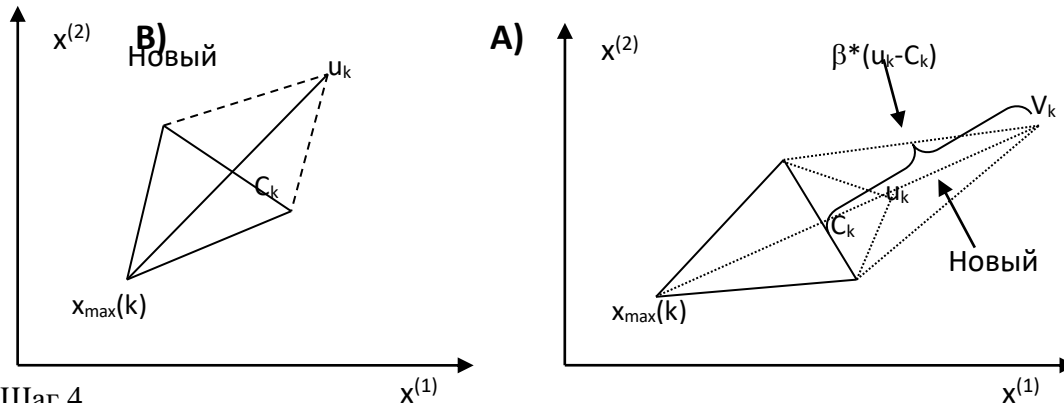
Тогда направление улучшения решения определяется векторов

$$C_k - x_{\max}(k).$$

Шаг 3.

Построение нового симплекса.

Замена вершины $x_{\max}(k)$ с максимальным значением целевой функции на новую точку с помощью операции отражения, результат которой является новая точка $u_k = C_k + \alpha * (C_k - x_{\max}(k))$, где α -коэффициент отражения.



Шаг 4.

Построение нового симплекса.

Вычисляем $f(u_k)$, при этом возможно один из трех случаев:

а) $f(u_k) < f(x_{\min}(k))$;

- б) $f(u_k) > f(x_m(k))$;
- в) $f(x_{\min}(k)) \leq f(u_k) \leq f(x_m(k))$;

Случай а): отражённая точка является точкой с наилучшим значением целевой функции. Поэтому направление отражение является перспективным и можно попытаться растянуть симплекс в этом направлении. Для этого строится точка

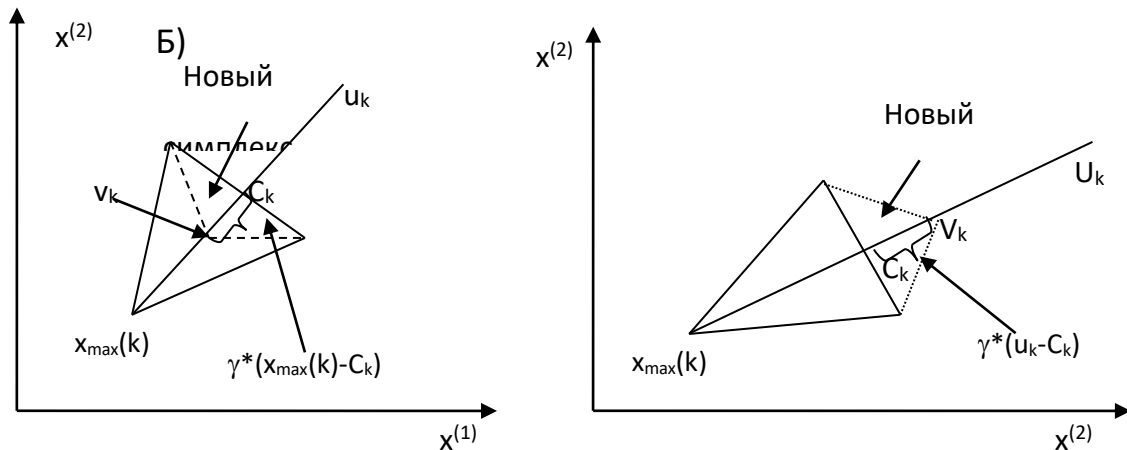
$$V_k = C_k + \beta^*(u_k - C_k), \text{ где } \beta > 1 \text{ – коэффициент расширения.}$$

Если $f(v_k) < f(u_k)$, то вершина $x_{\max}(k)$ заменяется на v_k , в противном случае на u_k и k -ая итерация заканчивается.

Случай б): в результате отражения получается новая точка u_k , которая, если заменить $x_{\max}(k)$, сама станет наихудшей. Поэтому в этом случае производится сжатие симплекса. Для этого строится точка v_k :

$$v_k = \begin{cases} C_k + \gamma^*(x_{\max}(k) - C_k), & \text{если } f(x_{\max}(k)) \leq f(u_k), \\ C_k + \gamma^*(u_k - C_k), & \text{если } f(x_{\max}(k)) > f(u_k), \end{cases}$$

где $0 < \gamma < 1$ – коэффициент сжатия.



Если $f(v_k) < \min\{f(x_{\max}(k)), f(u_k)\}$, то вершина $x_{\max}(k)$ заменяется на v_k .

В противном случае вершинам $x_i(k+1)$ ($i=0,1,2,\dots,n$) присваивается значение:

$$\tilde{x} = \frac{x_i(k) + x_{\min}(k)}{2}$$

и на этом k -ая итерация заканчивается.

в) вершина $x_{\max}(k)$ заменяется на u_k , чем определяется набор вершин $k+1$ -й итерации и k -ая итерация заканчивается.

Шаг 5.

Проверка сходимости.

Если:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i(k+1)) - f(x_0(k+1)))^2} \leq \varepsilon,$$

то поиск минимума заканчивается и полагается

$$\tilde{x} = x_0(k+1), \tilde{y} = f(x_0(k+1))$$

В противном случае $k=k+1$ и происходит переход к шагу 2.

Опыт использования описанного алгоритма показывает, что целесообразно брать следующие значения параметров:

$$\alpha=1, \beta=2, \gamma=0.5.$$

2. Задание на практическую работу

2.1 Изучить изложенные методы многомерной безусловной оптимизации.

2.2 В соответствии с вариантом задания, определенным преподавателем, найти точку минимума целевой функции $f(x)=f(x^{(1)}, x^{(2)})$ с заданной точностью ε указанными методами (составить программы, реализующие методы многомерной безусловной минимизации).

Целевая функция $f(x)=f(x^{(1)}, x^{(2)})$ зависит от двух аргументов. Функция $f(x)$ следующего вида:

$$f(x) = ax^{(1)} + bx^{(2)} + e^{cx+dx}.$$

№	Целевая функция				Начальное приближение	Точность решения
	a	b	c	d		
1	1	-1,4	0,01	0,11	(1;0)	0,0001
2	2	-1,3	0,04	0,12	(0;1)	0,00005
3	10	-0,5	0,94	0,2	(0;0)	0,0001
4	15	0	1,96	0,25	1,96	0,25
5	3	-1,2	0,02	1,3	(0;-1)	0,00005
6	11	-0,4	1	0,21	(-1;0)	0,0001
7	10	-1	1	2	(1;0)	0,0003
8	15	-0,5	2,25	2,5	(0;0)	0,0002
9	20	0,4	0,3	0,3	(0;-1)	0,0001
10	25	0,9	0,35	0,35	(1;0)	0,0004

Начальное приближение x_0 и точность ε приводятся в условии задачи. Сравнить результаты, полученные разными методами для одной и той же целевой функции (в частности, сравнить число вычисления целевой функции и её производных, понадобившихся для получения заданной точности). Для каждого применяемого метода построить траекторию промежуточных точек, получаемых на очередных шагах метода и сходящихся к точке минимума.

2.3 Оформить отчет о выполнении задания с приведением условия задачи, алгоритмов и программ, указанных в задании методов минимизации, графиков траекторий промежуточных приближений, таблицы результатов сравнения рассмотренных методов, заключения по результатам сравнения методов.

3. Вопросы к практической работе

- 1) Какова сущность прямых методов минимизации функции нескольких переменных?
- 2) Дайте характеристику метода минимизации по правильному симплексу.
- 3) Представьте алгоритм метода минимизации по правильному симплексу.
- 4) Дайте характеристику метода деформируемого симплекса (метод Нелдера – Мида)

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие/ А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

**МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ**
МЕТОДЫ ОПТИМИЗАЦИИ ДИСКРЕТНЫХ ЗАДАЧ

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

МЕТОДЫ ОПТИМИЗАЦИИ ДИСКРЕТНЫХ ЗАДАЧ: методические указания для
выполнения лабораторных работ по дисциплине «МЕТОДЫ ОПТИМИЗАЦИИ В
БИЗНЕС-АНАЛИТИКЕ» для студентов направления 09.04.03 «Прикладная
информатика» направленность «Искусственный интеллект и бизнес-аналитика» /
Составитель: Т. В. Хоменко. – Ульяновск: УлГТУ, 2021. – 37с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и бизнес-
аналитика». Цель данного практикума – ориентировать студентов на содержание
и порядок выполнения лабораторных работ во время прохождения студентами
курсов «Искусственный интеллект и бизнес-аналитика». Даются задания для
лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание	
ВВЕДЕНИЕ	4
Практическая работа №3	5
Тема: Переборные методы решения задач дискретной оптимизации.....	5
1. Краткие теоретические сведения	5
2. Задание на практическую работу	14
3. Вопросы к практической работе	15
Практическая работа №4	16
Тема: Метод динамического программирования в задачах дискретной оптимизации	16
1. Краткие теоретические сведения	16
2. Задание на практическую работу	34
3. Вопросы к практической работе	35
Список литературы	37

ВВЕДЕНИЕ

В учебно-методическом пособии описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.)

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Практическая работа №3

Тема: Переборные методы решения задач дискретной оптимизации

Цель работы: знакомство с оптимизационными задачами, изучение различных методов дискретной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Основные типы задач дискретной оптимизации

Задача дискретной оптимизации – это задача поиска максимума или минимума функции f , определенной на конечном или счетном множестве D :

$$f(x) \rightarrow \text{extr}, \quad x \in D.$$

Функция f называется целевой функцией, элементы множества D – допустимыми решениями. Если множество D задается системой ограничений:

$$\begin{aligned} g_i(x) &\leq 0, \quad i = 1, \dots, m_1, \quad g_i(x) = 0, \quad i = m_1 + 1, \dots, m, \\ x &= (x_1, \dots, x_n) \in R^n, \quad x_j \in \Omega_j \subset R, \quad j = 1, \dots, n_1, \quad n_1 \leq n, \end{aligned}$$

где каждое Ω_j – либо конечное множество, содержащее не менее двух элементов, либо счетное множество, то задача (1.1) называется задачей частично дискретного (или дискретного, если $n_1 = n$) математического программирования. Если $\Omega_j \subseteq Z$, $j = 1, \dots, n_1$, задача (1.1) называется задачей частично целочисленного (целочисленного при $n_1 = n$) программирования. В пособии будут рассматриваться, в основном, задачи целочисленного линейного программирования, которые формулируются следующим образом:

$$\begin{aligned} \sum_{j=1}^n c_j x_j &\rightarrow \min \\ \sum_{j=1}^n a_{ij} x_j &= b_i, \quad i = 1, \dots, m, \\ x_j &\geq 0, \quad j = 1, \dots, n, \\ x_j &\in Z, \quad j = 1, \dots, n_1, \quad n_1 \leq n. \end{aligned}$$

Обычно предполагается, что параметры задачи – коэффициенты целевой функции c_j , $j = 1, \dots, n$, элементы матрицы системы ограничений $A = (a_{ij})$ и вектора правых частей $b = (b_1, \dots, b_m)$ – целые числа.

Задача о рюкзаке

Задачей о рюкзаке или ранце называется задача целочисленного линейного программирования с одним ограничением:

$$\begin{aligned} \sum_{j=1}^n c_j x_j &\rightarrow \max \\ \sum_{j=1}^n a_j x_j &\leq b, \\ x_j &\geq 0 \text{ и целые, } j = 1, \dots, n. \end{aligned}$$

Числа c_j , a_j ($j = 1, \dots, n$), b можно считать положительными и целыми, $a_j \leq b$.

Имеется n видов неделимых предметов со стоимостями c_1, \dots, c_n и весами a_1, \dots, a_n . Требуется так упаковать ими рюкзак, чтобы его вес не превышал b , а суммарная

стоимость упакованных предметов была максимальной. Вводя переменные $x_j, j=1, \dots, n$, для количества упакованных предметов каждого типа, получаем представленную задачу. Иногда рассматривается задача о рюкзаке с ограничением типа равенства:

$$\sum_{j=1}^n a_j x_j = b,$$

что соответствует требованию полной загрузки рюкзака грузоподъемностью b .

Обобщением задачи о рюкзаке является задачей о многомерном рюкзаке:

$$\begin{aligned} \sum_{j=1}^n c_j x_j &\rightarrow \max \\ \sum_{j=1}^n a_{ij} x_j &\leq b_i, \quad i = 1, \dots, m, \\ x_j &\in \{0, 1\}, \quad j = 1, \dots, n, \end{aligned}$$

имеющая следующую экономическую интерпретацию.

Пусть есть n проектов, ожидаемая прибыль от реализации которых составляет c_1, \dots, c_n . Задан вектор ресурсов $b=(b_1, \dots, b_m)$, $b_i > 0, i=1, \dots, m$; количество единиц ресурса типа i , необходимое для реализации проекта с номером j , равно $a_{ij} > 0$. Для любого ресурса i выполнено условие:

$$\sum_{j=1}^n a_{ij} > b_i$$

то есть реализация всех проектов невозможна. Требуется выбрать набор проектов с максимальной суммарной прибылью.

Задача об упаковке

Задачу об упаковке в контейнеры можно сформулировать следующим образом: задано число $Q > 0$ и набор n положительных действительных чисел $L = \{w_1, \dots, w_n\}$, требуется разбить множество L на наименьшее возможное число подмножеств так, чтобы сумма чисел в каждом подмножестве не превосходила Q . Без ограничения общности можно считать, что:

$$Q=1, w_i \in (0, 1] \text{ при } i = 1, \dots, n \text{ и } \sum_{i=1}^n w_i > 1.$$

Задача интерпретируется как проблема упаковки предметов с размерами w_1, \dots, w_n в единичный контейнер.

Сформулируем задачу об упаковке как задачу целочисленного линейного программирования. Так как допустимым является решение, при котором в каждый контейнер помещается один предмет, можно считать, что имеется n контейнеров. Введем переменные $y_j, j=1, \dots, n$, принимающие значение 1, если j -й контейнер занят, и 0 в противном случае, и $x_{ij}, i, j=1, \dots, n$, равные 1, если i -й предмет помещен в j -й контейнер и 0, если нет. Задача тогда примет вид:

$$\begin{aligned} \sum_{j=1}^n y_j &\rightarrow \min \\ \sum_{i=1}^n w_i x_{ij} &\leq y_j, \quad j = 1, \dots, n, \\ \sum_{i=1}^n x_{ij} &= 1, \quad i = 1, \dots, n, \\ x_{ij} &\in \{0, 1\}, \quad y_j \in \{0, 1\}, \quad i, j = 1, \dots, n. \end{aligned}$$

Второе условие – ограничения на вместимость контейнера, третье условие

гарантирует, что каждый предмет помещен ровно в один контейнер.

Транспортная задача с фиксированными доплатами

Рассмотрим математическую модель следующей задачи организации перевозок однородного груза. Имеется m пунктов отправления (поставщиков) и n пунктов назначения (потребителей), предложение i -го поставщика составляет a_i , $i=1, \dots, m$, спрос j -го потребителя – b_j , $j=1, \dots, n$, при этом:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j.$$

Затраты на транспортировку x единиц груза от поставщика i до потребителя j составляют:

$$c_{ij}(x) = \begin{cases} 0, & x = 0, \\ c_{ij}x + d_{ij} & x > 0. \end{cases}$$

Числа $c_{ij} \geq 0$ интерпретируются как стоимость перевозки единицы груза и определяются, например, расстоянием между пунктами, постоянные затраты $d_{ij} \geq 0$ могут представлять собой оплату аренды транспортных средств или другие издержки, не зависящие от объема перевозок. Требуется организовать перевозку груза от поставщиков потребителям так, чтобы совокупные затраты на транспортировку были минимальны. Обозначив через x_{ij} количество единиц груза, перевозимого из i -го пункта производства в j -й пункт потребления, получим следующую задачу математического программирования с неоднородной разрывной линейной целевой функцией:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n c_{ij}(x_{ij}) &\rightarrow \min \\ \sum_{j=1}^n x_{ij} &= a_i, \quad i = 1, \dots, m, \\ \sum_{i=1}^m x_{ij} &= b_j, \quad j = 1, \dots, n, \\ x_{ij} &\geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \end{aligned}$$

Данная задача называется транспортной задачей с фиксированными доплатами или неоднородной транспортной задачей. Очевидно, если все $d_{ij}=0$, то рассматривается классическая транспортная задача. Данную задачу можно свести к задаче частично целочисленного линейного программирования. Для этого обозначим:

$$M_{ij} = \min\{a_i, b_j\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

и введем переменные:

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

принимающие значение 1, если $x_{ij} > 0$ и 0 в противном случае. Тогда задача при описанных ограничениях и дополнительных ограничениях:

$$x_{ij} \leq M_{ij} y_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

будет эквивалентна исходной.

Задачи коммивояжера

Коммивояжер, находящийся в родном городе, должен посетить несколько других населенных пунктов, побывав в каждом по одному разу, и вернуться обратно так, чтобы общая длина его пути была как можно меньше.

Пусть все пункты (включая начальный) перенумерованы и их общее количество

равно n . Расстояние от пункта i до пункта j обозначается c_{ij} . Полагая $c_{ii}=\infty$ (коммивояжеру запрещается оставаться на месте), определяем $n \times n$ матрицу $C=(c_{ij})$. Естественно считать, что $c_{ij} \geq 0$. Любой маршрут коммивояжера полностью определяется порядком посещенных пунктов и имеет, поэтому, вид $z=(i_1, i_2, \dots, i_n, i_1)$. Длина маршрута z – сумма соответствующих элементов матрицы C , то есть:

$$l(z) = \sum_{k=1}^n c_{i_k, i_{k+1}},$$

где $i_{n+1}=i_1$. Обозначив множество допустимых маршрутов через D , получаем задачу комбинаторной оптимизации:

$$l(z) \rightarrow \min, z \in D.$$

Очевидным практическим приложением задач коммивояжера является планирование маршрутов перевозки грузов.

Соответственно, c_{ij} может представлять собой не только расстояние, но и время, издержки, другой измеритель, служащий для определения «выгодности» маршрута. В общем случае, c_{ij} – стоимость следования j непосредственно после i .

Пусть $G=(V, A)$ – полный граф со множеством вершин $V=\{1, \dots, n\}$ и множеством дуг A , длина дуги (i, j) в котором равна c_{ij} . Контур (ориентированный цикл), включающий каждую вершину графа ровно один раз, называется гамильтоновым контуром. Таким образом, задача коммивояжера – это задача поиска гамильтонова контура, имеющего минимальную длину. В силу замкнутости маршрута можно считать, что начальный пункт i_1 задан, поэтому количество допустимых маршрутов равно $(n-1)!$. Если $c_{ij}=c_{ji}$ для всех $i, j=1, \dots, n$ (матрица C симметричная), граф G является неориентированным, и говорят о гамильтоновых циклах. Поскольку перестановки (i_2, \dots, i_n) и (i_n, \dots, i_2) определяют один маршрут, количество различных замкнутых маршрутов составит $(n-1)!/2$.

1.2. Переборные методы решения задач дискретной оптимизации

Для всех представленных выше задач множество решений конечно и, перебрав все из них, можно найти лучшее. Для реализации полного перебора нужна процедура генерации всех допустимых решений:

$$\begin{cases} Q(\bar{x}^*) = \max_{\bar{x} \in D} Q(\bar{x}), \\ 0 < |D| = N < \infty. \end{cases}$$

Здесь множество D – конечно, и количество возможных решений равно N . В силу этого формально все допустимые решения можно пронумеровать:

$$\bar{x}^1, \bar{x}^2, \dots, \bar{x}^N.$$

Поиск лучшего варианта из множества допустимых решений сводится к полному перебору. Именно этим обстоятельством объясняется название переборных задач.

Как правило, полный перебор невозможно осуществить на практике из-за огромного числа рассматриваемых вариантов. Методы полного перебора применяются в случае малых размеров входных данных или в виде частичного перебора в других алгоритмах. Для оценивания времени работы алгоритмов, основанных на полном переборе можно подсчитать число возможных решений для задач.

ПРИМЕР 1. Задача о рюкзаке

Каждый предмет может быть либо помещен в рюкзак, либо нет. Соответственно, каждая переменная x_j может принимать одно из двух возможных значений – 1 или 0.

Любому решению соответствует вектор из n нулей и единиц. Всего таких векторов 2^n .

Пусть $V=10$, $n=3$, $a_1=6$, $a_2=5$, $a_3=4$, $c_1=3$, $c_2=4$, $c_3=5$. Рассмотрим всевозможные комбинации значений переменных x_1 , x_2 , x_3 и соответствующие им наборы:

x_1	x_2	x_3	набор	вес	стоимость
0	0	0	\emptyset	0	0
1	0	0	{1}	6	3
0	1	0	{2}	5	4
0	0	1	{3}	4	5
1	1	0	{1, 2}	11	7
1	0	1	{1, 3}	10	8
0	1	1	{2, 3}	9	9
1	1	1	{1, 2, 3}	15	12

Сначала отбираем наборы, которые помещаются в рюкзак (1, 2, 3, 4, 6, 7), среди которых выбираем набор с наибольшей стоимостью. Очевидно, что можно взять 6-ой набор (первый и третий предметы), можно взять 7-ой набор (второй и третий предметы). Однако у 7-го набора стоимость выше.

ПРИМЕР 2. Задачи коммивояжера

Нужно найти перестановку городов A_1, \dots, A_n , для которой пройденное расстояние минимально. Первый элемент искомой перестановки всегда равен 1. Поэтому при полном переборе переставлять нужно оставшиеся $n-1$ города. Всего таких перестановок $(n-1)!$.

Пусть $n=5$, матрица расстояний между городами имеет вид:

$$C = \begin{pmatrix} 0 & 5 & 3 & 1 & 2 \\ 5 & 0 & 1 & 3 & 6 \\ 3 & 1 & 0 & 4 & 2 \\ 1 & 3 & 4 & 0 & 7 \\ 2 & 6 & 2 & 7 & 0 \end{pmatrix}.$$

Тогда получаем $4!=24$ различных перестановок. Если при этом матрица задачи симметрическая, то получаем 12 перестановок. Таблица всевозможных маршрутов имеет следующий вид:

маршрут	стоимость
1 → 2 → 3 → 4 → 5 → 1	19
1 → 2 → 3 → 5 → 4 → 1	16
1 → 2 → 4 → 3 → 5 → 1	16
1 → 2 → 4 → 5 → 3 → 1	20
1 → 2 → 5 → 3 → 4 → 1	18
1 → 2 → 5 → 4 → 3 → 1	25
1 → 3 → 2 → 4 → 5 → 1	16
1 → 3 → 2 → 5 → 4 → 1	18
1 → 3 → 4 → 2 → 5 → 1	18
1 → 3 → 5 → 2 → 4 → 1	15
1 → 4 → 2 → 3 → 5 → 1	9
1 → 4 → 3 → 2 → 5 → 1	14

Очевидно, оптимальный маршрут выглядит так:

$$1 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 5 \rightarrow 1.$$

При этом, пройденное коммивояжером расстояние равно 9.

1.3. Методы генетического поиска в решении задач дискретной оптимизации

Как правило, задачи дискретной оптимизации характеризуются большим числом переменных, и, как следствие, большим пространством поиска, что не дает возможности перебрать все многообразие решений за «разумное» время. С другой стороны, на практике зачастую и не требуется нахождения глобального решения, достаточно найти «приемлемое» решение по заданному критерию (или набору критериев). Этими обстоятельствами объясняется повышенный интерес к комбинаторным методам оптимизации. В основе многих комбинаторных алгоритмов лежат эвристические схемы нахождения решений, которые эксплуатируют особенности объекта исследования.

Однако и подобные методы не всегда удается использовать на практике, так как объект исследования может иметь очень непростую природу, и выявление закономерностей и особенностей его внутренней организации может оказаться нетривиальной задачей. В связи с этим возникла проблема практической разрешимости задач дискретной оптимизации: найти эффективный или хотя бы достаточно простой в практически важных случаях алгоритм ее решения. Другой прием, который позволил сократить вычислительные затраты и соответственно увеличить размерность задач, заключается в ведении элементов рандомизации. Эволюционные методы представляют целый спектр алгоритмов, которые активно эксплуатируют данный прием. Как уже было отмечено, методы эволюционных вычислений не гарантируют обнаружения оптимального решения за полиномиальное время. Эти методы позволяют исследовать и находить приемлемые решения таких задач, решение которых при помощи традиционных методов оказывается затруднительным, а в некоторых случаях и просто невозможным.

Введем ряд определений.

Репродукционно-популяционный алгоритм поиска – это алгоритм поисковой оптимизации, который начинается с начальной популяции P_0 (совокупности кодировок $(\chi_1^0, \chi_2^0, \dots, \chi_v^0)$) и итеративно выполняет следующий цикл операций:

1) *Оценивание* – вычисление значений функции приспособленности для любой кодировки $\chi \in S$.

2) *Селекция* – выбор из популяции P^t , $t=0,1,\dots$ репродукционного множества R^t (подмножества кодировок $R^t \subseteq P^t$).

3) *Воспроизводство (размножение)* – генерация из репродукционного множества R^t новых кодировок с помощью комбинаций следующих операций:

- *копирование* – создание тождественных копий некоторых или всех кодировок из R^t ;
- *скрещивание* – конструирование новых кодировок путем сцепления подстрок тех кодировок, которые выбираются путем копирования из R^t ;
- *мутация* – конструирование новых кодировок путем подстановки символов из алфавита V в выбранные позиции одной из кодировок $\chi \in R^t$.

4) *Замена* – формирование на очередном шаге (поколении) новой популяции P^{t+1} путем замены некоторых или всех кодировок $\chi \in P^t$.

Канонические генетические алгоритмы – это подкласс репродукционно-популяционных алгоритмов поиска, в которых основные операторы имитируют механизмы естественного отбора и популяционной генетики, удовлетворяющий следующим требованиям:

- все кодировки $\chi \in S$ имеют одну и ту же длину L ;

- каждая популяция P^t , $t=0,1,\dots$ имеет постоянную численность n ;
- мощность репродукционного множества $|R^t|=2$; элементы $\chi', \chi'' \in R^t$, называемые *родителями*, выбираются из популяции P^t случайным образом с вероятностями, пропорциональными значениям функции приспособленности;
- скрещивание осуществляется с помощью *одноточечного кроссовера*, когда символы $(\chi_1, \chi_2, \dots, \chi_i)$ в новой кодировке χ , называемой «потомком», являются символами кодировки $\chi' \in R^t$, а символы $(\chi_{i+1}, \chi_{i+2}, \dots, \chi_L)$ переходят от кодировки $\chi'' \in R^t$; индекс i , называемый *точкой кроссовера*, выбирается случайным образом с равной вероятностью из интервала $[1, L-1]$; мутации являются стохастическими операциями и обычно не зависят от значений функции приспособленности;
- замена сводится к исключению из популяции P^t особи с наименьшим значением функции приспособленности и включению в нее одной из новых кодировок с наибольшим значением функции приспособленности.

1.4. Обобщенная структура генетического алгоритма

Обобщенная структура генетического алгоритма может быть представлена следующим образом.

Шаг 1. Инициализация начальной популяции P^0 численностью v .

1.1. Установить номер текущего поколения $t:=0$.

1.2. Сгенерировать случайным образом хромосомный набор из v строковых кодировок фиксированной длины L , в котором Хэммингово расстояние между любой парой кодировок не равно нулю.

1.3. Оценить каждую строку из хромосомного набора с помощью функции приспособленности.

Шаг 2. Воспроизводство потомков с наследственными генетическими свойствами родителей.

2.1. Выбрать случайным образом из текущей популяции P^t согласно схеме скрещивания кодировку двух родителей, образующих «брачную пару».

2.2. Сгенерировать при помощи оператора кроссовера для выбранной «брачной пары» с вероятностью p_c одну или несколько кодировок потомков, которые наследуют генетические свойства родителей.

2.3. Оценить каждую кодировку потомков с помощью функции приспособленности.

2.4. Повторять все операции с п.2.1 до тех пор, пока не будет рассмотрено заданное число «брачных пар» N_c .

Шаг 3. Создание мутантов с генетическими свойствами, отличающимися от свойств родителей.

3.1. Выбрать случайным образом из числа потомков и/или родителей кодировку, наследующую генетические свойства одного или обоих родителей.

3.2. Сгенерировать при помощи оператора мутации для выбранной кодировки с вероятностью p_m кодировку-мутанта, обеспечивающую изменчивость генетических свойств родителей.

3.3. Оценить кодировку мутанта с помощью функции приспособленности.

3.4. Повторять все операции с п.3.1 до тех пор, пока не будет получено заданное число мутантов N_m .

Шаг 4. Замена текущей популяции P^t новой популяцией P^{t+1} .

4.1. Выбрать стратегию формирования популяции P^{t+1} .

4.2. Сформировать из родителей и/или «детей» (потомков и мутантов) репродукционное множество кодировок, различающихся между собой по Хэммингову расстоянию.

4.3. Скопировать при помощи оператора селекции из репродукционного множества кодировки, реализующие стратегию формирования популяции P^{t+1} .

Шаг 5. Условие выхода из итерационного цикла.

Сменить номер текущего поколения $t:=t+1$ и повторить все операции с п.2, если условие окончания генетического поиска не выполнено (например, эволюция популяции P^t считается законченной, если она исчерпала свой жизненный цикл T , т.е. если $t>T$).

Параметрами генетического алгоритма являются:

- n – численность популяции P^t , $t = 0, 1, \dots$;
- N_c – число «брачных пар»;
- N_m – число «мутантов»;
- p_c – вероятность кроссовера;
- p_m – вероятность мутации;
- T – число поколений, в течение которых осуществляется генетический поиск.

1.5. Сравнение качества решений

При сравнении методов формируется выборка $\Lambda = \{G_1, G_2, \dots, G_K\}$, из $K = |\Lambda|$ тестовых примеров, для каждого из которых производится построение решений с использованием выбранных методов $\Phi = \{m_1, m_2, \dots, m_T\}$, $T = |\Phi|$ и оценка их качества.

Рассмотрим в качестве примера сравнение качества решений эвристических методов в задаче поиска кратчайшего пути в графе. Пусть качество решения, полученного методом m_i , $i = \overline{1, T}$, на тестовом примере G_j , $j = \overline{1, K}$ определяется как $Q_{m_i}(G_j)$. Тогда простейшим способом сравнения качества решений является сопоставление их средневыборочных значений:

$$\gamma_{m_i}(\Lambda) = \frac{\sum_{j=1}^K Q_{m_i}(G_j)}{K} \rightarrow \min \quad (5.1)$$

Однако в случае, когда для некоторых тестовых примеров корректные решения не найдены (для них $Q_{m_i}(G_j) = \infty$), данная формула неприменима, и вместо нее при сравнении методов необходимо использовать критерий:

$$\Delta\gamma_{m_i}(\Lambda) = \frac{\sum_{j=1}^K (Q_{m_i}(G_j) - Q^*(G_j)) \phi_{m_i}(G_j)}{K} \rightarrow \min \quad (5.2)$$

где $Q^*(G_j)$ – качество наилучшего решения.

Критерий (5.2) представляет собой среднюю величину ухудшения решения по сравнению с известным оптимумом. При этом для рассматриваемых эвристических методов возможно оценить вероятность получения решения:

$$\rho_{m_i}(\Lambda) = \frac{\sum_{j=1}^K \phi_{m_i}(G_j)}{K} \rightarrow \max \quad (5.3)$$

причем метод, для которого $\rho_{m_i} \rightarrow 1$, является предпочтительным. Таким образом, для

объективного сравнения методов необходимо учитывать пару критериев, характеризующих среднее качество решений (или среднее отклонение от оптимума) и вероятность отыскания решения.

Еще одним возможным критерием сравнения эвристических методов является вероятность получения суб- или квазиоптимальных решений:

$$\rho_{m_i}^*(\Lambda) = \frac{\sum_{j=1}^K \theta_{m_i}(G_j)}{K} \rightarrow \max \quad (5.4)$$

где

$$\theta_{m_i}(G_j) = \begin{cases} 0, & Q_{m_i}(G_j) > Q^*(G_j), \\ 1, & Q_{m_i}(G_j) = Q^*(G_j) \end{cases} \quad (5.5)$$

– функция, принимающая единичное значение в случае совпадения качества текущего $Q_{m_i}(G_j)$ и наилучшего $Q^*(G_j)$ решений.

Отклонения качества решений $\Delta Q(G_j) = Q_{m_i}(G_j) - Q^*(G_j)$ имеют случайный характер, и для них может быть произведена оценка вероятности того, что найденное решение отличается от оптимума не более чем на заданную величину ΔQ :

$$\rho_{m_i}^\Delta(\Delta Q, \Lambda) = \frac{\sum_{j=1}^K \mu_\Delta(Q_{m_i}(G_j), Q^*(G_j), \Delta Q)}{K}, \quad (5.6)$$

где

$$\mu_\Delta(Q, Q^*, \Delta Q) = \begin{cases} 1, & Q - Q^* \leq \Delta Q, \\ 0, & Q - Q^* > \Delta Q \end{cases} \quad (5.7)$$

– функция, принимающая единичное значение в случае, если отклонение качества решения $\Delta Q(G_j) = Q(G_j) - Q^*(G_j)$ не превышает заданное пороговое значение ΔQ , и нулевое в противном случае. Обычно при сравнении эвристических методов используют

графики зависимости $\rho_{m_i}^\Delta$ от ΔQ . Несложно показать, что при $\Delta Q = 0$ $\rho_{m_i}^\Delta = \rho_{m_i}^*$.

При работе с малыми выборками возможна ситуация, в которой разница в значениях сопоставляемых критериев вызвана статистической погрешностью, а не различием в поведении методов. Для того, чтобы статистически корректно доказать преимущество одного метода по сравнению с другим, необходимо убедиться в том, что различие сопоставляемых средних оценок является статистически значимым. Для этого необходимо произвести расчет дисперсии средневывборочных оценок:

$$D_{m_i}(\Lambda) = \frac{1}{K} \sum_{j=1}^K (Q_{m_i}(G_j) - \gamma_{m_i})^2, \quad (5.8)$$

оценить величину среднеквадратичного отклонения:

$$\sigma_{m_i}(\Lambda) = \sqrt{D_{m_i}}$$

и рассчитать ширину доверительного интервала ΔQ для заданной доверительной вероятности. Например, для доверительной вероятности 0,95:

$$\Delta Q_{m_i}(\Lambda) = \frac{1,96\sigma_{m_i}}{\sqrt{K}}$$

где коэффициент 1,96 представляет собой квантиль распределения Стьюдента для заданной доверительной вероятности. Смысл указанных действий заключается в том, что

при записи средневывборочного значения в виде $\gamma_{m_i}(\Lambda) \pm \Delta Q_{m_i}(\Lambda)$ можно утверждать, что истинное среднее значение математического ожидания (γ_{m_i} при $K \rightarrow \infty$) находится в пределах доверительного интервала $[\gamma_{m_i}(\Lambda) - \Delta Q_{m_i}(\Lambda); \gamma_{m_i}(\Lambda) + \Delta Q_{m_i}(\Lambda)]_c$ заданной доверительной вероятностью.

2. Задание на практическую работу

2.1. Написать программу, реализующую решение, указанной в индивидуальном варианте, задачи полным перебором и генетическим алгоритмом. Сравнить качество решений путем организации вычислительного эксперимента над выборкой из K псевдослучайных тестовых примеров, для каждого из методов вычислить:

- средневывборочные оценки качества решений;
- соответствующие им дисперсии;
- среднеквадратичные отклонения;
- границы доверительных интервалов (для доверительной вероятности 0,95);
- вероятности отыскания решений;
- вероятности отыскания оптимального (или условно оптимального, в зависимости от задачи) решения.

Индивидуальные варианты

1	В заданном ориентированной графе найти кратчайший гамильтонов путь, соединяющий указанную пару вершин (гамильтонов путь – это путь, однократно проходящий через все вершины графа)
2	В заданном неориентированном графе найти кратчайший гамильтонов цикл (гамильтонов цикл – это цикл, однократно проходящий через все вершины графа)
3	Определить хроматическое число для заданного неориентированного графа (хроматическим числом неориентированного графа называется минимальное число цветов, в которое можно раскрасить вершины графа так, чтобы соединенные ребром вершины были раскрашены в разные цвета)
4	Найти такой способ перенумерации вершин (подстановку изоморфизма), который дает минимальное число расхождений между матрицами смежности рассматриваемых графов (графы называются изоморфными, если из одного можно получить другой путем перенумерации его вершин)
5	Найти такое разбиение графа на N подграфов, чтобы сумма ребер, связывающих полученные подграфы, была минимальна (для неориентированного графа заданы веса ребер)
6	Для заданного графа оценить мощность максимального независимого множества (максимальным независимым множеством называется такое максимальное по включению подмножество вершин графа, в котором ни одна пара вершин не соединена ребром)
7	Для заданного неориентированного графа найти мощность максимального по включению полносвязного подграфа
8	Для заданного графа определить его диаметр (диаметр графа – это максимальное расстояние между вершинами графа)
9	Для заданного ориентированного графа оценить мощность компоненты сильной связности (компонентой сильной связности ориентированного графа называется

	максимальное по включению множество взаимно достижимых вершин)
10	Найти минимальное остовное дерево заданного неориентированного графа (минимальным остовным деревом графа G называется дерево с минимальной суммарной длиной ребер, образованное из ребер графа G)

2.2. Построить графики зависимости вероятности получения решения с заданным отклонением от оптимума (или условного оптимума, в зависимости от задачи) от величины допустимого отклонения. Если в состав метода входят настроечные параметры, произвести метаоптимизацию, результаты которой представить в виде графиков зависимостей средневыборочных величин от значений настроечных параметров. Для указанных выше средневыборочных оценок произвести построение графиков их зависимостей от размерности задачи N и плотности графа d .

Произвести оценку времени отыскания решения, построить графики зависимости средневыборочного качества решений от числа итераций.

3. Вопросы к практической работе

- 1) Какое решение называется оптимальным?
- 2) Какие методы гарантируют получение оптимальных решений?
- 3) Дайте характеристику метода полного перебора
- 4) Какие алгоритмы называются генетическими?
- 5) Дайте характеристику обобщенного генетического алгоритма
- 6) Раскрыть процедуру сравнения качества решений

Практическая работа №4

Тема: Метод динамического программирования в задачах дискретной оптимизации

Цель работы: знакомство с оптимизационными задачами, изучение различных методов дискретной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Принцип динамического программирования

Метод динамического программирования – один из основных математических методов оптимизации. При его реализации одновременный выбор значений большого числа переменных решаемой экстремальной задачи заменяется поочередным определением каждой из них в зависимости от возможных обстоятельств. Процедуру выбора значений переменных будем трактовать как многоэтапный процесс управления некоторой системой. Далее объект Ω именуем дискретной управляемой системой, если множество его возможных состояний конечно и управления в нем осуществляются в дискретном времени, пошагово; каждое управление заключается в применении одного из конечного числа возможных воздействий; результатом любого воздействия является изменение состояния системы. С каждым изменением состояния системы связываем платеж (доход или расход); полагаем, что величина платежа на любом шаге зависит от текущего состояния системы и применяемого управляющего воздействия. Начальное состояние системы считается заданным. Определено также множество финальных состояний (достигнув финального состояния, система прекращает функционировать). Каждая последовательность управлений, переводящая систему из начального в одно из финальных состояний, фактически определяет некоторую полную траекторию движения системы. Требуется найти полную траекторию, оптимальную по значению суммарного платежа, т.е. обеспечивающую максимальный суммарный доход или минимальный суммарный расход. Возможно и другое требование – найти полную траекторию с минимальным значением максимального из пошаговых расходов (максимальным значением минимального из пошаговых доходов).

Формально дискретную управляемую систему определяем как совокупность:

$$\Omega = \{D; x_0; F; V(x), f(x, v), s(x, v)\}.$$

где D – множество состояний системы (множество D конечно); x_0 – начальное состояние; F – множество финальных состояний, $x_0 \in D$, $x_0 \notin F$, $F \subset D$; $V(x)$ – конечное множество возможных в состоянии x управлений, $x \in D \setminus F$; $f(x, v)$ – функция переходов (из состояния x под воздействием управления v система переходит в состояние $f(x, v)$), $x \in D \setminus F$, $v \in V(x)$, $f(x, v) \in D$; $s(x, v)$ – функция платежа, здесь $x \in D \setminus F$, $v \in V(x)$; значения функции платежа считаются неотрицательными.

Состояние x системы Ω называем промежуточным, если оно не является ни начальным, ни финальным. Конечную последовательность $T = \{x_0, x_1, x_2, \dots, x_n\}$ состояний системы Ω именуем траекторией системы, если выполняются условия:

$$x^t = f(x^{t-1}, v^t), \text{ где } v^t \in V(x^{t-1}), t = 1, 2, \dots, n.$$

Состояние x^0 – начальное состояние траектории T , а состояние x^n – ее конечное состояние. Состояния x^2, x^3, \dots, x^{n-1} из T являются промежуточными состояниями данной

траектории.

Траекторию называем полной, если начальным ее состоянием является начальное состояние системы Ω , а конечным – одно из финальных состояний этой системы. Таким образом, для полной траектории $T=\{x^0, x^1, x^2, \dots, x^n\}$ имеет место $x^0=x_0$ и $x^n \in F$.

Траекторию $T=\{x^0, x^1, x^2, \dots, x^n\}$ называем заключительной x -траекторией, если $x^0=x$ и $x^n \in F$. Заключительная x -траектория, имея своим начальным состоянием произвольное состояние x системы Ω , заканчивается в одном из финальных состояний системы.

Траекторию $T=\{x^0, x^1, x^2, \dots, x^n\}$ называем начальной x -траекторией, если $x^0=x_0$ и $x^n=x$. Начальная x -траектория, имея своим конечным состоянием x , начинается от начального состояния системы.

Состояние x системы Ω именуем достижимым, если имеет место $P(x^0, x)$. Отметим, что начальное состояние x^0 достижимо по определению.

Состояние x системы Ω именуем продуктивным, если для некоторого финального состояния f имеет место $P(x, f)$. Финальные состояния системы продуктивны по определению.

Пусть $T=\{x^0, x^1, x^2, \dots, x^n\}$ – произвольная траектория системы Ω , причем:

$$x^t = f(x^{t-1}, v^t), \text{ где } v^t \in V(x^{t-1}), t=1, 2, \dots, n.$$

Стоимость $C(T)$ траектории T определяем следующим образом:

$$C(T) = \sum_{t=1}^n s(x^{t-1}, v^t).$$

Таким образом, стоимость траектории – это сумма пошаговых платежей, имеющих место при реализации траектории. Центральной является задача построения полной траектории, оптимальной по значению суммарной стоимости, т.е. обеспечивающей максимальный суммарный доход или минимальный суммарный расход. Обе экстремальные задачи по способу решения идентичны. Для определенности проблему формулируем следующим образом.

Задача А. Для системы Ω найти полную траекторию T минимальной стоимости. Дополнительно введем в рассмотрение совокупность частных задач $A(x)$, где x принадлежит $D \setminus F$.

Задача $A(x)$. Для системы Ω и ее состояния x найти заключительную x -траекторию минимальной стоимости.

Решения сформулированных экстремальных задач A и $A(x)$ именуем оптимальной полной траекторией и оптимальной заключительной x -траекторией соответственно.

В основе метода динамического программирования лежит принцип оптимальности, сформулированный американским математиком Ричардом Беллманом следующим образом: «Оптимальное поведение обладает тем свойством, что каковы бы ни были первоначальное состояние и решение в начальный момент, последующие решения должны составлять оптимальное поведение относительно состояния, являющегося результатом первого решения».

Следующая теорема выражает принцип Беллмана в применении к задачам построения оптимальных траекторий.

Теорема. Если траектория $T=\{x^0, x^1, x^2, \dots, x^n\}$ оптимальна, то любая ее заключительная часть $T_k = \{x^k, x^{k+1}, \dots, x^n\}$ также оптимальна.

Стоимость оптимальной в задаче A полной траектории обозначим символом B_A . Стоимость заключительной x -траектории, оптимальной в задаче $A(x)$, обозначим $B(x)$. Функция $B(x)$ называется функцией Беллмана. Отметим, что $B_A = B(x_0)$.

Очевидны следующие равенства:

$$B(x) = 0, x \in F. \quad (6.1)$$

Пусть x - произвольное нефинальное состояние системы. В этом состоянии можно реализовать любое принадлежащее множеству $V(x)$ управление. Предположим, что выбрано управление $v, v \in V(x)$. Данное управление влечет платеж $s(x, v)$, а следующим состоянием системы оказывается $f(x, v)$. Если далее реализуется оптимальная заключительная $f(x, v)$ - траектория, то суммарная величина последующих платежей оказывается равной $B(f(x, v))$. Из принципа Беллмана получаем:

$$B(x) = \min_{v \in V(x)} \{s(x, v) + B(f(x, v))\}, x \in D \setminus F. \quad (6.2)$$

Вычисление значений функции Беллмана по соотношениям (6.1) - (6.2) выполняется поэтапно в следующем порядке.

На первом этапе фиксируются значения $B(x) = 0$ для всех $x \in F$.

Далее на каждом следующем этапе вычисление очередного значения функции Беллмана выполняется для произвольного состояния x такого, что $B(x)$ неизвестно, но значения $B(y)$ для всех непосредственно следующих за x состояний y уже найдены (состояние y системы Ω относим к числу непосредственно следующих за состоянием x , если пара $\{x, y\}$ является траекторией системы Ω).

Последним в процессе счета определяется значение $B(0)$.

Для синтеза оптимальной полной траектории системы при выполнении, определяемой соотношениями (6.1)-(6.2), процедуры счета следует дополнительно составлять список оптимальных переходов (СОП). Каждая запись этого списка имеет вид $[x_i, x_j]$, где x_i - произвольное нефинальное состояние, а $x_j = f(x_i, v^*)$ - состояние, в которое система переходит из x_i под воздействием управления v^* , возможного в состоянии x_i и обращающего в минимум правую часть (6.2); если таких управлений несколько, фиксируется любое из них (нашей целью считаем построение любой из оптимальных траекторий, но не всех таких траекторий). Записи СОП взаимно однозначно соответствуют нефинальным состояниям системы, при этом запись $[x_i, f(x_i, v)]$ считается соответствующей состоянию x_i . Составленный СОП полагаем упорядоченным по возрастанию индексов первых компонент входящих в него записей. Чтобы построить оптимальную траекторию, из СОП извлекаем последовательность записей W , первым элементом которой является запись, соответствующая состоянию x^0 ; каждая следующая запись последовательности W имеет своей первой компонентой вторую компоненту предыдущей записи этой последовательности. В последней записи последовательности W второй компонентой является некоторое финальное состояние системы Ω . Из составленного СОП последовательность W извлекается однозначно.

Пусть $W = \{[x_0, y_1], [y_1, y_2], [y_2, y_3], \dots, [y_{m-1}, y_m]\}$, здесь $y_m \in F$. Тогда $T = \{x_0, y_1, y_2, y_3, \dots, y_m\}$ - искомая оптимальная траектория.

Уравнения (6.1) - (6.2) - рекуррентные соотношения динамического программирования для решения задачи A . Основанный на этих соотношениях метод решения задачи A называется методом динамического программирования.

Пусть N - число состояний системы Ω . Тогда верхняя оценка числа вычисляемых значений функции Беллмана в процессе решения задачи A равна N . Число элементарных операций, выполняемых при определении по формуле (6.2) каждого конкретного значения этой функции не превышает CN , где C - некоторая не зависящая от N константа. Таким образом, верхней оценкой числа элементарных операций, выполняемых при решении задачи A методом динамического программирования является CN^2 , где N - число

состояний системы Ω , а C – не зависящая от N константа. Для запоминания в процессе вычислений всех найденных значений функции Беллмана необходим объем памяти, пропорциональный N ; объем памяти такого же порядка нужен для создания списка оптимальных переходов. В итоге получаем, что объем памяти, нужной для решения задачи A методом динамического программирования, линейно зависит от N .

Следует отметить, что в связи с конечностью числа состояний системы Ω число ее траекторий конечно и задача A в принципе может быть решена путем перебора конечного числа вариантов. Метод динамического программирования позволяет определенным образом упорядочить и существенно сократить перебор.

ПРИМЕР 1. Имеется дискретно управляемая система ΩI , определяющий ее граф $G(\Omega I)$ представлен на рисунке 1.

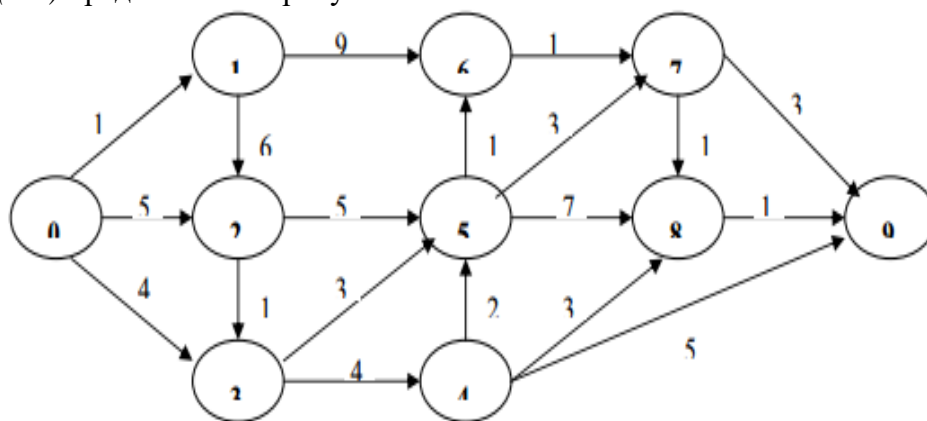


Рисунок 1 – Граф $G(\Omega I)$

Начальным является состояние 0, множество финальных состояний одноэлементно: $F=\{9\}$. Требуется найти полную траекторию минимальной стоимости.

Для состояний системы ΩI (вершин графа $G(\Omega I)$) вычисляем значения функции Беллмана. На основании (6.1) фиксируем $B(9)=0$. Далее, пользуясь (6.2), последовательно получаем:

$$\begin{aligned}
 B(8) &= 1; & B(7) &= \min[3, 1+B(8)]=2; & B(6) &= 1+B(7)=3; \\
 B(5) &= \min[1+B(6), 3+B(7), 7+B(8)]=\min[4, 6, 8]=4; \\
 B(4) &= \min[2+B(5), 3+B(8), 5+B(9)]=\min[6, 4, 5]=4; \\
 B(3) &= \min[3+B(5), 4+B(4)]=\min[7, 8]=7; \\
 B(2) &= \min[1+B(3), 5+B(5)]=\min[8, 9]=8; \\
 B(1) &= \min[9+B(6), 6+B(2)]=\min[12, 14]=12; \\
 B(0) &= \min[1+B(1), 5+B(2), 4+B(3)]=\min[13, 13, 11]=11.
 \end{aligned}$$

Таким образом, минимальная из стоимостей полных траекторий в рассматриваемой задаче равна 11. Получаемый результат представлен на рисунке 2.

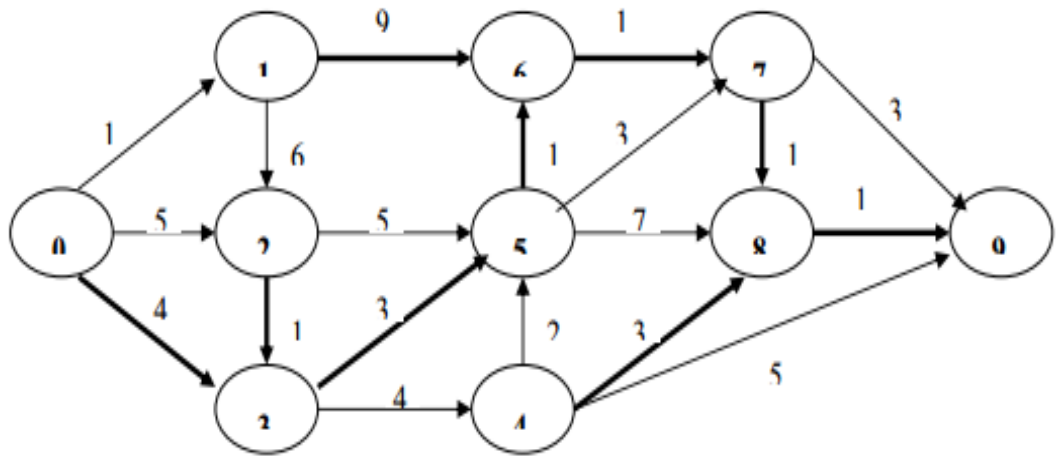


Рисунок 2 – Граф $G(\Omega I)$ с выделенными дугами

При подсчетах значений функции Беллмана в графе $G(\Omega I)$ для каждой нефинальной вершины специально выделяем (рисуем жирной линией) дугу, соответствующую управлению, в ней принимаемому (т.е. обращающему в минимум правую часть (6.2)).

Как показывают выделенные дуги, на первом такте процесса управления следует выполнить переход из вершины 0 в вершину 3, на втором такте - из вершины 3 в вершину 5, на третьем такте - из вершины 5 в вершину 6, на четвертом такте - из вершины 6 в вершину 7, на пятом такте - из вершины 7 в вершину 8, на шестом, заключительном такте - из вершины 8 в вершину 9.

Таким образом, оптимальной по критерию суммарной стоимости полной траекторией является $T = \{0, 3, 5, 6, 7, 8, 9\}$.

Далее введем совокупность частных задач следующего вида.

Задача $A^*(x)$. Для системы Ω и ее состояния x найти начальную x -траекторию минимальной стоимости.

Оптимальное решение задачи $A^*(x)$ именуем оптимальной начальной x -траекторией системы Ω .

Имеет место следующий факт.

Теорема 2. Если траектория $T = \{x^0, x^1, x^2, \dots, x^n\}$ оптимальна, то любая ее начальная часть также оптимальна.

Стоимость начальной x -траектории, оптимальной в задаче $A^*(x)$, обозначим $B^*(x)$. Очевидно, что:

$$B^*(x_0) = 0. \quad (6.3)$$

В системе Ω произвольное состояние x непосредственно предшествует состоянию y , если пара $\{x, y\}$ является траекторией системы Ω . Управление, переводящее систему Ω из состояния x непосредственно в состояние y , обозначим $v[x, y]$. Множество состояний системы, непосредственно предшествующих состоянию y , обозначим $\Gamma(y)$. Основываясь на теореме 2, запишем соотношение, позволяющее организовать процесс вычислений значений функции $B^*(y)$ для всех состояний y , отличных от начального:

$$B^*(y) = \min_{x \in \Gamma(y)} \{B^*(x) + s(x, v[x, y])\}, \quad y \in D \setminus x_0; \quad (6.4)$$

Отметим, что:

$$B_A = \min_{x \in F} B^*(x). \quad (6.5)$$

Вычисление значений функции $B^*(y)$ на основе рекуррентных соотношений (6.3) -

(6.4) выполняется поэтапно в следующем порядке. На первом этапе фиксируется значение $V^*(x_0)=0$. Далее на каждом следующем этапе вычисление очередного значения функции V^* выполняется для произвольного состояния y такого, что $V^*(y)$ неизвестно, но значения $V^*(x)$ для всех состояний x , непосредственно предшествующих состоянию y , уже найдены.

Метод решения задачи A , основанный на соотношениях (6.3) – (6.5), – *прямой метод Беллмана*. В отличие от него, метод, основанный на соотношениях (6.1) – (6.2), носит название *обратного метода Беллмана*. Функции $V^*(x)$ и $V(x)$ будем именовать функциями Беллмана для прямого и обратного счета соответственно. Реализация как обратного, так и прямого метода Беллмана для решения задачи A требует квадратично зависящего от числа состояний системы Ω количества элементарных операций. Иногда прямой и обратный метод Беллмана, подразумевая, что задача решается с использованием соотношений динамического программирования, называют *методами прямого и обратного счета* соответственно.

ПРИМЕР 2. Имеется дискретно управляемая система система ΩI , определяющий ее граф $G(\Omega I)$ представленный на рисунке 3.

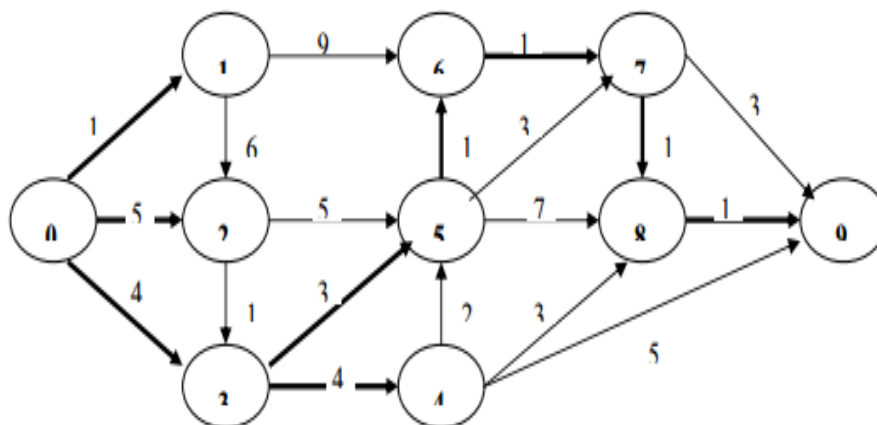


Рисунок 3. – Граф $G(\Omega I)$ с выделенными дугами.

Начальным является состояние 0, множество финальных состояний одноэлементно: $F=\{9\}$.

Методом прямого счета требуется найти полную траекторию минимальной стоимости.

Для состояний системы ΩI (вершин графа $G(\Omega I)$) вычисляем значения прямой функции Беллмана. На основании (6.3) фиксируем $V^*(0)=0$.

Далее, пользуясь (6.4), последовательно получаем:

$$V^*(1)=1;$$

$$V^*(2)=\min[5, V^*(1)+6] = 5;$$

$$V^*(3)=\min[4, V^*(2)+1] = 4;$$

$$V^*(4)=V^*(3)+4 = 8;$$

$$V^*(5)=\min[V^*(2)+5, V^*(3)+3, V^*(4)+2] = 7;$$

$$V^*(6)=\min[V^*(1)+9, V^*(5)+1] =8;$$

$$V^*(7)=\min[V^*(6)+1, V^*(5)+3] =9;$$

$$V^*(8)=\min[V^*(5)+7, V^*(7)+1] =10;$$

$$V^*(9)=\min[V^*(7)+3, V^*(8)+1, V^*(4)+5] =11.$$

При вычислении значений функции Беллмана в графе $G(\Omega I)$ для каждой вершины y , отличной от начальной, специально выделяем (рисуем жирной линией) дугу (x, y) , где x

- вершина для которой значение правой части (6.4) при подсчете $B^*(y)$ оказывается минимальным (этот способ выделения дуг с учетом используемого метода счета называем *прямым*; способ выделения, использованный в решении предыдущего примера, именуем *обратным*). Полученный результат представлен на рисунке 3.

Как показывают выделенные дуги, на последнем такте процесса управления следует выполнить переход из вершины 8 в вершину 9, при этом переход в вершину 8 реализуется из вершины 7, в вершину 7 - из вершины 6, в вершину 6 - из вершины 5, в вершину 5 - из вершины 3, в вершину 3 - из вершины 0. Оптимальной является траектория $\{0,3,5,6,7,8,9\}$.

Полученные в результате решения примеров 1 и 2 оптимальные в системе ΩI траектории совпали.

Отметим следующее обстоятельство. Пусть в процессе вычислений по соотношениям (6.3) - (6.4) для некоторого финального состояния x^* оказалось найденным значение $B^*(x^*)$; полную траекторию стоимости $B^*(x^*)$, имеющую своим конечным состоянием x^* , обозначим T^* . Если каждая из остальных траекторий имеет в своем составе как минимум одно состояние из некоторого подмножества состояний Q , причем все значения $B^*(x)$ при $x \in Q$ уже вычислены и оказались не меньшими $B^*(x^*)$, то T^* - полная оптимальная траектория. Сделанное утверждение нередко дает возможность ограничить объем требуемых вычислений.

Подмножество промежуточных состояний M системы Ω назовем *разрезающим*, если каждая полная траектория T этой системы имеет в своем составе состояния из M .

Разрезающее подмножество M^* назовем *минимальным разрезающим*, если среди его собственных подмножеств нет разрезающих.

Минимальная, из стоимостей полных траекторий системы Ω равна:

$$\min_{x \in M^*} (B^*(x) + B(x)), \quad (6.6)$$

здесь M^* - любое минимальное разрезающее подмножество состояний системы Ω .

Основанный на данном факте метод решения задачи синтеза полной траектории минимальной стоимости называется *методом встречного счета*.

Для системы ΩI , определяемой представленным на рисунке 2 графом $G(\Omega I)$, минимальным разрезающим является, в частности, подмножество вершин-состояний $M^* = \{4, 5, 6\}$. При решении задачи синтеза оптимальной по стоимости полной траектории методом встречного счета выполняем следующие действия.

Прямым счетом последовательно определяем:

$$B^*(1)=1; B^*(2)= 5; B^*(3) = 4; B^*(4)= 8; B^*(5) = 7; B^*(6) = 8.$$

Обратным счетом находим:

$$B(8)=1; B(7) = 2; B(6) =3; B(5)= 4; B(4) = 4.$$

Для принадлежащих подмножеству M^* состояний находим:

$$B^*(4)+ B(4)= 12; B^*(5)+ B(5)=11; B^*(6)+ B(6)= 11.$$

Получаем, что минимальная из стоимостей полных траекторий системы ΩI равна 11.

Оптимальные полные траектории системы ΩI (их, в принципе, может быть несколько) проходят через вершины 5 и 6. При выполненном подсчете значений функции $B^*(x)$, см. решение примера 2, определено, что оптимальной начальной 5-траекторией системы ΩI является $T^*(5) = \{0,3,5\}$, а оптимальной начальной 6-траекторией этой системы является $T^*(6) = \{0,3,5,6\}$. При подсчете значений функции $B(x)$, см. решение примера 1, установлено, что оптимальной заключительной 5-траекторией системы ΩI

является $T(5) = \{5,6,7,8,9\}$, а оптимальной заключительной 6-траекторией этой системы является $T(6) = \{6,7,8,9\}$. Получаем, что единственной оптимальной полной траекторией системы ΩI является $T = \{0,3,5,6,7,8,9\}$.

1.2. Решение дискретных оптимизационных задач методом динамического программирования

Множество вариантов конечно, но число их, как правило, уже при относительно малых размерностях решаемых задач решение оказывается столь большим, что полный перебор практически неосуществим. Использование схемы динамического программирования дает возможность определенным образом упорядочить перебор, существенно сократив при этом число подлежащих рассмотрению вариантов.

Задача оптимального распределения капиталовложений

Пусть имеются денежная сумма (капитал) C и возможные сферы вложений S_1, S_2, \dots, S_n . Считается, что в каждую сферу S_t может быть вложен целочисленный вклад, не превосходящий константы C_t . Для каждой сферы S_t полагается известной монотонно возрастающая в нестрогом смысле функция $f_t(u)$, определяющая величину дохода, получаемого при вложении в данную сферу вклада u .

Требуется определить обеспечивающее максимальный суммарный доход распределение суммы C или ее части по сферам вложений.

Математическая модель задачи следующая:

$$\sum_{i=1}^n f_i(u_i) \rightarrow \max, \quad (6.7)$$

при условиях:

$$\sum_{i=1}^n u_i \leq C;$$

$$u_t \in \{0, 1, \dots, C_t\}, t = \overline{1, n}. \quad (6.8-6.9)$$

Сформулированную задачу можно трактовать как задачу поиска полной траектории, обеспечивающей максимальный суммарный доход соответствующим образом построенной дискретной управляемой системе.

Исходные данные задачи (ЗОРК) определяют систему $\Omega(\text{ЗОРК})$, управления в которой осуществляется в дискретном времени $t=0, 1, 2, \dots, n-1$ (в момент t определяется вклад в сферу S_{t+1}).

Состояния системы - пары (t, U) , где t – момент дискретного времени, а U – сумма, которая к данному моменту еще не распределена, здесь $t \in \{0, 1, \dots, n\}$, $U \in \{0, 1, \dots, C\}$. Начальное состояние системы $(0, C)$; финальными являются состояния вида (n, U) . В любом нефинальном состоянии (t, U) выбор реализуемого в этом состоянии управления u_{t+1} осуществляется из совокупности $V(t, U) = \{0, 1, \dots, \min(C_{t+1}, U)\}$; выбранная величина u_{t+1} - размер вклада в сферу S_{t+1} . Под воздействием управления u_{t+1} система $\Omega(\text{ЗОРК})$ из состояния (t, U) переходит в состояние $(t+1, U - u_{t+1})$, получаемый при этом доход равен $f_{t+1}(u_{t+1})$. Каждая последовательность управлений, переводящая систему из начального состояния в одно из финальных состояний, определяет некоторое допустимое решение ЗОРК (распределение (u_1, u_2, \dots, u_n) капиталовложений).

Исходную ЗОРК трактуем как задачу отыскания траектории системы $\Omega(\text{ЗОРК})$, обеспечивающей максимальный суммарный доход. Функция Беллмана для системы $\Omega(\text{ЗОРК})$ определяется соотношениями:

$$B(n, U) = 0, \quad U \in \{0, 1, \dots, C\}; \quad (6.10)$$

$$B(t, U) = \max_{u_{t+1} \in \{0, 1, \dots, \min(C_{t+1}, U)\}} \{f_{t+1}(u_{t+1}) + B(t+1, U - u_{t+1})\}, \quad (6.11)$$

здесь $t \in \{0, 1, \dots, n-1\}$, $U \in \{0, 1, \dots, C\}$.

При реализации обратного метода Беллмана процесс вычислений реализуется в соответствии с записанными соотношениями. Зная, что $B(n, U) = 0$ при всех U , по формуле (6.11) находим все значения $B(n-1, U)$ при $U \in \{0, 1, \dots, S\}$; здесь оказывается справедливым тождество $B(n-1, U) = f_n(\min(C_n, U))$. Далее, имея вычисленные значения $B(n-1, U)$, по той же формуле (6.11) определяем все значения $B(n-2, U)$, где $U \in \{0, 1, \dots, S\}$, и т.д. до тех пор, пока не окажется найденной величина $B(0, C)$ - оптимальное значение критерия в решаемой ЗОПК. В процессе вычислений, с целью обеспечения возможности определить оптимальное распределение капиталовложений, после отыскания каждого следующего значения $B(t, U)$, где $t \in \{0, 1, \dots, n-1\}$ и $U \in \{0, 1, \dots, C\}$, для пары (t, U) отдельно фиксируем значение u_{t+1} , на котором достигается максимум правой части (6.11).

Прямой метод Беллмана для решения ЗОПК заключается в последовательном вычислении значений функции $B^*(t, U)$ в порядке возрастания значений первого аргумента. При этом:

$$B^*(0, U) = 0, U \in \{0, 1, \dots, C\}; \quad (6.12)$$

$$B^*(t+1, U) = \max_{u_{t+1} \in \{0, 1, \dots, \min(C_{t+1}, U)\}} \{B^*(t, U - u_{t+1}) + f_{t+1}(u_{t+1})\}, \quad (6.13)$$

здесь $t \in \{0, 1, \dots, n-1\}$, $U \in \{0, 1, \dots, C\}$.

Основанные на формуле (6.13) вычисления заканчиваются отысканием величин $B^*(n, U)$, где $U \in \{1, 2, \dots, C\}$; максимальная из найденных величин – оптимальное значение критерия в решаемой ЗОПК.

ПРИМЕР 3. Имеются сферы капиталовложений S_1, S_2, S_3, S_4 ; в каждую из них может быть внесен целочисленный вклад, не превосходящий 5. Величина подлежащего распределению капитала равна 6. Функции, определяющие доходы, получаемые при вложении в различные сферы вклада u , заданы таблицей 1.

Таблица 1.

Значения функций дохода $f_i(u)$.

	$F_1(u)$	$F_2(u)$	$F_3(u)$	$F_4(u)$
$u = 0$	0	0	0	0
$u = 1$	0,2	0	0,1	0
$u = 2$	0,2	0	0,2	0
$u = 3$	0,3	0,4	0,3	0,3
$u = 4$	0,3	0,4	0,4	0,4
$u = 5$	0,5	0,4	0,5	0,5

Требуется найти оптимальное распределение капитала между сферами вложений.

Решение данного примера можно выполнить как прямым, так и обратным методом Беллмана. Будем основываться на рекуррентных соотношениях (6.10) - (6.11), т.е. применим обратный метод. Отыскиваемые в процессе счета значения функции $B(t, U)$ вносим в таблицу 2. В каждую заполняемую клетку (t, U) этой таблицы (параметр t определяет столбец, параметр U – строку) одновременно с $B(t, U)$ вносится и записанное в квадратных скобках значение u_{t+1} , на котором достигается максимум правой части (6.11) при подсчете $B(t, U)$; если таких значений несколько, вносится любое из них (нашей целью является отыскание, какого-либо из оптимальных решений, а не полной их совокупности).

В связи с тем, что $B(4, U)$ тождественно равно нулю, процесс вычислений начинаем с определения значений $B(3, U)$ при всех возможных значениях U . Из равенства (6.11)

следует, что $B(3, U) = f4(U)$ при $U = 0,1,2,3,4,5$ и $B(3, 6) = f4(5)$. Так заполняется последний столбец таблицы 2.

Таблица 2.

Значения функции $B(t, U)$.

	t=0	t=1	t=2	t=3
U=0		0 [0]	0 [0]	0 [0]
U=1		0,1 [0]	0,1 [1]	0 [0]
U=2		0,2 [0]	0,2 [2]	0 [0]
U=3		0,4 [3]	0,3 [3]	0,3 [3]
U=4		0,5 [3]	0,4 [4]	0,4 [4]
U=5		0,6 [3]	0,5 [5]	0,5 [5]
U=6	0,8 [1]	0,7 [3]	0,6 [1]	0,5 [5]

То же равенство (6.11) дает возможность последовательного (справа налево) заполнения остальных столбцов. Заметим, что для решения задачи в первом столбце достаточно найти только один, нижний элемент. В рассматриваемом примере оказалось, что $B(0,6) = 0,8$. Это оптимальное значение критерия (максимально возможный доход). В клетке (0,6) в квадратных скобках записано число 1. Значит, в оптимальном распределении капиталовложений взнос в первую сферу должен равняться единице. В таком случае суммарный взнос в остальные сферы не превосходит 5. В клетке (1,5) в квадратных скобках записано число 3. Значит, в синтезируемом оптимальном распределении капиталовложений взнос во вторую сферу равняется трем. Получаем, что суммарный взнос в третью и четвертую сферы не превосходит 2. В клетке (2,2) в квадратных скобках записано число 2. В оптимальном распределении взнос в третью сферу должен равняться двум. В итоге получаем, что оптимальное распределение делит сумму $C = 6$ между первой, второй и третьей сферами капиталовложений; вносимые в них вклады равны соответственно 1, 3, 2. При этом первая сфера приносит доход 0,2; вторая сфера дает доход 0,4 и третья сфера - доход 0,2. Суммарный доход действительно равен 0,8.

Классическая задача коммивояжера

Имеется полный взвешенный ориентированный граф без петель G с множеством вершин $N = \{1, 2, \dots, n\}$; веса всех дуг неотрицательны. Требуется найти гамильтонов цикл минимального веса.

Исходную информацию по $3K$ считаем представленной в виде $(n \times n)$ - матрицы $S = \{s_{ij}\}$, где s_{ij} – вес дуги (i, j) графа G , $i = 1, n$, $j = 1, n$, $i \neq j$; все элементы главной диагонали матрицы S являются нулями.

В типовой интерпретации вершины $1, 2, \dots, n$ графа G – это города. Дуги отображают возможные элементарные переходы. Коммивояжеру, изначально находящемуся в городе 1 , необходимо обойти все остальные города, побывав в каждом из них ровно по одному разу, и затем вернуться в город 1 . Веса дуг графа трактуются как длины соответствующих элементарных переходов. Требуется найти имеющий минимальную длину допустимый (т.е. удовлетворяющий наложенным требованиям) маршрут коммивояжера.

С учетом других возможных интерпретаций, на матрицу S требование симметричности не налагается, не считается обязательным и выполнение неравенства треугольника.

Один из основных алгоритмов решения $3K$ основан на принципе динамического программирования. При изложении этого алгоритма будем придерживаться терминологии, соответствующей приведенной типовой интерпретации задачи.

Пусть i – произвольный город ($i \in N$), а V – любое подмножество городов, не содержащее города 1 и города i . Через $M(i, V)$ обозначим совокупность путей, каждый из которых начинается в городе i , завершается в городе 1 и проходит в качестве промежуточных только через города множества V , заходя в каждый из них ровно по одному разу. Через $B(i, V)$ обозначим длину кратчайшего пути множества $M(i, V)$.

Для решаемой задачи $B(i, V)$ – функция Беллмана. Как очевидно, $B(1, \{2, 3, \dots, n\})$ – искомая минимальная длина простого (без самопересечений) замкнутого пути, проходящего через все города.

Если V – одноэлементное множество, $V = \{j\}$, где $j \neq 1$ и $j \neq i$, то совокупность $M(i, V)$ состоит из единственного пути $\pi = (i, j, 1)$. Поэтому:

$$B(i, \{j\}) = s_{ij} + s_{j1}, i \in N, j \in \{2, 3, \dots, n\}, j \neq i. \quad (6.18)$$

Предположим, что значения функции $B(i, V)$ для всех $i \in N \setminus \{1\}$ и всех возможных k -элементных ($k < n-1$) множеств V уже вычислены. Тогда значение $B(i, V')$, где V' – произвольное $(k+1)$ -элементное подмножество совокупности $N \setminus \{1, i\}$, вычисляется по формуле:

$$B(i, V') = \min_{j \in V'} (s_{ij} + B(j, V' \setminus \{j\})). \quad (6.19)$$

Уравнения (6.18)-(6.19) – рекуррентные соотношения динамического программирования для решения задачи коммивояжера, они реализуют обратный метод Беллмана.

ПРИМЕР 4. Решить задачу коммивояжера, определяемую матрицей:

$$S = \begin{pmatrix} 0 & 5 & 1 & 4 \\ 2 & 0 & 3 & 5 \\ 1 & 4 & 0 & 2 \\ 3 & 5 & 4 & 0 \end{pmatrix}$$

Сначала, пользуясь формулой (6.18), определяем значения $B(i, \{j\})$:

$$\begin{aligned} B(2, \{3\}) &= 4; & B(2, \{4\}) &= 8; & B(3, \{2\}) &= 6; \\ B(3, \{4\}) &= 5; & B(4, \{2\}) &= 7; & B(4, \{3\}) &= 5. \end{aligned}$$

Далее по формуле (6.19) последовательно получаем (в левой части каждого из ниже записанных равенств подчеркнуты те значения параметра j , на которых при подсчете реализуется указанный в правой части (6.19) минимум):

$$\begin{aligned} B(2, \{3, 4\}) &= \min(\underline{3} + 5, 5 + 5) = 8; \\ B(3, \{2, 4\}) &= \min(4 + \underline{8}, 2 + 7) = 9; \\ B(4, \{2, 3\}) &= \min(5 + 4, \underline{4} + 6) = 9; \\ B(1, \{2, 3, 4\}) &= \min(5 + 8, 1 + 9, 4 + 9) = 10. \end{aligned}$$

Итак, оптимальное значение критерия в рассматриваемом примере равно 10. Выполненные подчеркивания позволяют определить оптимальный маршрут. Он следующий:

$$1 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1.$$

Для записи соотношений, по которым реализуется прямой метод Беллмана, введем новые обозначения. Пусть $M'(V, i)$ – совокупность путей, каждый из которых начинается в городе 1 , проходит в качестве промежуточных только через города подмножества V , заходя в каждый из них ровно по одному разу, и завершается в городе i ; здесь, как и ранее, i – произвольный город ($i \in N$), а V – любое подмножество N , не содержащее городов 1 и i . Длину кратчайшего пути множества $M'(V, i)$ обозначим $B^*(V, i)$. Как очевидно, $B^*(\{2, 3, \dots, n\}, 1)$ – искомая минимальная длина простого (без самопересечений) замкнутого пути, проходящего через все города. Если V – одноэлементное множество, $V = \{j\}$, где $j \neq 1$ и $j \neq i$, то совокупность $M'(V, i)$ состоит из единственного пути $\pi = (1, j, i)$. Поэтому:

$$B^*(\{j\}, i) = s_{1j} + s_{ji}, i \in N, j \in \{2, 3, \dots, n\}, j \neq i. \quad (6.20)$$

Предположим, что значения функции $B^*(V, i)$ для всех $i \in N$ и всех возможных k -элементных ($k < n-1$) множеств V уже вычислены. Тогда значение $B^*(V', i)$, где V' – произвольное $(k+1)$ -элементное подмножество совокупности $N \setminus \{1, i\}$, вычисляется по формуле:

$$B^*(V', i) = \min_{j \in V'} (B^*(V' \setminus \{j\}, j) + s_{ji}). \quad (6.21)$$

Уравнения (6.20)-(6.21) – рекуррентные соотношения динамического программирования для решения классической задачи коммивояжера, они реализуют прямой метод Беллмана. Заметим, что при решении задачи коммивояжера обратным методом состояния моделирующей системы – пары вида (i, V) , где i – город, в котором сейчас находится коммивояжер, а V – множество городов, которые ему еще предстоит обойти перед тем, как вернуться в исходный город 1. При решении задачи коммивояжера прямым методом состояния моделирующей системы – пары вида (V, i) , где i – город, в котором коммивояжер находится сейчас, а V – множество городов, которые коммивояжер уже обошел после того, как вышел из города 1. По сути, (i, V) и $(\{2, 3, \dots, n\} \setminus (V \cup \{i\}), i)$ – это разные обозначения одной и той же ситуации. Данный факт используется при решении рассматриваемой задачи методом встречного счета.

ПРИМЕР 1.5. Методом встречного счета решить задачу коммивояжера, определяемую матрицей:

$$S = \begin{pmatrix} 0 & 5 & 1 & 4 \\ 2 & 0 & 3 & 5 \\ 1 & 4 & 0 & 2 \\ 3 & 5 & 4 & 0 \end{pmatrix}$$

В качестве разрезающего выберем множество, включающее следующие состояния:

- (2, {3}) /совпадет с ({4},2)/;
- (2, {4}) /совпадет с ({3},2)/;
- (3, {2}) /совпадет с ({4},3)/;
- (3, {4}) /совпадет с ({2},3)/;
- (4, {2}) /совпадет с ({3},4)/;
- (4, {3}) /совпадет с ({2},4)/;

Реализуя метод обратного счета, определяем:

$$B(2, \{3\}) = 4; B(2, \{4\}) = 8; B(3, \{2\}) = 6; B(3, \{4\}) = 5; B(4, \{2\}) = 7; B(4, \{3\}) = 5.$$

Реализуя метод прямого счета, получаем:

$$B^*(\{4\}, 2) = 9; B^*(\{3\}, 2) = 5; B^*(\{4\}, 3) = 8; B^*(\{2\}, 3) = 8; B^*(\{3\}, 4) = 3; B^*(\{2\}, 4) = 10.$$

Далее имеем:

$$\begin{aligned} B(2, \{3\}) + B^*(\{4\}, 2) &= 13; \\ B(2, \{4\}) + B^*(\{3\}, 2) &= 13; \\ B(3, \{2\}) + B^*(\{4\}, 3) &= 14; \\ B(3, \{4\}) + B^*(\{2\}, 3) &= 13; \\ B(4, \{2\}) + B^*(\{3\}, 4) &= 10; \\ B(4, \{3\}) + B^*(\{2\}, 4) &= 15. \end{aligned}$$

Минимальное, равное 10, значение правая часть принимает в пятом равенстве. Отсюда следует:

- а) оптимальное значение критерия в рассматриваемой задаче равно 10;
- б) оптимальная траектория системы проходит через состояние $(\{3\}, 4)$ или, что то же самое, через состояние $(4, \{2\})$. Последнее дает возможность установить оптимальный

маршрут коммивояжера:

$$1 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1.$$

Применение метода встречного счета в сравнении с прямым или обратным счетом тем более выгодно, чем больше размер решаемой задачи.

Классическая задача о ранце (КЗР).

Имеются ранец и предметы $П1, П2, \dots, Пn$; каждый предмет $Пi$ характеризуется двумя целыми положительными показателями: c_i – стоимость предмета; v_i – вес предмета, $i = 1, n$. Предметы недробимы, т.е. каждый из них можно поместить в ранец только целиком. Требуется найти совокупность предметов, которые следует поместить в ранец с тем, чтобы суммарная стоимость предметов в ранце оказалась максимальной при условии, что суммарный вес предметов в ранце не может превысить заданной натуральной константы W (предполагается, что суммарный вес всех имеющихся предметов больше W). КЗР формулируется как следующая задача булева линейного программирования (БЛП):

$$\sum_{i=1}^n c_i x_i \rightarrow \max \tag{6.29}$$

$$\sum_{i=1}^n v_i x_i \leq W;$$

$$x_i \in \{0, 1\}, i = \overline{1, n}. \tag{1.30-1.31}$$

Задачу (1.29)-(1.31) будем именовать задачей Z .

По начальным данным задачи Z формулируем совокупность частных задач о ранце $Z(k, p)$, где $k \in \{1, 2, \dots, n\}$; $p \in \{1, 2, \dots, W\}$.

Задача $Z(k, p)$ записывается следующим образом:

$$\sum_{i=1}^k c_i x_i \rightarrow \max$$

при условиях:

$$\sum_{i=1}^k v_i x_i \leq p;$$

$$x_i \in \{0, 1\}, i = \overline{1, k}.$$

(в частной задаче $Z(k, p)$ имеющимися в наличии считаются только первые k предметов из совокупности $\{П1, П2, \dots, Пn\}$, суммарный вес предметов в ранце не может превышать константы p). Оптимальное значение критерия в частной задаче $Z(k, p)$ обозначим $B^*(k, p)$. Как очевидно, задача $Z(n, W)$ совпадает с исходной задачей Z ; поэтому $B^*(n, W)$ – оптимальное значение критерия в задаче Z .

Легко видеть, что:

$$B^*(1, p) = \begin{cases} 0 & \text{при } p \in \{0, 1, 2, \dots, v_1 - 1\}; \\ c_1 & \text{при } p \in \{v_1, v_1 + 1, \dots, W\}. \end{cases} \tag{6.32}$$

Пусть для некоторого k , принадлежащего множеству $\{1, 2, \dots, n-1\}$ и всех $p \in \{1, 2, \dots, W\}$ значения функции $B^*(k, p)$ уже вычислены. При $p \in \{0, 1, 2, \dots, vk+1-1\}$ ситуация, возникающая при нахождении $B^*(k+1, p)$, фактически не отличается от ситуации, имевшей место при подсчете $B^*(k, p)$, ибо дополнительно появившийся при переходе от

задачи $Z(k, p)$ к задаче $Z(k+1, p)$ предмет $П_{k+1}$ в ранец поместить невозможно. В случае, когда $p \in \{vk+1, vk+1+1, \dots, W\}$, при переходе от задачи $Z(k, p)$ к задаче $Z(k+1, p)$ мы получаем дополнительную возможность поместить в ранец предмет $П_{k+1}$. Этой возможностью можно: а) не воспользоваться; б) воспользоваться. Очевидно, что в случае а) мы получим уже вычисленную величину $B^*(k, p)$. В случае б) в ранец кладется предмет $П_{k+1}$ стоимостью $ck+1$ и так как вес этого предмета равен $vk+1$, то далее мы оказываемся в условиях задачи $Z(k, p - vk+1)$. В итоге получаем:

$$B^*(k, p), \text{ если } p \in \{0, 1, 2, \dots, vk+1-1\};$$

$$B^*(k+1, p) =$$

$$\max \{ B^*(k, p), c_{k+1} + B^*(k, p - vk+1) \} \text{ при } p \in \{vk+1, vk+1+1, \dots, W\}; \quad (6.33)$$

Равенства (6.32) - (6.33) – рекуррентные соотношения динамического программирования для решения задачи о ранце. Пользуясь ими, мы находим сначала величины $B^*(1, p)$, $p \in \{1, 2, \dots, W\}$, затем величины $B^*(2, p)$, $p \in \{1, 2, \dots, W\}$, и т.д., вплоть до отыскания $B^*(n, W)$ – оптимального значения критерия в исходной задаче Z . Процедуру вычисления значений функции $B^*(k, p)$ целесообразно реализовывать как процесс последовательного заполнения строк таблицы стоимостей. В этой таблице строки соответствуют значениям параметра k (по возрастанию), а столбцы – значениям параметра p (также по возрастанию). В каждую клетку, расположенную в пересечении произвольной строки k и произвольного столбца p , вносится значение $B^*(k, p)$. Первая строка заполняется согласно формуле (6.32). Далее заполнение каждой $(k+1)$ -ой строки, $k = 1, 2, \dots, n-1$, осуществляется в соответствии с формулой (1.33). Одновременно целесообразно фиксировать множества предметов $M(k, p)$, которые обеспечивают значения $B^*(k, p)$ критериев рассматриваемых частных задач. Как очевидно, $M(1, p) = \emptyset$, если $p < v_1$; $M(1, p) = \{1\}$ в противном случае. Далее можно считать $M(k, p)$, если $B^*(k+1, p) = B^*(k, p)$; $M(k+1, p) = M(k, p - vk+1) \cup \{k+1\}$ в противном случае. В заполненной таблице стоимостей содержимое клетки (n, W) – оптимальное значение критерия в задаче Z . Для обеспечения этого значения в ранец следует поместить предметы совокупности $M(n, W)$. Изложенный алгоритм решения КЗР далее будем называть алгоритмом $Акзр$.

ПРИМЕР 6. Решить КЗР:

$$5x_1 + 2x_2 + 4x_3 + 7x_4 + 5x_5 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 \leq 8;$$

$$x_i \in \{0, 1\}, i=1, 5.$$

Задача решается путем последовательного, сверху вниз, заполнения строк таблицы стоимостей (таблица 4).

Таблица 4.

Таблица стоимостей

	1	2	3	4	5	6	7	8
1	0	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)
2	2 (2)	5 (1)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)
3	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	11 (1,2,3)	11 (1,2,3)	11 (1,2,3)
4	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	12 (1,4)	14 (1,2,4)	16 (1,3,4)
5	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	12 (1,4)	14 (1,2,4)	16 (1,3,4)

В каждой клетке (k, p) данной таблицы вслед за полученным значением $B^*(k, p)$ в скобках малыми цифрами перечисляются номера предметов, которые согласно оптимальному

решению частной задачи $Z(k, p)$ следует положить в ранец. Если задача $Z(k, p)$ имеет несколько оптимальных решений, то указывается только одно из них.

Заполнение первой строки таблицы выполняется в соответствии с формулой (1.32); здесь считается, что в нашем распоряжении имеется только предмет III , его стоимость равна 5 и вес равен 2. Предмет III можно положить в ранец, если дозволённый вес предметов в ранце не меньше чем 2. Поэтому $B^*(1, 1) = 0$ и $B^*(1, p) = 5$, если $p \geq 2$.

Заполнение второй и каждой из нижеследующих строк выполняется в соответствии с формулой (1.33). Рассмотрим в качестве иллюстрации, причем на содержательном уровне, как заполняется клетка (4, 8). Данная клетка соответствует задаче $Z(4, 8)$. В сравнении с задачей $Z(3, 8)$, которой соответствует клетка (3, 8), в $Z(4, 8)$ появляется новая возможность – в ранец можно положить предмет IV .

Согласно формуле (1.33), выбирается лучший из двух вариантов.

Первый вариант предполагает, что возможностью положить в ранец предмет IV мы не воспользовались. Тогда имеет место абсолютно та же ситуация, что в задаче $Z(3, 8)$, и мы можем обеспечить суммарную стоимость предметов в ранце равную $B^*(3, 8)$, т.е. числу 11, записанному в клетке, расположенной непосредственно над заполняемой. Итак, первый вариант обеспечивает суммарную стоимость предметов в ранце, равную 11; реализация данного варианта означает, что в ранец кладутся предметы III , II и III .

Второй вариант предусматривает, что предмет IV , его стоимость 7, кладется в ранец. Вес предмета IV равен 4. Поэтому от заполняемой клетки в четвертой строке перемещаемся на 4 клетки влево, оказываемся в столбце, соответствующем оставшемуся резерву веса ранца после того, как в него положен предмет IV ; непосредственно над полученной клеткой, а именно в клетке (3, 4) записано число 9, это максимально возможная суммарная стоимость предметов, взятых из совокупности $\{III, II, III\}$, при условии, что их суммарный вес не превышает четырех (в этой ситуации как показывает заполнение клетки (3, 4), из указанной совокупности следует взять первый и третий предмет).

Таким образом, второй вариант обеспечивает суммарную стоимость предметов в ранце $7 + 9 = 16$; реализация данного варианта означает, что в ранец кладутся предметы III , III и IV . Второй вариант дает лучший результат, заполнение клетки (4, 8) осуществляется по этому варианту.

Задача отыскания кратчайших путей (ЗОКП).

Считается заданным конечный взвешенный ориентированный граф G с множеством вершин $N = \{1, 2, \dots, n\}$, веса всех дуг неотрицательны. Веса дуг трактуем как их длины. Последовательность i_1, i_2, \dots, i_k вершин графа G определяет путь из вершины i_1 в вершину i_k , если для каждого $t = 1, 2, \dots, k-1$ в данном графе имеется дуга (i_t, i_{t+1}) ; указанные дуги образуют путь i_1, i_2, \dots, i_k . Сумма длин дуг, образующих путь, называется длиной этого пути. Для каждой вершины x графа требуется найти путь минимальной длины (кратчайший путь) из вершины 1 в вершину x . Вес дуги (i, j) графа G обозначаем $c(i, j)$. Длину кратчайшего пути из 1 в x будем называть расстоянием от вершины 1 до вершины x . Расстояние от вершины 1 до вершины x будем обозначать $s(x)$. Ясно, что $s(1) = 0$.

Пусть H – множество вершин, длины кратчайших путей из вершины 1 в которые уже известны. В начальной ситуации это множество одноэлементно: $H = \{1\}$. Через $s(1, M)$ обозначим минимальную из длин кратчайших путей от вершины 1 до вершин множества M , $M \subseteq \{2, 3, \dots, n\}$. Как очевидно, для множества вершин H , содержащего вершину 1, имеет место:

$$s(I, N \setminus H) = \min_{i \in H, j \in N \setminus H} (s(i) + c(i, j)). \quad (6.40)$$

Пусть минимум правой части (6.40) реализуется на паре (i^0, j^0) . Тогда кратчайший путь из вершины I в вершину j^0 получаем добавлением к кратчайшему пути из вершины I в вершину i^0 дуги (i^0, j^0) , длина $s(j^0)$ этого пути равна $s(i^0) + c(i^0, j^0)$.

Формула (6.40) – рекуррентное соотношение динамического программирования для решения ЗОКП. Пользуясь этой формулой, на первом шаге определяем ближайшую к вершине I вершину i^1 из совокупности $\{2, 3, \dots, n\}$. На втором шаге определяем ближайшую к вершине I вершину i^2 из совокупности $\{2, 3, \dots, n\} \setminus \{i^1\}$. На третьем шаге определяем ближайшую к вершине I вершину i^3 совокупности $\{2, 3, \dots, n\} \setminus \{i^1, i^2\}$, и т.д. В процессе счета строится дерево D кратчайших путей из вершины I в остальные вершины. Корнем D является вершина I ; если на произвольном шаге алгоритма минимум правой части (6.40) реализуется на паре (i^0, j^0) , к конструируемому дереву добавляется ребро (i^0, j^0) . Ход выполняемых вычислений можно оформить в виде процесса заполнения таблицы, строки которой соответствуют вершинам графа G (i -ая строка – вершине $i, i=1, n$). Таблица заполняется по столбцам, слева направо (каждый столбец для определенности заполняется сверху вниз). При этом первый (левый) столбец таблицы имеет имя " I ". При заполнении столбца " I " в клетку, стоящую в пересечении с j -ой строкой, $j \in \{2, 3, \dots, n\}$, вносится $c(I, j)$, если дуга (I, j) в графе G имеется, и $+\infty$, если такой дуги в графе нет. После заполнения этого столбца, в нем символом * отмечается наименьший элемент (в случае, когда таких элементов несколько, отмечается любой из них). Если отмеченный символом * элемент равен u и находится в строке номер v , то длина кратчайшего пути от вершины I до вершины v найдена, $s(v) = u$.

Первый шаг работы алгоритма реализован. Каждый следующий шаг алгоритма предусматривает заполнение очередных двух столбцов таблицы. Результатом каждого следующего шага является определение расстояния от вершины I до еще одной вершины. После того, как очередное расстояние $s(x)$ определилось, соответствующая строка x таблицы считается отмеченной, дальнейшее заполнение клеток отмеченных строк не производится. В момент, когда строка становится отмеченной, во всех незаполненных ее клетках ставим специальный символ z . Изначально отмеченной считается первая строка таблицы, ибо значение $s(I)$ известно с самого начала: $s(I) = 0$.

На втором шаге алгоритма второй столбец таблицы, а на последующих шагах – каждый следующий четный столбец, приобретает название последней вершины x графа G , для которой реализуемая вычислительная процедура в результате выполнения предшествующего шага определила расстояние от I до x .

В заголовках первого, второго и далее каждого следующего четного столбца рядом с его именем " x " указывается значение $s(x)$. При заполнении второго и каждого следующего четного столбца " x " в каждую клетку, находящуюся в пересечении с неотмеченной строкой j вносится сумма $s(x) + c(x, j)$; если в графе G дуга (x, j) отсутствует, в клетку вносится символ $+\infty$.

Третий и далее каждый следующий нечетный столбец имеет наименование " \min ". В каждую клетку такого столбца, стоящую в пересечении с j -ой неотмеченной строкой, вносится минимальное из расстояний, записанных в двух соседних слева клетках данной строки.

После заполнения каждого очередного столбца " \min ", в нем отмечается символом * наименьший элемент (если таких элементов несколько, отмечается любой из них). Если отмеченный символом * элемент последнего заполненного столбца " \min " равен u и

находится в строке номер v , то, согласно (6.40), считаем найденной длину кратчайшего пути от вершины 1 до вершины v , $s(v)=u$. С этого момента v -ая строка таблицы переходит в число отмеченных строк, дальнейшее заполнение ее клеток производиться не будет. Следующий подлежащий заполнению столбец четный, ему назначается имя " v ", в заголовке этого столбца указывается также найденное значение $s(v)$. Далее заполняется очередной столбец "min". В результате заполнения каждой следующей пары столбцов определяется расстояние от вершины 1 до еще одной вершины графа, таким образом завершается реализация следующего шага алгоритма. Изначально известно, что $s(1) = 0$. Поэтому общее число заполняемых пар столбцов $n-1$. Заполнение каждого очередного столбца требует выполнения линейно зависящего от n числа элементарных операций. Таким образом, реализация изложенного алгоритма требует выполнения квадратично зависящего от n числа элементарных операций.

ПРИМЕР 7. Взвешенный ориентированный граф G задан матрицей:

$$M = \begin{matrix} & \begin{matrix} 0 & 9 & 7 & \infty & 2 & 1 & 9 \end{matrix} \\ \begin{matrix} \infty & 0 & 4 & \infty & 3 & 8 & 5 \\ 6 & 3 & 0 & 1 & \infty & \infty & 6 \\ \infty & 1 & \infty & 0 & 4 & 5 & 6 \\ \infty & 9 & 1 & 3 & 0 & 7 & \infty \\ \infty & 5 & 9 & 9 & \infty & 0 & 2 \\ 9 & 8 & 7 & 6 & \infty & 5 & 0 \end{matrix} \end{matrix}$$

Числовой элемент m_{ij} этой матрицы равен весу дуги (i,j) графа G ; если дуга в графе отсутствует, то $m_{ij} = \infty$. Веса дуг трактуются как их длины.

Требуется найти расстояния от вершины 1 до остальных вершин графа.

Ход выполняемых в процессе решения вычислений отображаем как последовательное заполнение столбцов таблицы 5.

Таблица 5.

Таблица определения расстояний

	"1" 0	"6" 1	min	"5" 2	min	"3" 3	Min	"7" 3	min	"4" 4	min
1	z	z	z	z	z	z	Z	z	z	z	z
2	9	9	9	11	9	6	6	11	6	5	5*
3	7	8	7	3	3*	z	Z	z	z	z	z
4	∞	7	7	5	5	4	4	9	4*	z	z
5	2	∞	2*	z	z	z	z	z	z	z	z
6	1*	z	z	z	z	z	z	z	z	z	z
7	9	3	3	∞	3	9	3*	z	z	z	z

В итоге получаем, что в рассматриваемом графе G расстояния от вершины 1 до вершин $2, 3, 4, 5, 6$ и 7 равны соответственно $5, 3, 4, 2, 1$ и 3 . Отметим, что ранее для определяющего дискретную управляемую систему Ω графа $G(\Omega)$ был изложен другой, основанный на рекуррентных соотношениях (6.3)-(6.5) и использующий ацикличность графа, алгоритм поиска кратчайших путей.

Действительно, если в графе $G(\Omega)$ вес каждой дуги (x, y) трактуется как длина соответствующего одношагового перехода, то значение функции Беллмана $B^*(y)$ – длина кратчайшего пути из начальной вершины в вершину y .

Выше отмечалось, что метод динамического программирования относят к числу схем инвариантного погружения. Суть указанных схем в том, что каждая подлежащая решению задача заменяется упорядоченным семейством одноподобных задач, последней из которых является решаемая. Далее в порядке возрастания значений одного или нескольких параметров последовательно решаются задачи введенного семейства; знание результатов решения предыдущих задач существенно облегчает решение следующих; в итоге решенной оказывается исходная задача. Удачная замена исходной задачи семейством порождаемых ею задач – главное условие эффективности такого подхода.

Классическая задача о назначениях (КЗН).

Имеются множество исполнителей $I = \{1, 2, \dots, n\}$, множество работ $R = \{r_1, r_2, \dots, r_n\}$ и $(n \times n)$ -матрица численных оценок (например, производительностей) $A = \{a_{ij}\}$, где a_{ij} – оценка закрепления исполнителя i за работой r_j , $i=1, n, j=1, n$.

Назначения – взаимно однозначные отображения множества $\{1, 2, \dots, n\}$ в себя; назначение π исполнителю i предписывает работу $r_{\pi(i)}$; численной оценкой такого закрепления является $a_{i\pi(i)}$ – элемент матрицы A , здесь $i=1, 2, \dots, n$. Каждое назначение π оцениваем по критерию:

$$K(\pi) = \sum_{i=1}^n a_{i\pi(i)}$$

в указанной интерпретации значение этого критерия – суммарная производительность исполнителей при назначении π .

Требуется найти назначение, при котором суммарная производительность исполнителей максимальна.

КЗН формулируется в виде:

$$\max_{\pi} \sum_{i=1}^n a_{i\pi(i)}$$

Рассмотрим вопрос о решении записанной *КЗН*, обозначим ее символом Z , методом динамического программирования. С этой целью введем в рассмотрение совокупность частных задач $Z(i, W_i)$, формулируемых по исходным данным задачи Z ; здесь $i \in \{1, 2, \dots, n\}$, а W_i – i -элементные подмножества совокупности $\{1, 2, \dots, n\}$.

В задаче $Z(i, W_i)$ между исполнителями $\{1, 2, \dots, i\}$ следует распределить совокупность работ, индексы которых перечислены в W_i ; критерий задачи прежний – суммарная производительность имеющихся исполнителей. Оптимальное значение критерия в задаче $Z(i, W_i)$ обозначим $B^*(i, W_i)$. В таком случае $B^*(n, \{1, 2, \dots, n\})$ – оптимальное значение критерия в исходной задаче Z .

Как очевидно:

$$B^*(1, \{j\}) = a_{1j} \text{ для всех } j \in \{1, 2, \dots, n\}. \tag{6.44}$$

Если $i > 1$, имеем:

$$B^*(i, W_i) = \max_{j \in W_i} (a_{ij} + B^*(i-1, W_i \setminus \{j\})), \quad i = 2, 3, \dots, n; \tag{6.45}$$

здесь W_i – произвольное i -элементное подмножество из совокупности $\{1, 2, \dots, n\}$.

Записанные равенства (6.44)-(6.45) являются соотношениями динамического программирования для решения *КЗН*. Легко видеть, что оценкой числа элементарных операций, выполняемых основанным на соотношениях (6.44) - (6.45) алгоритмом, является функция, которая экспоненциально зависит от n . В отличие от всех остальных рассматриваемых в данном разделе задач, для *КЗН* имеются алгоритмы, обладающие

существенно лучшими в сравнении с процедурой счета по соотношениям динамического программирования, характеристиками быстродействия.

2. Задание на практическую работу

1) Имеются сферы капиталовложений $S1, S2, S3, S4$; в каждую из них может быть внесен целочисленный вклад, не превосходящий 5. Величина подлежащего распределению капитала равна 12. Функции, определяющие доходы, получаемые при вложении в имеющиеся сферы вклада u , заданы приведенной таблицей. Прямым методом Беллмана требуется найти оптимальное распределение капитала между сферами вложений.

Таблица значений функций дохода $f_t(u)$.

	$f_1(u)$	$f_2(u)$	$f_3(u)$	$f_4(u)$
$u = 0$	0	0	0	0
$u = 1$	0,1	0,1	0,2	0
$u = 2$	0,3	0,1	0,2	0
$u = 3$	0,3	0,4	0,3	0,3
$u = 4$	0,4	0,4	0,3	0,4
$u = 5$	0,5	0,4	0,3	0,4

2) Автомобиль, начиная от базы и заканчивая на базе замкнутый маршрут, должен доставить с базы в точки T_1, T_2, \dots, T_m некоторые грузы. Известно, что вес направляемого в точку T_i груза равен $c_i, i = 1, m$. Выполняя тот же маршрут, автомобиль должен собрать в точках R_1, R_2, \dots, R_n грузы, адресованные базе. Известно, что вес груза, направляемого из точки R_j на базу, равен $w_j, j = 1, n$. Через каждую из перечисленных точек маршрут должен проходить однократно. Матрица расстояний S размера $(m+n+1) \times (m+n+1)$ задана. Требуется записать рекуррентные соотношения динамического программирования для определения маршрута, минимального по количеству выполняемых тонно-километров.

3) Записать рекуррентные соотношения динамического программирования для классической задачи о ранце, дополненной условием: из каждой пары предметов $(P1, Pn), (P2, Pn-1), \dots, (Pk, Pn-k+1)$ в ранец можно поместить только один предмет, здесь $k \leq n/2$.

4) Решить задачу о ранце:

$$7x_1 + 4x_2 + 4x_3 + 5x_4 + 2x_5 + 3x_6 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 + 2x_6 \leq 10;$$

$$x_i \in \{0, 1\}, i = \overline{1, 5}.$$

5) Решить следующую модификацию классической задачи о ранце:

$$5x_1 + 2x_2 + 4x_3 + 7x_4 + 5x_5 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 \leq 12;$$

$$x_i \in \{0, 1\}, i = 1, 2, 5;$$

$$x_i \in \{0, 1, 2\}, i = 3, 4.$$

6) Задача о путешественнике. На местности имеется сеть дорог, связывающих несколько населенных пунктов. Путешественник находится в пункте a_0 , из которого, двигаясь по одной из трех дорог, можно попасть в пункты a_1, a_2, a_3 . Из каждого пункта

опять выходят ровно три дороги, ведущие в a_4, a_5, a_6 . Из них – в a_7, a_8, a_9 и так далее, вплоть до конечных пунктов $b_1=a_3N-2, b_2=a_3N-1, b_3=a_3 \cdot N$. Длины всех дорог заданы. Найти наиболее короткий путь из a_0 в один из конечных пунктов. Решить задачу при $N=5$. Оцените количество операций сложения и сравнения при ее решении по методу Беллмана, а также при полном переборе всех путей.

7) Задача о распределении инвестиций. Нужно распределить между N предприятиями сумму a , выделенную для их инвестирования. Известно, что вложение средств в размере y в k -ое предприятие обеспечивает прибыль в размере $d_k(y)$. Целью распределения является получение максимального суммарного дохода. Решить задачу при $N=4, a=300$ при условии, что суммы инвестиций всегда кратны 50, а функции $d_k(y)$ для $y=50j$ ($j=0, 1, \dots, 6$) принимают значения, заданные в таблице.

Таблица

Значения функции $d_k(y)$							
y	0	50	100	150	200	250	300
$d_1(y)$	0	50	120	140	150	200	250
$d_2(y)$	0	60	130	140	130	160	200
$d_3(y)$	0	30	60	100	130	200	250
$d_4(y)$	0	40	100	110	120	160	220

8) Имеется множество недобрымых камней $\{K_1, K_2, \dots, K_8\}$. Веса камней (в порядке возрастания индексов) следующие: 7, 5, 8, 4, 3, 6, 3, 1. Требуется разбить совокупность $\{K_1, K_2, \dots, K_8\}$ на два подмножества так, чтобы суммарный вес камней одного подмножества минимально отличался от суммарного веса камней другого подмножества.

9) Задача о замене оборудования. Частное предприятие планирует в течение N лет заниматься выпуском изделий, используя некоторое оборудование. В начале, можно либо купить новое оборудование возраста $x_0 = 0$ лет и стоимостью p , либо подержанное оборудование возраста $x_0 > 0$ лет по его ликвидной стоимости $\phi(x_0)$. Показатели эксплуатации оборудования включают: $f(t)$ – стоимость произведенных за год изделий на оборудовании возраста t лет; $r(t)$ – затраты на эксплуатацию в течение года оборудования возраста t лет.

В процессе эксплуатации оборудование можно менять, продавая старое по ликвидной стоимости $\phi(t)$ и покупая новое стоимостью p . В конце N -го года оборудование продается по ликвидной стоимости. Определить оптимальный возраст оборудования x_0 при начальной покупке и оптимальный график его замены. Выполнить расчеты при $N = 8, x_0 \in \{0, 1, 2\}$,

$$f(t) = 30 - t/2, r(t) = 13 + t/2, p = 17, \phi(t) = \begin{cases} 6 & \text{при } 0 \leq t \leq 6 \\ 2 & \text{при } 7 \leq t \leq 10 \end{cases}$$

10) Задача о загрузке судна. Судно, имеющее грузоподъемность a , загружается предметами N типов. Один предмет i -го типа имеет стоимость y_i и вес z_i . Требуется найти вариант загрузки судна, при котором стоимость взятых на борт предметов максимальна.

Решить задачу для $N = 3, a = 200, y_1 = 25, y_2 = 40, y_3 = 80, z_1 = 40, z_2 = 50, z_3 = 70$.

3. Вопросы к практической работе

- 1) Каковы особенности применения генетических алгоритмов
- 2) Опишите операторы скрещивания в задаче коммивояжера.
- 3) Опишите операторы мутации в решении задачи коммивояжера
- 4) Раскрыть принцип динамического программирования.
- 5) Описать прямой метод Беллмана
- 6) Описать обратный метод Беллмана
- 7) Основные рекуррентные соотношения динамического программирования.

- 8) Каким образом будет выглядеть формула метода динамического программирования для решения задачи о ранце?
- 9) В чем заключается задача коммивояжера?
- 10) Приведите детерминированный алгоритм решения задачи о назначениях.
- 11) Что является в задаче о назначениях геном, генотипом и особью?
- 12) Как осуществляется генерация начальной популяции в задаче оптимального распределения капиталовложений?

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие / А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Методы оптимизации в бизнес-аналитике

Профиль подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ
МЕТОДЫ ОДНОМЕРНОГО И МНОГОМЕРНОГО БЕЗУСЛОВНОГО ПОИСКА

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

МЕТОДЫ ОДНОМЕРНОГО И МНОГОМЕРНОГО БЕЗУСЛОВНОГО ПОИСКА:
методические указания для выполнения лабораторных работ по дисциплине
«МЕТОДЫ ОПТИМИЗАЦИИ В БИЗНЕС-АНАЛИТИКЕ» для студентов
направления 09.04.03 «Прикладная информатика» направленность «Искусственный
интеллект и бизнес-аналитика» / Составитель: Т. В. Хоменко. – Ульяновск: УлГТУ,
2021. – 24с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и бизнес-
аналитика». Цель данного практикума – ориентировать студентов на содержание
и порядок выполнения лабораторных работ во время прохождения студентами
курсов «Искусственный интеллект и бизнес-аналитика». Даются задания для
лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание

ВВЕДЕНИЕ	4
Лабораторная работа №1	5
Тема: Методы одномерного поиска.....	5
1. Краткие теоретические сведения.....	5
2. Задание на лабораторную работу	15
3. Вопросы к лабораторной работе.....	15
Лабораторная работа № 2	17
Тема: Методы многомерного безусловного поиска: методы прямого поиска.....	17
1. Краткие теоретические сведения.....	17
2. Задание на лабораторную работу	22
3. Вопросы к лабораторной работе.....	22
Список литературы	23

ВВЕДЕНИЕ

В учебно-методическом пособии описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.).

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Лабораторная работа №1

Тема: Методы одномерного поиска

Цель работы: знакомство с оптимизационными задачами, изучение различных методов одномерной оптимизации и сравнение эффективности их применения для конкретных целевых функций.

1. Краткие теоретические сведения

1.1 Постановка задачи одномерной безусловной оптимизации

Задача поиска минимума целевой функции формулируется в виде:

$$x^* = \arg \min f(x), x \in X,$$

где X – множество допустимых точек, среди которых ищется точка x^* , доставляющая минимум $f(x)$ целевой функции.

Другая распространенная запись задачи минимизации

$$\begin{aligned} \min f(x) \\ x \in X \end{aligned}$$

Когда $X=R$, мы имеем дело с одномерной безусловной задачей минимизации, т.е. когда целевая функция $f(x)$ имеет только один простой аргумент и область X есть вся вещественная ось чисел.

В методах одномерной оптимизации вместо $X=R$ рассматривается отрезок $X=[a,b]$, содержащий искомое решение x^* . Такой отрезок называется отрезком неопределенности, или отрезком локализации. Относительно целевой функции $f(x)$ часто предполагается, что она унимодальная.

Определение: Функция $f(x)$ называется унимодальной на $X=[a,b]$, если существует такая точка $x^* \in X$, что:

$$\begin{aligned} f(x_1) > f(x_2), \text{ если } x_1 < x_2 < x^*, x_1, x_2 \in X, \\ f(x_1) < f(x_2), \text{ если } x^* < x_1, x_2, x_1, x_2 \in X. \end{aligned}$$

Если ограничиваться рассмотрением лишь непрерывных функций $f(x)$, то свойство унимодальности функции попросту означает наличие у нее единственного локального минимума и этот минимум достигается в точке $x=x^*$.

В ряде методов относительно целевой функции $f(x)$ предполагается, что она выпуклая на X .

Определение: Функция $f(x)$ называется выпуклой на $x=[a,b]$, если $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$ при любых $x_1, x_2 \in X$ и всех $\alpha, 0 \leq \alpha \leq 1$.

Если при любых $x_1, x_2 \in X$ неравенство будет строгим, то функция $f(x)$ называется строго выпуклой.

1.2 Алгоритм пассивного поиска минимума

Пусть N - число точек, в которых необходимо провести вычисления целевой функции $f(x)$, т.е. N экспериментов. Точки, в которых необходимо провести эксперименты, определяются следующим образом:

$$\left\{ \begin{array}{l} \text{Если } N = 2k - 1 - \text{нечётное, то} \\ x_i = a + \frac{b-a}{N+1} \cdot i \\ i = 1, 2, 3, \dots, N \\ \text{Если } N = 2k - \text{чётное, то} \\ x_{2i} = a + \frac{b-a}{k+1} \cdot i \\ x_{2i-1} = x_{2i} - \delta \\ i = 1, 2, \dots, k \end{array} \right.$$

Среди вычисленных значений $\{f(x_i)\}$ ($i=1, N$), ищется точка x_j , в которой достигается минимум:

$$f(x_j) = \min_{1 \leq i \leq N} f(x_i), \text{ где } 1 \leq i \leq N$$

Найденная точка принимается за приближенное решение задачи $\tilde{x} = x_j$. Исходный отрезок неопределенности $[a, b]$ после экспериментов в N точках сужается до $[x_{j-1}, x_{j+1}]$, длина которого равна:

$$L_N = L_N(x_1, x_2, \dots, x_N) = \max_{1 \leq i \leq N} (x_{i+1} - x_{i-1}) = x_{j+1} - x_{j-1} = \begin{cases} 2 \frac{b-a}{N+1}, & \text{если } N = 2k - 1 \\ \frac{b-a}{N/2+1} + \delta, & \text{если } N = 2k \end{cases}$$

Точность найденного решения \tilde{x} равна половине отрезка неопределенности, т.е. $|x^* - \tilde{x}| \leq \varepsilon$, где $\varepsilon = \frac{1}{2} L_N$ и x^* - точное решение.

1.3 Алгоритмы активного поиска минимума

В алгоритмах активного поиска очередная точка, в которой производится эксперимент, выбирается с учетом информации, полученной в предыдущих опытах. Рассмотрение этих алгоритмов начнем с методов блочного поиска, которые сочетают в себе пассивные и последовательные стратегии поиска. При этом вычисления в точках объединены в блоки, в каждом из которых проводится одновременно n_i экспериментов), общее число экспериментов будет $\sum_{i=1}^m n_i = N$, т.е. блок - это совокупность из нескольких экспериментов, которые проводятся одновременно (пассивный поиск). Результаты, полученные в $(i-1)$ -м блоке, становятся известны перед началом экспериментов в i -м блоке $\{x_{ij} \ (j=1, 2, \dots, n_i)\}$ (последовательный поиск). Если размеры блоков равны единице, т.е. $n_i=1$, то мы имеем обычный последовательный алгоритм поиска.

1.3.1 Алгоритм равномерного блочного поиска

Схема алгоритма

Шаг 1. Задаются исходный отрезок неопределенности $[a, b]$, ε - точность приближенного решения \tilde{x} , число экспериментов в блоке - n (нечетное, $n=2k-1$). Проводим эксперимент в середине отрезка $[a, b]$, т.е. вычисляем $y_k = f(x_k)$, где $x_k = (a+b)/2$.

Шаг 2. Проводим эксперименты в остальных точках блока: $y_i = f(x_i)$, где $x_i = a + i \cdot (b-a)/(n+1)$, $i=1, 2, \dots, n$, $i \neq k$. Находим точку x_j , в которой достигается минимум среди вычисленных значений: $f(x_j) = \min f(x_i)$, следовательно, точное значение минимума x^*

содержится на отрезке $[x_{j-1}, x_{j+1}]$.

Шаг 3. Полагаем $a=x_j-1$, $b=x_j+1$, $x_k=x_j$, $y_k=y_j$. Если $b-a \leq 2\varepsilon$, то $\tilde{x} = x_k$, $\tilde{y} = y_k$ и поиск заканчивается. Иначе перейти к шагу 2. Если заданная точность ε достигнута после m итераций, т.е. после экспериментов в m блоках, то длина отрезка неопределенности после всех N вычислений ($N=n+(m-1)(n-1)=(n-1)m+1$) будет:

$$L_N = \left(2 \frac{b-a}{n+1}\right)^m \text{ и } |x^* - \tilde{x}| \leq \frac{1}{2} L_N.$$

1.3.2 Алгоритм деления интервала пополам

Это вариант предыдущего алгоритма при $n=3$.

Схема алгоритма.

Шаг 1. Задаются a, b, ε . Производим эксперимент в точке $x_2=(a+b)/2$, т.е. вычисляем $y_2=f(x_2)$.

Шаг 2. Проводим эксперименты в остальных точках блока: $x_1=(a+x_2)/2$, $y_1=f(x_1)$, $x_3=(x_2+b)/2$, $y_3=f(x_3)$.

Находим x_j такую, что $f(x_j)=\min \{f(x_i)\}$, где $1 \leq i \leq 3$

Тогда точное решение x^* содержится на отрезке $[x_{j-1}, x_{j+1}]$. Предполагается $x_0 = a, x_4 = b$.

Шаг 3. Полагаем $a=x_j-1$, $b=x_j+1$, $x_2=x_j$, $y_2=y_j$. Если $b-a \leq 2\varepsilon$, то $\tilde{x} = x_2$, $\tilde{y} = y_2$ и поиск заканчивается. Иначе перейти к шагу 2.

После k итераций общее число проведенных экспериментов равно $N=2k+1$, а длина получившегося отрезка неопределенности будет $L_N = \frac{b-a}{2^k} = \frac{b-a}{2^{\lfloor N/2 \rfloor}}$, где $\lfloor z \rfloor$ – целая часть числа z . Следовательно, достигнутая точность будет $|x^* - \tilde{x}| \leq \varepsilon$, $\varepsilon = 1/2L_N$.

1.3.3 Метод дихотомии

Это алгоритм блочного поиска для $n_i=n=2$, т.е. когда в блоке два эксперимента. Так как пассивная составляющая алгоритма, т.е. блок, содержит четное число экспериментов, то оптимальный выбор точек x_{ij} , в которых необходимо провести эксперименты, будет неравномерным, в отличие от предыдущих алгоритмов, где число экспериментов в блоке было нечетным и, соответственно, расположение точек равномерным. Если блок содержит два эксперимента, то оптимальное (дельта оптимальное) расположение точек, в которых будут проводиться эксперименты, это как можно ближе к середине отрезка. Такое расположение точек позволяет получить наименьший отрезок неопределенностей после экспериментов в блоке.

Схема алгоритма.

Шаг 1. Задаются a, b, ε и δ - малое положительное число, значительно меньшее чем ε .

Шаг 2. Определяется середина отрезка $x=(a+b)/2$. Производятся эксперименты в двух точках близких середине: $y_1=f(x-\delta)$, $y_2=f(x+\delta)$.

Шаг 3. Определяется следующий отрезок локализации, т.е. определяется какой из отрезков $[a, x+\delta]$ или $[x-\delta, b]$ содержит точное решение x^* . Если $y_1 \leq y_2$, то это отрезок $[a, x+\delta]$ и $b=x+\delta$, иначе это отрезок $[x-\delta, b]$ и $a=x-\delta$, т.е. выбранный отрезок локализации мы снова обозначили как $[a, b]$.

Шаг 4. Если $b-a \leq 2\varepsilon$, то $x=(a+b)/2$, $\tilde{y} = f(\tilde{x})$ и поиск заканчивается. Иначе перейти к шагу 2.

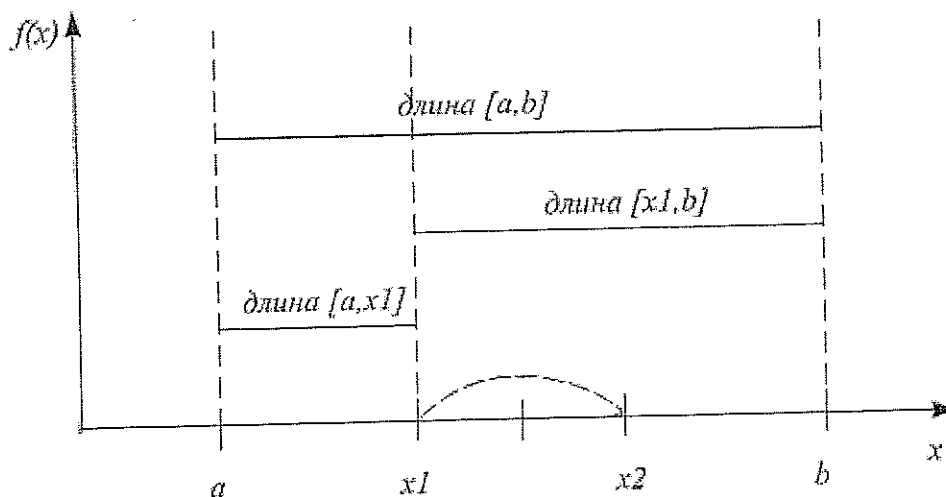
После k итераций общее число экспериментов будет $N=2k$, а длина получившегося отрезка неопределенности $L_N < \frac{b-a}{2^{N/2}} + \delta$. Следовательно, $|x^* - \tilde{x}| < \frac{1}{2} L_N$.

А теперь перейдём к рассмотрению чисто последовательных методов поиска.

1.3.4 Метод золотого сечения

Для того чтобы уменьшить отрезок неопределённости $[a, b]$, нам необходимо вычислить значение целевой функции $f(x)$, по крайней мере, в двух точках на отрезке $[a, b]$.

В результате этих двух экспериментов отрезок неопределённости сузится до отрезка $[a, x_2]$ или $[x_1, b]$.



Так как у нас нет никаких оснований предпочесть один из этих вариантов, то точки x_1 и x_2 должны быть симметричны относительно середины отрезка $[a, b]$. В этом случае длины отрезков $[a, x_2]$ и $[x_1, b]$ будут равны. Таким образом, остаётся вопрос, как выбрать точку x_1 .

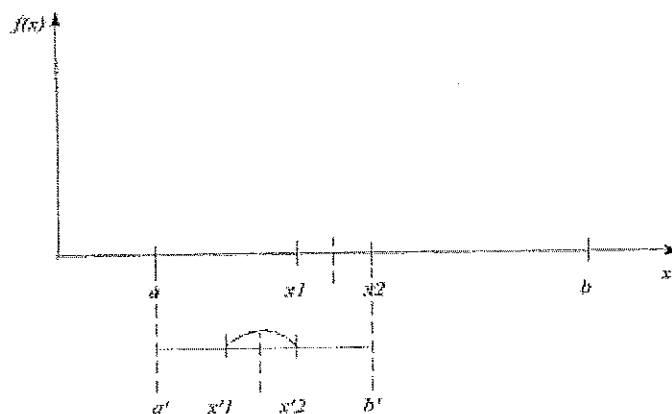
В методе золотого сечения точка x_1 выбирается из соображения, что должно выполняться соотношение:

$$\frac{\text{длина}[a, b]}{\text{длина}[x_1, b]} = \frac{\text{длина}[x_1, b]}{\text{длина}[a, x_1]} = \lambda = 1.618033989\dots$$

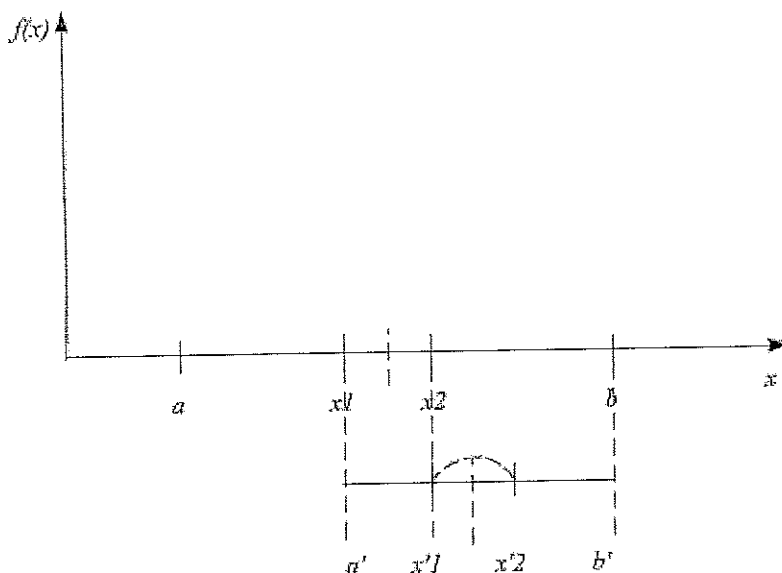
т.е. точка x_1 делит отрезок $[a, b]$ по правилу «золотого сечения», где λ - есть «золотое отношение». Точка x_2 определяется как точка симметричная к x_1 относительно середины отрезка.

В результате экспериментов у нас получается отрезок неопределённости $[a, x_2]$, содержащий точку x_1 , или отрезок неопределённости $[x_1, b]$, содержащий точку x_2 . Оказывается, что остающаяся точка на суженном отрезке неопределённости делит его вновь по правилу «золотого сечения». Следовательно, чтобы, в свою очередь, уменьшить новый отрезок неопределённости, нам не достаёт одного эксперимента, а именно, вычисления целевой функции в точке, симметричной к оставшейся точке относительно середины этого нового отрезка. Всё продемонстрировано на рисунке,

а)



б)



где буквы со штрихами обозначают новый отрезок неопределённости. Вариант а) соответствует случаю, если новым отрезком неопределённости будет $[a, x_2]$, а вариант б) – отрезку $[x_1, b]$.

В приводимой ниже схеме алгоритма остающиеся отрезки неопределённости переименовываются каждый раз как $[a, b]$, а точки, в которых проводятся эксперименты на этом отрезке, обозначается через x_1 и x_2 , причём $x_1 < x_2$. Кроме того, y_1 и y_2 имеют следующие значения: $y_1 = f(x_1)$ и $y_2 = f(x_2)$.

Схема алгоритма

Шаг1. Задаются a, b, ε и $\lambda = 1.618\dots$ Вычисляют

$$x_1 = b - \frac{b-a}{\lambda}, x_2 = a + \frac{b-a}{\lambda}, y_1 = f(x_1), y_2 = f(x_2).$$

Шаг2. а) Если $y_1 \leq y_2$, то полагают $b = x_2, x_2 = x_1, y_2 = y_1$ и вычисляют $x_1 = a + b - x_2, y_1 = f(x_1)$.

б) Если $y_1 > y_2$, то полагают $a = x_1, x_1 = x_2, y_1 = y_2$ и вычисляют $x_2 = a + b - x_1, y_2 = f(x_2)$.

Шаг3. Если $b - a > \varepsilon$, то переходят к шагу 2. Иначе если $y_1 < y_2$, то полагают $\tilde{x} = x_1$

и $\tilde{y} = y_1$, если $y_1 \geq y_2$, то полагают $\tilde{x} = x_2$ и $\tilde{y} = y_2$.

Закончить поиск.

После каждой итерации длина отрезка неопределённости уменьшается в λ раз. Так как первая итерация начинается после двух экспериментов, то после N экспериментов длина отрезка неопределённости будет

$$L_N = \frac{b-a}{\lambda^{N-1}}.$$

1.3.4 Метод чисел Фибоначчи

Этот метод применяется, когда число экспериментов N заранее задано. Метод чисел Фибоначчи, также как и метод золотого сечения относится к симметричным методам, т.е. точки, в которых выполняются два эксперимента, на основе которых происходит уменьшение отрезка неопределённости, расположены симметрично относительно середины отрезка. Вот только выбор точки x_1 происходит на основе других соотношений. Для этого используются числа Фибоначчи: $F_0, F_1, F_2, F_3, \dots$, где $F_i = F_{i-2} + F_{i-1}$ ($i = 2, 3, \dots$) и $F_0 = F_1 = 1$.

Точка x_1 определяется из соотношения:

$$\frac{\text{длина } [a, x_1]}{\text{длина } [a, b]} = \frac{F_{N-2}}{F_N},$$

т.е. $x_1 = a + (b-a) \frac{F_{N-2}}{F_N}$. Точка x_1 делит $[a, b]$ на две неравные части. Отношение

малого отрезка к большему равно F_{N-2}/F_{N-1} . Точка x_2 определяется как точка, симметричная к x_1 относительно середины отрезка $[a, b]$. Поэтому

$$x_2 = b - (b-a) \frac{F_{N-2}}{F_N} = a + (b-a) \frac{F_{N-1}}{F_N}. \text{ При этом будет выполняться условие } x_1 < x_2.$$

В результате экспериментов в точках x_1 и x_2 у нас получится отрезок неопределённости $[a, x_2]$, содержащий точку x_1 , или отрезок неопределённости $[x_1, b]$, содержащий точку x_2 . Остающаяся точка делит новый отрезок неопределённости на две неравные части в отношении:

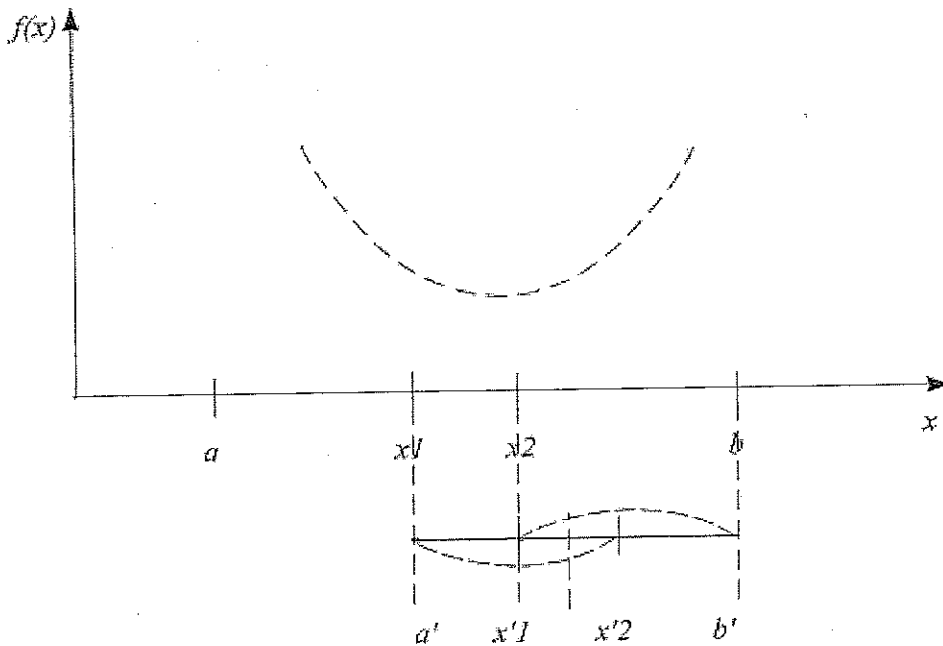
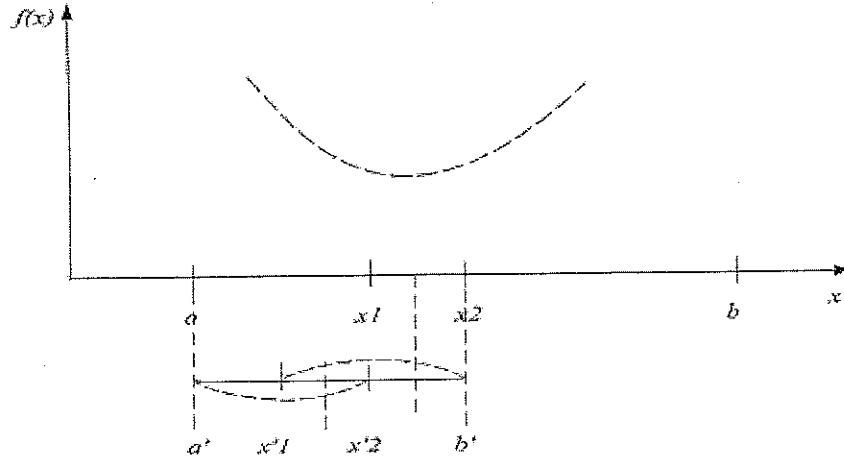
$$\frac{\text{меньшая часть}}{\text{большая часть}} = F_{N-3}/F_{N-2}.$$

То есть в методе Фибоначчи остающаяся точка делит отрезок на две неравные части в пропорциях определяемых числами Фибоначчи. Так на k -ом шаге это отношение равно

$$\frac{\text{меньшая часть}}{\text{большая часть}} = \frac{F_{N-k-2}}{F_{N-k}}, \text{ а длины отрезков равны: } \text{меньшая часть} = \frac{F_{N-k-1}}{F_N}(b-a) \text{ и}$$

$$\text{большая часть} = \frac{F_{N-k}}{F_N}(b-a). \text{ Всё это показано на рисунке:}$$

a)



б)

$$\frac{\text{длина } [a, x_1]}{\text{длина } [x_1, b]} = \frac{F_{N-2}}{F_{N-1}}$$

$$\frac{\text{длина } [x'_2, b']}{\text{длина } [a', x'_2]} = \frac{F_{N-3}}{F_{N-2}}$$

$$\frac{\text{длина } [a', x'_1]}{\text{длина } [x'_1, b']}} = \frac{F_{N-3}}{F_{N-2}}$$

Для того чтобы в свою очередь уменьшить получившийся отрезок неопределённости, надо определить симметричную точку относительно середины отрезка и произвести эксперимент в ней. Этот процесс продолжается, пока не будет проведено N экспериментов.

Схема алгоритма

Шаг 1. Задаются a, b, N . Вычисляются числа Фибоначчи F_0, F_1, \dots, F_N . Определяется:

$$x_1 = a + (b-a)F_{N-2}/F_N,$$

$$x_2 = a + (b-a)F_{N-1}/F_N,$$

$$y_1 = f(x_1), y_2 = f(x_2).$$

Шаг 2. а) Если $y_1 \leq y_2$, то полагают $b = x_2, x_2 = x_1, y_2 = y_1$ и вычисляют $x_1 = a + b - x_2, y_1 = f(x_1)$.

б) Если $y_1 > y_2$, то полагают $a = x_1, x_1 = x_2, y_1 = y_2$ и вычисляют $x_2 = a + b - x_1, y_2 = f(x_2)$.

Повторить шаг 2 $N-2$ раза.

Шаг 3. Если $y_1 < y_2$, то полагают $\tilde{x} = x_1$ и $\tilde{y} = y_1$. Если $y_1 \geq y_2$, то полагают $\tilde{x} = x_2$ и $\tilde{y} = y_2$.

Закончить поиск.

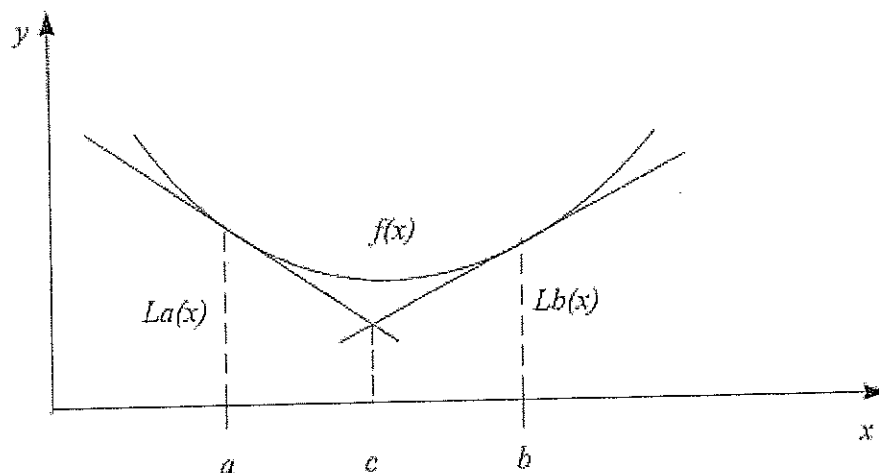
Длина отрезка неопределённости в методе Фибоначчи $L_N \approx (b-a)/F_N$.

1.4 Методы поиска, основанные на аппроксимации целевой функции

Суть этих методов заключается в том, что по полученной в ходе вычислений информации строится аппроксимирующая функция и её минимум принимается за точку очередного вычисления. Такие методы дают хорошие результаты при минимизации достаточно гладких унимодальных функций.

1.4.1 Метод касательных

Пусть функция $f(x)$ выпукла и дифференцируема на $[a, b]$. Идея метода состоит в следующем. Пусть $[a, b]$ - отрезок неопределённости и $f(a), f'(a), f(b), f'(b)$ - результаты вычислений в точках a и b . По этой информации строится аппроксимирующая функция, представляющую из себя кусочно-линейную функцию, состоящую из касательной $L_a(x) = f(a) + f'(a)(x-a)$ к $f(x)$ в точке a и касательной $L_b(x) = f(b) + f'(b)(x-b)$ к $f(x)$ в точке b .



Полученная аппроксимирующая функция есть ломанная состоящая из прямой $L_a(x)$ на $[a, c]$ и $L_b(x)$ на $[c, b]$, где c - точка пересечения касательных. Легко заметить, что при $f(a) < 0$ и $f(b) > 0$ минимум аппроксимирующей функции достигается в точке c . Значение точки пересечения c можно определить по формуле

$$c = \frac{(bf'(b) - af'(a)) - (f(b) - f(a))}{f'(b) - f'(a)}$$

В точке c производятся вычисления $f(c)$ и $f'(c)$. Если $f'(c) = 0$, то решением задачи будет $x^* = c$. Если же $f'(c) > 0$, то в качестве следующего отрезка неопределённости будет $[a, c]$. Если же $f'(c) < 0$, то – отрезок $[c, b]$. Процесс повторяется до тех пор, пока $f'(c) = 0$ или отрезок неопределённости не достигнет заданной точности.

Схема алгоритма

Шаг 1. Заданы a, b, ε . Вычислить $y_1 = f(a), y_2 = f(b), z_1 = f'(a), z_2 = f'(b)$.

Шаг 2. Если $b - a \leq 2\varepsilon$, то полагаем $\tilde{x} = (a + b)/2, \tilde{y} = f(\tilde{x})$. Поиск окончен. Если $b - a > 2\varepsilon$, то вычислить $c = \frac{(bz_2 - az_1) - (y_2 - y_1)}{z_2 - z_1}, y = f(c), z = f'(c)$. Если $z = 0$, то полагаем $\tilde{x} = c, \tilde{y} = y$ и поиск окончен. Если $z < 0$, то $a = c, y_1 = y, z_1 = z$. Если $z > 0$, то $b = c, y_2 = y, z_2 = z$. Повторить шаг 2.

1.4.2 Метод парабол

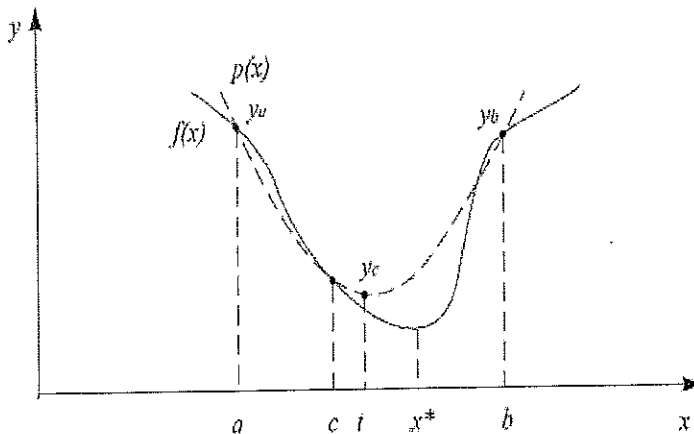
Рассмотрим алгоритм квадратичной интерполяции или метод парабол, т.е. в качестве аппроксимирующей функции используется парабола. Для однозначного задания параболы необходимы три точки. Пусть имеются три точки, для которых выполняется $a < c < b, f(c) \leq f(a), f(c) \leq f(b)$. Так как $[a, b]$ – отрезок неопределённости и $f(x)$ – унимодальная функция, то найти такую точку c нетрудно. Парабола, проходящая через три точки $(a, f(a)), (c, f(c)), (b, f(b))$, имеет вид

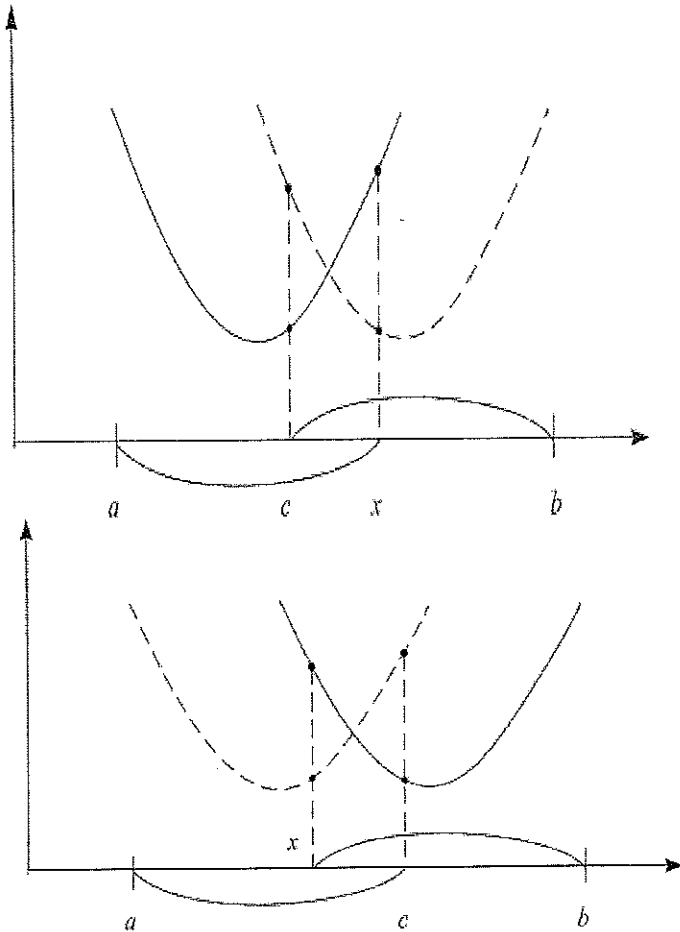
$$P(x) = \left(\frac{f(b) - f(c)}{b - c} + \frac{f(a) - f(c)}{a - c} \right) \frac{(x - c)(x - b)}{b - a} + \frac{f(b) - f(c)}{b - c} (x - c) + f(c)$$

Поскольку f – унимодальная функция, то хотя бы одно из неравенств $f(c) \leq f(a), f(c) \leq f(b)$ строгое и, следовательно, коэффициент при старшем члене $P(x)$ положителен. Тогда $P(x)$ достигает минимума в точке

$$t = c + \frac{1}{2} \frac{(b - c)^2 (f(a) - f(c)) - (c - a)^2 (f(b) - f(c))}{(b - c)(f(a) - f(c)) + (c - a)(f(b) - f(c))},$$

причем $(a + c)/2 \leq t \leq (c + b)/2$. Эта точка и выбирается в качестве точки очередного вычисления значения функции.





Если оказалось, что $t=c$, условимся в качестве точки очередного вычисления выбирать точку $(a+c)/2$. Итак, следующее вычисление проводится в точке

$$x = \begin{cases} t, & \text{если } t \neq c \\ (a+c)/2, & \text{если } t = c. \end{cases}$$

Определим новый отрезок неопределенности с лежащей внутри него точкой, для которой выполняются условия аналогичные условиям, которым удовлетворяла точка c . В силу унимодальности функции f и в зависимости от выполнения или невыполнения условий $x < c, f(x) < f(c), f(x) = f(c)$ это будут отрезки с точкой внутри: $[a, c]$ и x ; $[x, b]$ и c ; $[x, c]$ и $(x+c)/2$; $[a, x]$ и c ; $[c, b]$ и x ; $[c, x]$ и $(x+c)/2$. Смотри рисунок. Далее строится парабола, определяется ее минимум, и т.д., до тех пор, пока длина отрезка неопределенности не удовлетворяет заданной точности.

Схема алгоритма

Шаг 1. Задаются a, c, b и ε . Вычислить $y_a = f(a), y_c = f(c), y_b = f(b)$.

Шаг 2. Вычислить $x = \begin{cases} t, & \text{если } t \neq c \\ (a+c)/2, & \text{если } t = c. \end{cases}, y = f(x)$, где

$$t = c + \frac{1}{2} \cdot \frac{(b-c)^2(y_a - y_c) - (c-a)^2(y_b - y_c)}{(b-c)(y_a - y_c) + (c-a)(y_b - y_c)}.$$

Шаг 3. А) При $x < c$.

Если $y < y_c$, то $b=c, c=x, y_b=y_c, y_c=y$.

Если $y > y_0$, то $a=x$, $y_a=y$.

Если $y=y_0$, то $a=x$, $b=c$, $c=(x+c)/2$, $y_a=y$, $y_b=y_0$, $y_c=f(c)$.

Б) При $x > c$.

Если $y < y_0$, то $a=c$, $c=x$, $y_a=y_0$, $y_c=y$.

Если $y > y_0$, то $b=x$, $y_b=y$.

Если $y=y_0$, то $a=c$, $b=x$, $c=(x+c)/2$, $y_a=y_0$, $y_b=y$, $y_c=f(c)$.

Шаг 4. Если $b-a \leq \varepsilon$, то закончить поиск, положив $\tilde{x} = x$, $\tilde{y} = y$, иначе перейти к шагу 2.

2. Задание на лабораторную работу

2.1. Изучить предлагаемые методы одномерной безусловной оптимизации.

2.2. В соответствии с вариантом задания, определенным преподавателем, найти точку минимума функции $f(x)$ на отрезке $[a, b]$ одним из рассмотренных методов (составить программу, реализующую метод поиска).

№	Целевая функция	Отрезок $[a, b]$	Точность ε или число экспериментов N
1	$x^2 + 6 \cdot e^{0,15x}$	$[-1, 0]$	$N=22$
2	$x^2 + 4 \cdot e^{-0,25x}$	$[0, 1]$	$N=23$
3	$x^4 + 0,4 \cdot \arctg 5x$	$[-1, 0]$	$N=20$
4	$x^4 - 1,5 \arctg x$	$[0, 1]$	$N=21$
5	$x^2 + 8 \cdot e^{0,55x}$	$[-2, 0]$	$\varepsilon=10^{-3}$
6	$-4x + e^{ x-0,2 }$	$[0, 2]$	$\varepsilon=1,5 \cdot 10^{-3}$
7	$1,4x + e^{ x-2 }$	$[0, 2]$	$\varepsilon=5 \cdot 10^{-3}$
8	$x^2 + e^x$	$[-1, 0]$	$N=18$
9	$ x + e^{10x}$	$[-1, 0]$	$\varepsilon=10^{-3}$
10	$10 \cos x + e^x$	$[0, 3]$	$\varepsilon=5 \cdot 10^{-4}$

а) Если в варианте задания указано число экспериментов N , то сравнить заданные в варианте методы по получаемой длине отрезка неопределенности.

б) Если же указана точность искомого решения, то сравнить методы по количеству экспериментов, понадобившихся для достижения заданной точности.

2.3. Оформить отчет о выполнении задания с приведением условия задачи (алгоритмов и программ указанных методов поиска), таблицы результатов сравнения рассмотренных методов, заключения по результатам сравнения методов.

3. Вопросы к лабораторной работе

1) Приведите математическую модель задачи безусловной оптимизации функции одной переменной.

2) Сформулируйте необходимые условия оптимальности в задачах безусловной оптимизации функции одной переменной.

3) Представьте алгоритм классического метода в задачах безусловной оптимизации функции одной переменной.

4) Охарактеризуйте сущность прямых методов минимизации функции одной переменной.

5) Представьте характеристику методов исключения отрезков.

6) Дайте характеристику метода деления отрезка пополам (дихотомии).

7) Представьте алгоритм метода деления отрезка пополам (дихотомии).

- 8) Дайте характеристику метода золотого сечения.
- 9) Представьте алгоритм метода золотого сечения.
- 10) Перечислите методы безусловной минимизации функции одной переменной с использованием производных.
- 11) Дайте характеристику метода касательных.
- 12) Приведите алгоритм метода касательных.
- 13) Какие существуют достоинства и недостатки метода касательных?

Лабораторная работа № 2

Тема: Методы многомерного безусловного поиска: методы прямого поиска

Цель работы: знакомство с методами многомерной безусловной оптимизации нулевого порядка и их освоение, сравнение эффективности применения этих методов конкретных целевых функций.

1. Краткие теоретические сведения.

1.1 О численных методах многомерной оптимизации

Задача многомерной безусловной оптимизации формулируется в виде:

$$\min_{x \in X} f(x),$$

где $x = \{x^{(1)}, x^{(2)}, \dots, x^{(n)}\}$ – точка в n -мерном пространстве $X = \mathbb{R}^n$, то есть целевая функция $f(x) = f(x^{(1)}, \dots, x^{(n)})$ – функция n аргументов.

Так же как и в первой лабораторной работе мы рассматриваем задачу минимизации. Численные методы отыскания минимума, как правило, состоят в построении последовательности точек $\{x_k\}$, удовлетворяющих условию $f(x_0) > f(x_1) > \dots > f(x_n)$. Методы построения таких последовательностей называются методами спуска. В этих методах точки последовательности $\{x_k\}$ вычисляются по формуле:

$$x_{k+1} = x_k + \alpha_k p_k, \quad k=0,1,2,\dots,$$

где p_k – направление спуска, α_k – длина шага в этом направлении.

Различные методы спуска отличаются друг от друга способами выбора направления спуска p_k и длины шага α_k вдоль этого направления.

1.2 Общая характеристика

Методы прямого поиска – это методы, в которых используются только значения целевой функции (методы нулевого порядка). Рассмотрим следующие методы, основанные на эвристических соображениях. Эти методы довольно часто применяются на практике, позволяя в ряде случаев получить удовлетворительные решения.

Основное достоинство методов нулевого порядка состоит в том, что они не требуют непрерывности целевой функции и существования производных.

1.3 Метод конфигураций (метод Хука и Дживса)

Алгоритм включает в себя два основных этапа поиска.

- В начале обследуется окрестность выбранной точки (базисной точки), в результате находится приемлемое направление спуска;
- Затем в этом направлении находится точка с наименьшим значением целевой функции. Таким образом находится новая базисная точка.

Эта процедура продолжается пока в окрестностях базисных точек удастся находить приемлемые направления спуска.

Схема алгоритма

Шаг 1.

Задаются начальное приближение (первая базисная точка):

$$x_0 = \{x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)}\}$$

начальный шаг h для поиска направления спуска, точность решения δ (предельное значение для шага h). Присваивается $k=0$.

Шаг 2. (Первый этап).

Определяется направление минимизации целевой функции $f(x)=f(x^{(1)},x^{(2)},\dots,x^{(n)})$ в базисной точке

$$x_0 = \{x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(n)}\}$$

Для этого последовательно дают приращение переменным $x^{(j)}$ в точке x_k . Присвоим $z=x_k$. Циклически даем приращение переменным $x^{(j)}$ и формируем $z^{(j)}=x_k^{(j)}+h$, если $f(z)<f(x_k)$, если же нет, то $z^{(j)}=x_k^{(j)}-h$, если $f(z)<f(x_k)$, иначе $z^{(j)}=x_k^{(j)}$. Так для всех $j(j=1,2,\dots,n)$.

Шаг 3.

Если $z=x_k$, то есть не определилось подходящее направление, то обследование окрестности базисной точки x_k повторяется, но с меньшим шагом h (например, $h=h/2$).

Если $h>\delta$, то перейти к шагу 2, то есть повторить обследование точки x_k .

Если $h\leq\delta$, то поиск заканчивается, то есть достигнуто предельное значение для шага h и найти приемлемое направление спуска не удастся. В этом случае полагаются:

$$\tilde{x} = x_k, \tilde{y} = f(x_k)$$

Шаг 4. (Второй этап).

Если $z\neq x_k$, то требуется найти новую базисную точку в направлении вектора:

$$z-x_k: x_{k+1}=x_k + \lambda(z-x_k), \text{ где } \lambda - \text{коэффициент «ускорения поиска»}.$$

Определяется такое значение $\lambda=\lambda_k$, при котором достигается наименьшее значение целевой функции в выбранном направлении, то есть функции:

$$f(x_k + \lambda(z-x_k)) = \varphi(\lambda).$$

В зависимости от способа выбора λ_k возможны варианты метода:

- а) $\lambda_k=\lambda=\text{const}$ постоянная для всех итераций;
- б) задается начальное $\lambda_0=\lambda$, а далее $\lambda_k=\lambda_{k-1}$, если $f(x_{k+1})<f(x_k)$, иначе дробим λ_k , пока не выполнится это условие;
- в) λ_k определяется решением задачи одномерной минимизации функции $\varphi(\lambda)$.

Таким образом, определяется новая базисная точка $x_{k+1}=x_k + \lambda(z-x_k)$. Полагаем $k=k+1$ и поиск оптимального решения повторяется с шага 2.

1.4 Метод симплекса

Под симплексом понимается n -мерный выпуклый многогранник n -мерного пространства, имеющий $n+1$ вершину. Для $n=2$ это треугольник, а при $n=3$ это тетраэдр.

Идея метода состоит в сравнении значений функции в $n+1$ вершинах симплекса и перемещении симплекса в направлении лучшей точки. В рассматриваемом методе симплекс перемещается с помощью операций отражения. Далее принято следующее: $x_0(k), x_1(k), \dots, x_n(k)$ – вершины симплекса, где k – номер итерации.

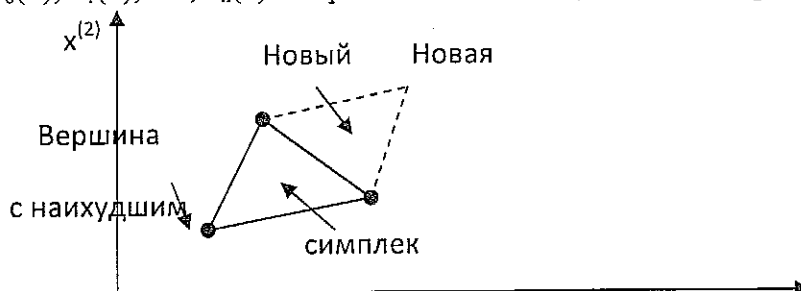


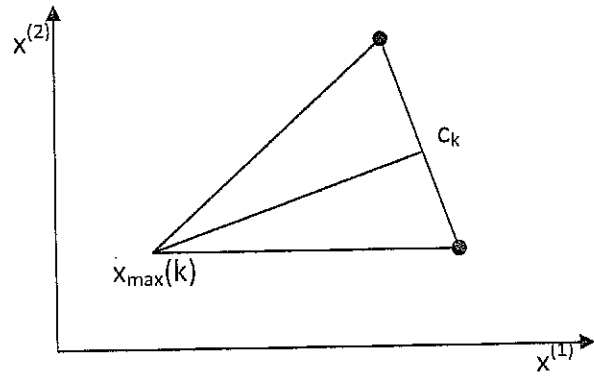
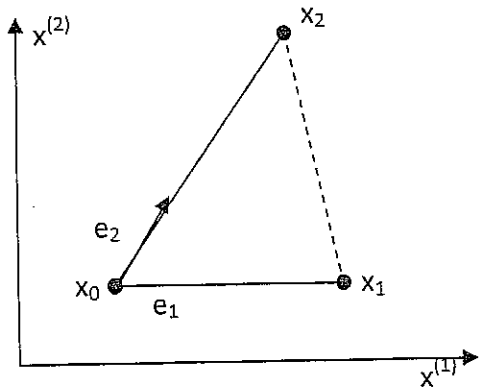
Схема алгоритма

Шаг 1.

Построение начального симплекса.

Для этого задаются начальная точка $x_0(0)$ и длина ребра симплекса l . Формируются остальные вершины симплекса:

$$x_i(0) = x_0(0) + l \cdot e_i \quad (i=1,2,\dots,n), \text{ где } e_i - \text{единичные векторы}.$$



Шаг 2.

Определение направления улучшения решения.

Для этого на k -й итерации вычисляются значения целевой функции в каждой точке симплекса. Пусть для всех i :

$$f(x_{\min}(k)) \leq f(x_i(k)) \leq f(x_{\max}(k)),$$

где \min , \max , i – номера соответствующих вершин симплекса. Определим центр тяжести всех точек, исключая точку $x_{\max}(k)$,

$$C_k = (\sum x_i(k)) / n.$$

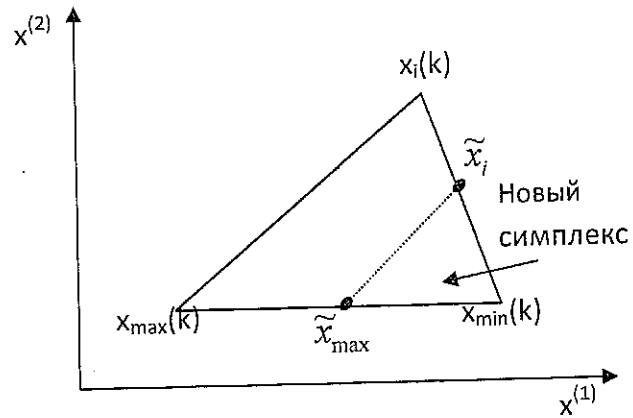
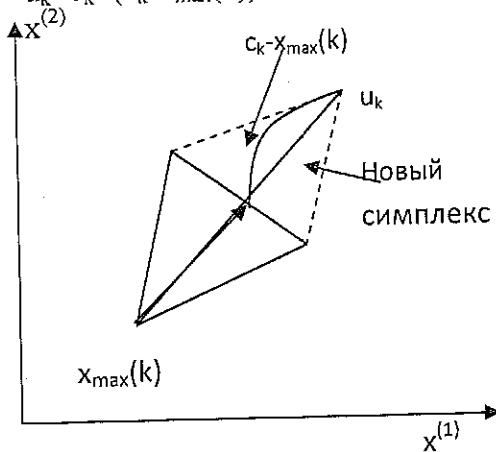
Тогда направление улучшения решения определяется вектором $C_k - x_{\max}(k)$.

Шаг 3.

Построение отраженной точки.

Замена вершины $x_{\max}(k)$ с максимальным значением целевой функции на новую точку с помощью операции отражения, результатом которой является новая точка:

$$u_k = C_k + (C_k - x_{\max}(k)) = 2C_k - x_{\max}(k)$$



Шаг 4.

Построение нового симплекса.

Вычисляем $f(u_k)$. При этом возможен один из двух случаев:

а) $f(u_k) < f(x_{\max}(k))$;

б) $f(u_k) \geq f(x_{\max}(k))$.

Случай а): вершина x_{\max} заменяется на u_k , чем определяется набор вершин $k+1$ -й итерации и k -я итерация заканчивается.

Случай б): в результате отражения получается новая точка u_k , значение функции в которой еще хуже, чем в точке x_{\max} , то есть отражать симплекс некуда. Поэтому в этом случае производится пропорциональное уменьшение симплекса (например, в 2 раза) в сторону вершины $x_{\min}(k)$:

$$x_i(k+1) = \hat{x}_i = (x_i(k) + x_{\min}(k)) / 2, \text{ где } i=0, 1, \dots, n.$$

На этом k -я итерация заканчивается.

Шаг 5.

Проверка сходимости.

Если

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i(k+1)) - f(x_0(k+1)))^2} \leq \varepsilon,$$

то поиск минимума заканчивается и полагается

$$\tilde{x} = x_0(k+1), \tilde{y} = f(x_0(k+1))$$

В противном случае $k=k+1$ и происходит переход к шагу 2.

1.5 Метод деформируемого симплекса (метод Нелдера – Мида).

В рассматриваемом методе симплекс перемещается с помощью трех основных операций над симплексом: отражение, растяжение и сжатие.

Схема алгоритма.

Шаг 1.

Построение начального симплекса.

Задаются начальная точка $x_0(0)$ и длина ребра l . Формируются остальные вершины симплекса: $x_i(0) = x_0(0) + l e_i$ ($i=1, 2, \dots, n$), где e_i – единичные векторы.

Шаг 2.

Определение направления улучшения решения.

Для этого на каждой итерации вычисляются значения целевой функции в каждой вершине симплекса. Пусть для всех i

$$f(x_{\min}(k)) \leq f(x_i(k)) \leq f(x_m(k)) \leq f(x_{\max}(k)),$$

где \min, m, \max, i – номера соответствующих вершин симплекса. Определим центр тяжести всех точек, исключая точку $x_{\max}(k)$,

$$C_k = \frac{1}{n} \sum_{i \neq \max} x_i(k)$$

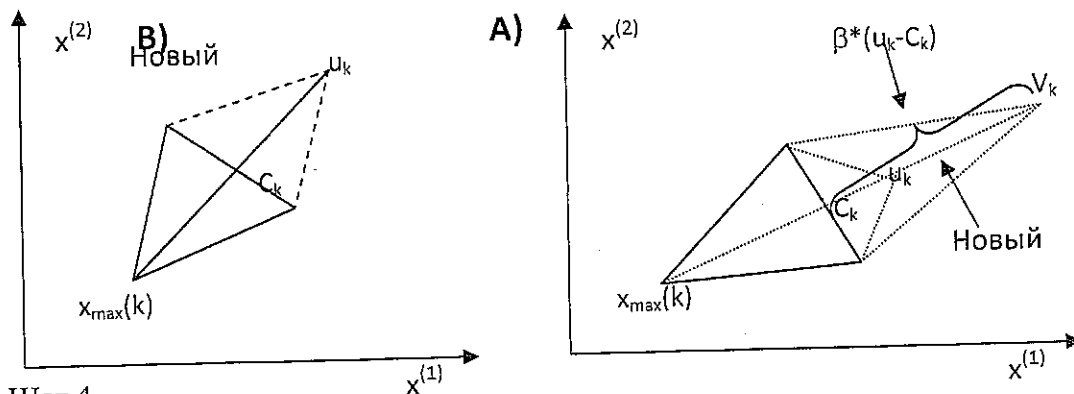
Тогда направление улучшения решения определяется векторов

$$C_k - x_{\max}(k).$$

Шаг 3.

Построение нового симплекса.

Замена вершины $x_{\max}(k)$ с максимальным значением целевой функции на новую точку с помощью операции отражения, результат которой является новая точка $u_k = C_k + \alpha^*(C_k - x_{\max}(k))$, где α – коэффициент отражения.



Шаг 4.

Построение нового симплекса.

Вычисляем $f(u_k)$, при этом возможно один из трех случаев:

а) $f(u_k) < f(x_{\min}(k))$;

- б) $f(u_k) > f(x_m(k))$;
 в) $f(x_{\min}(k)) \leq f(u_k) \leq f(x_m(k))$;

Случай а): отражённая точка является точкой с наилучшим значением целевой функции. Поэтому направление отражение является перспективным и можно попытаться растянуть симплекс в этом направлении. Для этого строится точка

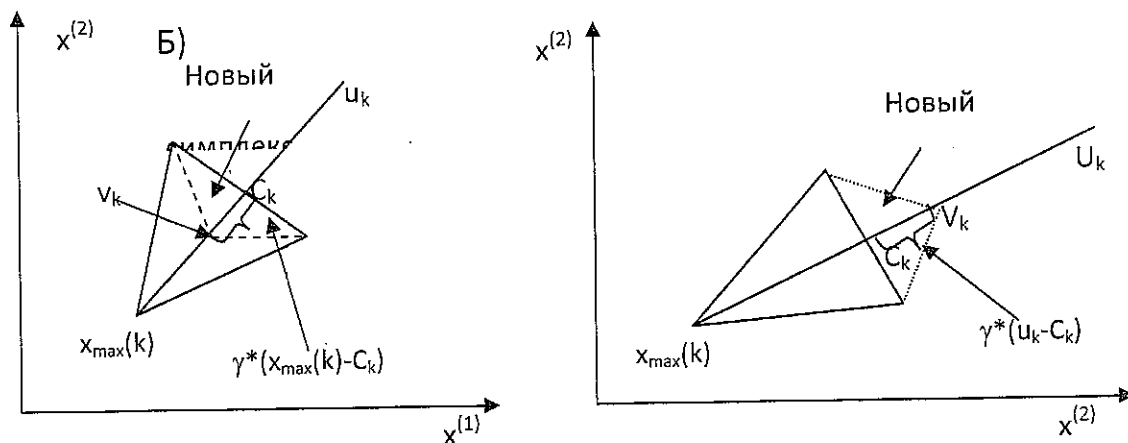
$$V_k = C_k + \beta^*(u_k - C_k), \text{ где } \beta > 1 \text{ — коэффициент расширения.}$$

Если $f(v_k) < f(u_k)$, то вершина $x_{\max}(k)$ заменяется на v_k , в противном случае на u_k и k -ая итерация заканчивается.

Случай б): в результате отражения получается новая точка u_k , которая, если заменить $x_{\max}(k)$, сама станет наихудшей. Поэтому в этом случае производится сжатие симплекса. Для этого строится точка v_k :

$$v_k = \begin{cases} C_k + \gamma^*(x_{\max}(k) - C_k), & \text{если } f(x_{\max}(k)) \leq f(u_k), \\ C_k + \gamma^*(u_k - C_k), & \text{если } f(x_{\max}(k)) > f(u_k), \end{cases}$$

где $0 < \gamma < 1$ — коэффициент сжатия.



Если $f(v_k) < \min\{f(x_{\max}(k)), f(u_k)\}$, то вершина $x_{\max}(k)$ заменяется на v_k .
 В противном случае вершинам $x_i(k+1)$ ($i=0,1,2,\dots,n$) присваивается значение:

$$\tilde{x} = \frac{x_i(k) + x_{\min}(k)}{2}$$

и на этом k -ая итерация заканчивается.

в) вершина $x_{\max}(k)$ заменяется на u_k , чем определяется набор вершин $k+1$ -й итерации и k -ая итерация заканчивается.

Шаг 5.

Проверка сходимости.

Если:

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i(k+1)) - f(x_0(k+1)))^2} \leq \varepsilon,$$

то поиск минимума заканчивается и полагается

$$\tilde{x} = x_0(k+1), \tilde{y} = f(x_0(k+1))$$

В противном случае $k=k+1$ и происходит переход к шагу 2.

Опыт использования описанного алгоритма показывает, что целесообразно брать следующие значения параметров:

$$\alpha=1, \beta=2, \gamma=0.5.$$

2. Задание на лабораторную работу

2.1 Изучить изложенные методы многомерной безусловной оптимизации.

2.2 В соответствии с вариантом задания, определенным преподавателем, найти точку минимума целевой функции $f(x)=f(x^{(1)}, x^{(2)})$ с заданной точностью ε указанными методами (составить программы, реализующие методы многомерной безусловной минимизации).

Целевая функция $f(x)=f(x^{(1)}, x^{(2)})$ зависит от двух аргументов. Функция $f(x)$ следующего вида:

$$f(x) = ax^{(1)} + bx^{(2)} + e^{cx+dx}.$$

№	Целевая функция				Начальное приближение	Точность решения
	a	b	c	d		
1	1	-1,4	0,01	0,11	(1;0)	0,0001
2	2	-1,3	0,04	0,12	(0;1)	0,00005
3	10	-0,5	0,94	0,2	(0;0)	0,0001
4	15	0	1,96	0,25	1,96	0,25
5	3	-1,2	0,02	1,3	(0;-1)	0,00005
6	11	-0,4	1	0,21	(-1;0)	0,0001
7	10	-1	1	2	(1;0)	0,0003
8	15	-0,5	2,25	2,5	(0;0)	0,0002
9	20	0,4	0,3	0,3	(0;-1)	0,0001
10	25	0,9	0,35	0,35	(1;0)	0,0004

Начальное приближение x_0 и точность ε приводятся в условии задачи. Сравнить результаты, полученные разными методами для одной и той же целевой функции (в частности, сравнить число вычислений целевой функции и её производных, понадобившихся для получения заданной точности). Для каждого применяемого метода построить траекторию промежуточных точек, получаемых на очередных шагах метода и сходящихся к точке минимума.

2.3 Оформить отчет о выполнении задания с приведением условия задачи, алгоритмов и программ, указанных в задании методов минимизации, графиков траекторий промежуточных приближений, таблицы результатов сравнения рассмотренных методов, заключения по результатам сравнения методов.

3. Вопросы к лабораторной работе

- 1) Какова сущность прямых методов минимизации функции нескольких переменных?
- 2) Дайте характеристику метода минимизации по правильному симплексу.
- 3) Представьте алгоритм метода минимизации по правильному симплексу.
- 4) Дайте характеристику метода деформируемого симплекса (метод Нелдера – Мида)

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие / А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

**МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ**
МЕТОДЫ ОПТИМИЗАЦИИ ДИСКРЕТНЫХ ЗАДАЧ

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

МЕТОДЫ ОПТИМИЗАЦИИ ДИСКРЕТНЫХ ЗАДАЧ: методические указания для
выполнения лабораторных работ по дисциплине «МЕТОДЫ ОПТИМИЗАЦИИ В
БИЗНЕС-АНАЛИТИКЕ» для студентов направления 09.04.03 «Прикладная
информатика» направленность «Искусственный интеллект и бизнес-аналитика» /
Составитель: Т. В. Хоменко. – Ульяновск: УлГТУ, 2021. – 37с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и бизнес-
аналитика». Цель данного практикума – ориентировать студентов на содержание
и порядок выполнения лабораторных работ во время прохождения студентами
курсов «Искусственный интеллект и бизнес-аналитика». Даются задания для
лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание

ВВЕДЕНИЕ	4
Лабораторная работа №3	5
Тема: Переборные методы решения задач дискретной оптимизации.....	5
1. Краткие теоретические сведения.....	5
2. Задание на лабораторную работу	14
3. Вопросы к лабораторной работе.....	15
Лабораторная работа №4	16
Тема: Метод динамического программирования в задачах дискретной оптимизации.....	16
1. Краткие теоретические сведения.....	16
2. Задание на лабораторную работу	34
3. Вопросы к лабораторной работе.....	35
Список литературы	37

ВВЕДЕНИЕ

В учебно-методическом пособии описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.)

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Лабораторная работа №3

Тема: Переборные методы решения задач дискретной оптимизации

Цель работы: знакомство с оптимизационными задачами, изучение различных методов дискретной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Основные типы задач дискретной оптимизации

Задача дискретной оптимизации – это задача поиска максимума или минимума функции f , определенной на конечном или счетном множестве D :

$$f(x) \rightarrow \text{extr}, \quad x \in D.$$

Функция f называется целевой функцией, элементы множества D – допустимыми решениями. Если множество D задается системой ограничений:

$$g_i(x) \leq 0, \quad i = 1, \dots, m_1, \quad g_i(x) = 0, \quad i = m_1 + 1, \dots, m, \\ x = (x_1, \dots, x_n) \in R^n, \quad x_j \in \Omega_j \subset R, \quad j = 1, \dots, n_1, \quad n_1 \leq n,$$

где каждое Ω_j – либо конечное множество, содержащее не менее двух элементов, либо счетное множество, то задача (1.1) называется задачей частично дискретного (или дискретного, если $n_1 = n$) математического программирования. Если $\Omega_j \subseteq Z, j = 1, \dots, n_1$, задача (1.1) называется задачей частично целочисленного (целочисленного при $n_1 = n$) программирования. В пособии будут рассматриваться, в основном, задачи целочисленного линейного программирования, которые формулируются следующим образом:

$$\sum_{j=1}^n c_j x_j \rightarrow \min \\ \sum_{j=1}^n a_{ij} x_j = b_i, \quad i = 1, \dots, m, \\ x_j \geq 0, \quad j = 1, \dots, n, \\ x_j \in Z, \quad j = 1, \dots, n_1, \quad n_1 \leq n.$$

Обычно предполагается, что параметры задачи – коэффициенты целевой функции $c_j, j = 1, \dots, n$, элементы матрицы системы ограничений $A = (a_{ij})$ и вектора правых частей $b = (b_1, \dots, b_m)$ – целые числа.

Задача о рюкзаке

Задачей о рюкзаке или ранце называется задача целочисленного линейного программирования с одним ограничением:

$$\sum_{j=1}^n c_j x_j \rightarrow \max \\ \sum_{j=1}^n a_j x_j \leq b, \\ x_j \geq 0 \text{ и целые, } j = 1, \dots, n.$$

Числа $c_j, a_j (j=1, \dots, n), b$ можно считать положительными и целыми, $a_j \leq b$.

Имеется n видов неделимых предметов со стоимостями c_1, \dots, c_n и весами a_1, \dots, a_n . Требуется так упаковать ими рюкзак, чтобы его вес не превышал b , а суммарная

стоимость упакованных предметов была максимальной. Вводя переменные $x_j, j=1, \dots, n$, для количества упакованных предметов каждого типа, получаем представленную задачу. Иногда рассматривается задача о рюкзаке с ограничением типа равенства:

$$\sum_{j=1}^n a_j x_j = b,$$

что соответствует требованию полной загрузки рюкзака грузоподъемностью b .

Обобщением задачи о рюкзаке является задачей о многомерном рюкзаке:

$$\begin{aligned} \sum_{j=1}^n c_j x_j &\rightarrow \max \\ \sum_{j=1}^n a_{ij} x_j &\leq b_i, \quad i=1, \dots, m, \\ x_j &\in \{0, 1\}, \quad j=1, \dots, n, \end{aligned}$$

имеющая следующую экономическую интерпретацию.

Пусть есть n проектов, ожидаемая прибыль от реализации которых составляет c_1, \dots, c_n . Задан вектор ресурсов $b=(b_1, \dots, b_m)$, $b_i > 0, i=1, \dots, m$; количество единиц ресурса типа i , необходимое для реализации проекта с номером j , равно $a_{ij} > 0$. Для любого ресурса i выполнено условие:

$$\sum_{j=1}^n a_{ij} > b_i$$

то есть реализация всех проектов невозможна. Требуется выбрать набор проектов с максимальной суммарной прибылью.

Задача об упаковке

Задачу об упаковке в контейнеры можно сформулировать следующим образом: задано число $Q > 0$ и набор n положительных действительных чисел $L=\{w_1, \dots, w_n\}$, требуется разбить множество L на наименьшее возможное число подмножеств так, чтобы сумма чисел в каждом подмножестве не превосходила Q . Без ограничения общности можно считать, что:

$$Q=1, w_i \in (0, 1] \text{ при } i=1, \dots, n \text{ и } \sum_{i=1}^n w_i > 1.$$

Задача интерпретируется как проблема упаковки предметов с размерами w_1, \dots, w_n в единичный контейнер.

Сформулируем задачу об упаковке как задачу целочисленного линейного программирования. Так как допустимым является решение, при котором в каждый контейнер помещается один предмет, можно считать, что имеется n контейнеров. Введем переменные $y_j, j=1, \dots, n$, принимающие значение 1, если j -й контейнер занят, и 0 в противном случае, и $x_{ij}, i, j=1, \dots, n$, равные 1, если i -й предмет помещен в j -й контейнер и 0, если нет. Задача тогда примет вид:

$$\begin{aligned} \sum_{j=1}^n y_j &\rightarrow \min \\ \sum_{i=1}^n w_i x_{ij} &\leq y_j, \quad j=1, \dots, n, \\ \sum_{i=1}^n x_{ij} &= 1, \quad i=1, \dots, n, \\ x_{ij} &\in \{0, 1\}, \quad y_j \in \{0, 1\}, \quad i, j=1, \dots, n. \end{aligned}$$

Второе условие – ограничения на вместимость контейнера, третье условие

гарантирует, что каждый предмет помещен ровно в один контейнер.

Транспортная задача с фиксированными доплатами

Рассмотрим математическую модель следующей задачи организации перевозок однородного груза. Имеется m пунктов отправления (поставщиков) и n пунктов назначения (потребителей), предложение i -го поставщика составляет $a_i, i=1, \dots, m$, спрос j -го потребителя – $b_j, j=1, \dots, n$, при этом:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$$

Затраты на транспортировку x единиц груза от поставщика i до потребителя j составляют:

$$c_{ij}(x) = \begin{cases} 0, & x = 0, \\ c_{ij}x + d_{ij} & x > 0. \end{cases}$$

Числа $c_{ij} \geq 0$ интерпретируются как стоимость перевозки единицы груза и определяются, например, расстоянием между пунктами, постоянные затраты $d_{ij} \geq 0$ могут представлять собой оплату аренды транспортных средств или другие издержки, не зависящие от объема перевозок. Требуется организовать перевозку груза от поставщиков потребителям так, чтобы совокупные затраты на транспортировку были минимальны. Обозначив через x_{ij} количество единиц груза, перевозимого из i -го пункта производства в j -й пункт потребления, получим следующую задачу математического программирования с неоднородной разрывной линейной целевой функцией:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^n c_{ij}(x_{ij}) &\rightarrow \min \\ \sum_{j=1}^n x_{ij} &= a_i, \quad i = 1, \dots, m, \\ \sum_{i=1}^m x_{ij} &= b_j, \quad j = 1, \dots, n, \\ x_{ij} &\geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, n. \end{aligned}$$

Данная задача называется транспортной задачей с фиксированными доплатами или неоднородной транспортной задачей. Очевидно, если все $d_{ij}=0$, то рассматривается классическая транспортная задача. Данную задачу можно свести к задаче частично целочисленного линейного программирования. Для этого обозначим:

$$M_{ij} = \min\{a_i, b_j\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

и введем переменные:

$$y_{ij} \in \{0, 1\}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

принимающие значение 1, если $x_{ij} > 0$ и 0 в противном случае. Тогда задача при описанных ограничениях и дополнительных ограничениях:

$$x_{ij} \leq M_{ij} y_{ij}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

будет эквивалентна исходной.

Задачи коммивояжера

Коммивояжер, находящийся в родном городе, должен посетить несколько других населенных пунктов, побывав в каждом по одному разу, и вернуться обратно так, чтобы общая длина его пути была как можно меньше.

Пусть все пункты (включая начальный) перенумерованы и их общее количество

равно n . Расстояние от пункта i до пункта j обозначается c_{ij} . Полагая $c_{ii} = \infty$ (коммивояжеру запрещается оставаться на месте), определяем $n \times n$ матрицу $C = (c_{ij})$. Естественно считать, что $c_{ij} \geq 0$. Любой маршрут коммивояжера полностью определяется порядком посещенных пунктов и имеет, поэтому, вид $z = (i_1, i_2, \dots, i_n, i_1)$. Длина маршрута z – сумма соответствующих элементов матрицы C , то есть:

$$l(z) = \sum_{k=1}^n c_{i_k, i_{k+1}},$$

где $i_{n+1} = i_1$. Обозначив множество допустимых маршрутов через D , получаем задачу комбинаторной оптимизации:

$$l(z) \rightarrow \min, z \in D.$$

Очевидным практическим приложением задач коммивояжера является планирование маршрутов перевозки грузов.

Соответственно, c_{ij} может представлять собой не только расстояние, но и время, издержки, другой измеритель, служащий для определения «выгодности» маршрута. В общем случае, c_{ij} – стоимость следования j непосредственно после i .

Пусть $G = (V, A)$ – полный граф со множеством вершин $V = \{1, \dots, n\}$ и множеством дуг A , длина дуги (i, j) в котором равна c_{ij} . Контур (ориентированный цикл), включающий каждую вершину графа ровно один раз, называется гамильтоновым контуром. Таким образом, задача коммивояжера – это задача поиска гамильтонова контура, имеющего минимальную длину. В силу замкнутости маршрута можно считать, что начальный пункт i_1 задан, поэтому количество допустимых маршрутов равно $(n-1)!$. Если $c_{ij} = c_{ji}$ для всех $i, j = 1, \dots, n$ (матрица C симметричная), граф G является неориентированным, и говорят о гамильтоновых циклах. Поскольку перестановки (i_2, \dots, i_n) и (i_n, \dots, i_2) определяют один маршрут, количество различных замкнутых маршрутов составит $(n-1)!/2$.

1.2. Переборные методы решения задач дискретной оптимизации

Для всех представленных выше задач множество решений конечно и, перебрав все из них, можно найти лучшее. Для реализации полного перебора нужна процедура генерации всех допустимых решений:

$$\begin{cases} Q(\bar{x}^*) = \max_{\bar{x} \in D} Q(\bar{x}), \\ 0 < |D| = N < \infty. \end{cases}$$

Здесь множество D – конечно, и количество возможных решений равно N . В силу этого формально все допустимые решения можно пронумеровать:

$$\bar{x}^1, \bar{x}^2, \dots, \bar{x}^N.$$

Поиск лучшего варианта из множества допустимых решений сводится к полному перебору. Именно этим обстоятельством объясняется название переборных задач.

Как правило, полный перебор невозможно осуществить на практике из-за огромного числа рассматриваемых вариантов. Методы полного перебора применяются в случае малых размеров входных данных или в виде частичного перебора в других алгоритмах. Для оценивания времени работы алгоритмов, основанных на полном переборе можно подсчитать число возможных решений для задач.

ПРИМЕР 1. Задача о рюкзаке

Каждый предмет может быть либо помещен в рюкзак, либо нет. Соответственно, каждая переменная x_j может принимать одно из двух возможных значений – 1 или 0.

Любому решению соответствует вектор из n нулей и единиц. Всего таких векторов 2^n . Пусть $V=10$, $n=3$, $a_1=6$, $a_2=5$, $a_3=4$, $c_1=3$, $c_2=4$, $c_3=5$. Рассмотрим всевозможные комбинации значений переменных x_1, x_2, x_3 и соответствующие им наборы:

x_1	x_2	x_3	набор	вес	стоимость
0	0	0	\emptyset	0	0
1	0	0	{1}	6	3
0	1	0	{2}	5	4
0	0	1	{3}	4	5
1	1	0	{1, 2}	11	7
1	0	1	{1, 3}	10	8
0	1	1	{2, 3}	9	9
1	1	1	{1, 2, 3}	15	12

Сначала отбираем наборы, которые помещаются в рюкзак (1, 2, 3, 4, 6, 7), среди которых выбираем набор с наибольшей стоимостью. Очевидно, что можно взять 6-ой набор (первый и третий предметы), можно взять 7-ой набор (второй и третий предметы). Однако у 7-го набора стоимость выше.

ПРИМЕР 2. Задачи коммивояжера

Нужно найти перестановку городов A_1, \dots, A_n , для которой пройденное расстояние минимально. Первый элемент искомой перестановки всегда равен 1. Поэтому при полном переборе переставлять нужно оставшиеся $n-1$ города. Всего таких перестановок $(n-1)!$.

Пусть $n=5$, матрица расстояний между городами имеет вид:

$$C = \begin{pmatrix} 0 & 5 & 3 & 1 & 2 \\ 5 & 0 & 1 & 3 & 6 \\ 3 & 1 & 0 & 4 & 2 \\ 1 & 3 & 4 & 0 & 7 \\ 2 & 6 & 2 & 7 & 0 \end{pmatrix}.$$

Тогда получаем $4!=24$ различных перестановок. Если при этом матрица задачи симметрическая, то получаем 12 перестановок. Таблица всевозможных маршрутов имеет следующий вид:

маршрут	стоимость
1 → 2 → 3 → 4 → 5 → 1	19
1 → 2 → 3 → 5 → 4 → 1	16
1 → 2 → 4 → 3 → 5 → 1	16
1 → 2 → 4 → 5 → 3 → 1	20
1 → 2 → 5 → 3 → 4 → 1	18
1 → 2 → 5 → 4 → 3 → 1	25
1 → 3 → 2 → 4 → 5 → 1	16
1 → 3 → 2 → 5 → 4 → 1	18
1 → 3 → 4 → 2 → 5 → 1	18
1 → 3 → 5 → 2 → 4 → 1	15
1 → 4 → 2 → 3 → 5 → 1	9
1 → 4 → 3 → 2 → 5 → 1	14

Очевидно, оптимальный маршрут выглядит так:

$$1 \rightarrow 4 \rightarrow 2 \rightarrow 3 \rightarrow 5 \rightarrow 1.$$

При этом, пройденное коммивояжером расстояние равно 9.

1.3. Методы генетического поиска в решении задач дискретной оптимизации

Как правило, задачи дискретной оптимизации характеризуются большим числом переменных, и, как следствие, большим пространством поиска, что не дает возможности перебрать все многообразие решений за «разумное» время. С другой стороны, на практике зачастую и не требуется нахождения глобального решения, достаточно найти «приемлемое» решение по заданному критерию (или набору критериев). Этими обстоятельствами объясняется повышенный интерес к комбинаторным методам оптимизации. В основе многих комбинаторных алгоритмов лежат эвристические схемы нахождения решений, которые эксплуатируют особенности объекта исследования.

Однако и подобные методы не всегда удается использовать на практике, так как объект исследования может иметь очень непростую природу, и выявление закономерностей и особенностей его внутренней организации может оказаться нетривиальной задачей. В связи с этим возникла проблема практической разрешимости задач дискретной оптимизации: найти эффективный или хотя бы достаточно простой в практически важных случаях алгоритм ее решения. Другой прием, который позволил сократить вычислительные затраты и соответственно увеличить размерность задач, заключается в ведении элементов рандомизации. Эволюционные методы представляют целый спектр алгоритмов, которые активно эксплуатируют данный прием. Как уже было отмечено, методы эволюционных вычислений не гарантируют обнаружения оптимального решения за полиномиальное время. Эти методы позволяют исследовать и находить приемлемые решения таких задач, решение которых при помощи традиционных методов оказывается затруднительным, а в некоторых случаях и просто невозможным.

Введем ряд определений.

Репродукционно-популяционный алгоритм поиска – это алгоритм поисковой оптимизации, который начинается с начальной популяции P_0 (совокупности кодировок $(x_1^0, x_2^0, \dots, x_n^0)$) и итеративно выполняет следующий цикл операций:

1) *Оценивание* – вычисление значений функции приспособленности для любой кодировки $x \in S$.

2) *Селекция* – выбор из популяции P^t , $t=0,1,\dots$ репродукционного множества R^t (подмножества кодировок $R^t \subseteq P^t$).

3) *Воспроизводство (размножение)* – генерация из репродукционного множества R^t новых кодировок с помощью комбинаций следующих операций:

- *копирование* – создание тождественных копий некоторых или всех кодировок из R^t ;
- *скрещивание* – конструирование новых кодировок путем сцепления подстрок тех кодировок, которые выбираются путем копирования из R^t ;
- *мутация* – конструирование новых кодировок путем подстановки символов из алфавита V в выбранные позиции одной из кодировок $x \in R^t$.

4) *Замена* – формирование на очередном шаге (поколении) новой популяции P^{t+1} путем замены некоторых или всех кодировок $x \in P^t$.

Канонические генетические алгоритмы – это подкласс репродукционно-популяционных алгоритмов поиска, в которых основные операторы имитируют механизмы естественного отбора и популяционной генетики, удовлетворяющий следующим требованиям:

- все кодировки $x \in S$ имеют одну и ту же длину L ;

- каждая популяция $P^t, t=0, 1, \dots$ имеет постоянную численность n ;
- мощность репродукционного множества $|R^t|=2$; элементы $\chi' \chi'' \in R^t$, называемые *родителями*, выбираются из популяции P^t случайным образом с вероятностями, пропорциональными значениям функции приспособленности;
- скрещивание осуществляется с помощью *одноточечного кроссовера*, когда символы $(\chi_1, \chi_2, \dots, \chi_i)$ в новой кодировке χ , называемой «потомком», являются символами кодировки $\chi' \in R^t$, а символы $(\chi_{i+1}, \chi_{i+2}, \dots, \chi_L)$ переходят от кодировки $\chi'' \in R^t$; индекс i , называемый *точкой кроссовера*, выбирается случайным образом с равной вероятностью из интервала $[1, L-1]$; мутации являются стохастическими операциями и обычно не зависят от значений функции приспособленности;
- замена сводится к исключению из популяции P^t особи с наименьшим значением функции приспособленности и включению в нее одной из новых кодировок с наибольшим значением функции приспособленности.

1.4. Обобщенная структура генетического алгоритма

Обобщенная структура генетического алгоритма может быть представлена следующим образом.

Шаг 1. Инициализация начальной популяции P^0 численностью v .

1.1. Установить номер текущего поколения $t:=0$.

1.2. Сгенерировать случайным образом хромосомный набор из v строковых кодировок фиксированной длины L , в котором Хэммингово расстояние между любой парой кодировок не равно нулю.

1.3. Оценить каждую строку из хромосомного набора с помощью функции приспособленности.

Шаг 2. Воспроизводство потомков с наследственными генетическими свойствами родителей.

2.1. Выбрать случайным образом из текущей популяции P^t согласно схеме скрещивания кодировки двух родителей, образующих «брачную пару».

2.2. Сгенерировать при помощи оператора кроссовера для выбранной «брачной пары» с вероятностью p_c одну или несколько кодировок потомков, которые наследуют генетические свойства родителей.

2.3. Оценить каждую кодировку потомков с помощью функции приспособленности.

2.4. Повторять все операции с п.2.1 до тех пор, пока не будет рассмотрено заданное число «брачных пар» N_c .

Шаг 3. Создание мутантов с генетическими свойствами, отличающимися от свойств родителей.

3.1. Выбрать случайным образом из числа потомков и/или родителей кодировку, наследующую генетические свойства одного или обоих родителей.

3.2. Сгенерировать при помощи оператора мутации для выбранной кодировки с вероятностью p_m кодировку-мутанта, обеспечивающую изменчивость генетических свойств родителей.

3.3. Оценить кодировку мутанта с помощью функции приспособленности.

3.4. Повторять все операции с п.3.1 до тех пор, пока не будет получено заданное число мутантов N_m .

Шаг 4. Замена текущей популяции P^t новой популяцией P^{t+1} .

4.1. Выбрать стратегию формирования популяции P_{t+1} .

4.2. Сформировать из родителей и/или «детей» (потомков и мутантов) репродукционное множество кодировок, различающихся между собой по Хэммингову расстоянию.

4.3. Скопировать при помощи оператора селекции из репродукционного множества кодировки, реализующие стратегию формирования популяции P_{t+1} .

Шаг 5. Условие выхода из итерационного цикла.

Сменить номер текущего поколения $t:=t+1$ и повторить все операции с п.2, если условие окончания генетического поиска не выполнено (например, эволюция популяции P^t считается законченной, если она исчерпала свой жизненный цикл T , т.е. если $t > T$).

Параметрами генетического алгоритма являются:

- n – численность популяции P^t , $t = 0, 1, \dots$;
- N_c – число «брачных пар»;
- N_m – число «мутантов»;
- p_c – вероятность кроссовера;
- p_m – вероятность мутации;
- T – число поколений, в течение которых осуществляется генетический поиск.

1.5. Сравнение качества решений

При сравнении методов формируется выборка $\Lambda = \{G_1, G_2, \dots, G_K\}$, из $K = |\Lambda|$ тестовых примеров, для каждого из которых производится построение решений с использованием выбранных методов $\Phi = \{m_1, m_2, \dots, m_T\}$, $T = |\Phi|$ и оценка их качества.

Рассмотрим в качестве примера сравнение качества решений эвристических методов в задаче поиска кратчайшего пути в графе. Пусть качество решения, полученного методом m_i , $i = \overline{1, T}$, на тестовом примере G_j , $j = \overline{1, K}$ определяется как $Q_{m_i}(G_j)$. Тогда простейшим способом сравнения качества решений является сопоставление их средневыборочных значений:

$$\gamma_{m_i}(\Lambda) = \frac{\sum_{j=1}^K Q_{m_i}(G_j)}{K} \rightarrow \min \quad (5.1)$$

Однако в случае, когда для некоторых тестовых примеров корректные решения не найдены (для них $Q_{m_i}(G_j) = \infty$), данная формула неприменима, и вместо нее при сравнении методов необходимо использовать критерий:

$$\Delta\gamma_{m_i}(\Lambda) = \frac{\sum_{j=1}^K (Q_{m_i}(G_j) - Q^*(G_j)) \phi_{m_i}(G_j)}{K} \rightarrow \min \quad (5.2)$$

где $Q^*(G_j)$ – качество наилучшего решения.

Критерий (5.2) представляет собой среднюю величину ухудшения решения по сравнению с известным оптимумом. При этом для рассматриваемых эвристических методов возможно оценить вероятность получения решения:

$$\rho_{m_i}(\Lambda) = \frac{\sum_{j=1}^K \phi_{m_i}(G_j)}{K} \rightarrow \max \quad (5.3)$$

причем метод, для которого $\rho_{m_i} \rightarrow 1$, является предпочтительным. Таким образом, для

объективного сравнения методов необходимо учитывать пару критериев, характеризующих среднее качество решений (или среднее отклонение от оптимума) и вероятность отыскания решения.

Еще одним возможным критерием сравнения эвристических методов является вероятность получения суб- или квазиоптимальных решений:

$$\rho_{m_i}^*(\Lambda) = \frac{\sum_{j=1}^K \theta_{m_i}(G_j)}{K} \rightarrow \max \quad (5.4)$$

где

$$\theta_{m_i}(G_j) = \begin{cases} 0, & Q_{m_i}(G_j) > Q^*(G_j), \\ 1, & Q_{m_i}(G_j) = Q^*(G_j) \end{cases} \quad (5.5)$$

– функция, принимающая единичное значение в случае совпадения качества текущего $Q_{m_i}(G_j)$ и наилучшего $Q^*(G_j)$ решений.

Отклонения качества решений $\Delta Q(G_j) = Q_{m_i}(G_j) - Q^*(G_j)$ имеют случайный характер, и для них может быть произведена оценка вероятности того, что найденное решение отличается от оптимума не более чем на заданную величину ΔQ :

$$\rho_{m_i}^\Delta(\Delta Q, \Lambda) = \frac{\sum_{j=1}^K \mu_\Delta(Q_{m_i}(G_j), Q^*(G_j), \Delta Q)}{K}, \quad (5.6)$$

где

$$\mu_\Delta(Q, Q^*, \Delta Q) = \begin{cases} 1, & Q - Q^* \leq \Delta Q, \\ 0, & Q - Q^* > \Delta Q \end{cases} \quad (5.7)$$

– функция, принимающая единичное значение в случае, если отклонение качества решения $\Delta Q(G_j) = Q(G_j) - Q^*(G_j)$ не превышает заданное пороговое значение ΔQ , и нулевое в противном случае. Обычно при сравнении эвристических методов используют

графики зависимости $\rho_{m_i}^\Delta$ от ΔQ . Несложно показать, что при $\Delta Q = 0$ $\rho_{m_i}^\Delta = \rho_{m_i}^*$.

При работе с малыми выборками возможна ситуация, в которой разница в значениях сопоставляемых критериев вызвана статистической погрешностью, а не различием в поведении методов. Для того, чтобы статистически корректно доказать преимущество одного метода по сравнению с другим, необходимо убедиться в том, что различие сопоставляемых средних оценок является статистически значимым. Для этого необходимо произвести расчет дисперсии средневыворочных оценок:

$$D_{m_i}(\Lambda) = \frac{1}{K} \sum_{j=1}^K (Q_{m_i}(G_j) - \gamma_{m_i})^2, \quad (5.8)$$

оценить величину среднеквадратичного отклонения:

$$\sigma_{m_i}(\Lambda) = \sqrt{D_{m_i}}$$

и рассчитать ширину доверительного интервала ΔQ для заданной доверительной вероятности. Например, для доверительной вероятности 0,95:

$$\Delta Q_{m_i}(\Lambda) = \frac{1,96 \sigma_{m_i}}{\sqrt{K}}$$

где коэффициент 1,96 представляет собой квантиль распределения Стьюдента для заданной доверительной вероятности. Смысл указанных действий заключается в том, что

при записи средневыборочного значения в виде $\gamma_{m_i}(\Lambda) \pm \Delta Q_{m_i}(\Lambda)$ можно утверждать, что истинное среднее значение математического ожидания (γ_{m_i} при $K \rightarrow \infty$) находится в пределах доверительного интервала $[\gamma_{m_i}(\Lambda) - \Delta Q_{m_i}(\Lambda); \gamma_{m_i}(\Lambda) + \Delta Q_{m_i}(\Lambda)]_c$ заданной доверительной вероятностью.

2. Задание на лабораторную работу

2.1. Написать программу, реализующую решение, указанной в индивидуальном варианте, задачи полным перебором и генетическим алгоритмом. Сравнить качество решений путем организации вычислительного эксперимента над выборкой из K псевдослучайных тестовых примеров, для каждого из методов вычислить:

- средневыборочные оценки качества решений;
- соответствующие им дисперсии;
- среднеквадратичные отклонения;
- границы доверительных интервалов (для доверительной вероятности 0,95);
- вероятности отыскания решений;
- вероятности отыскания оптимального (или условно оптимального, в зависимости от задачи) решения.

Индивидуальные варианты

1	В заданном ориентированной графе найти кратчайший гамильтонов путь, соединяющий указанную пару вершин (гамильтонов путь – это путь, однократно проходящий через все вершины графа)
2	В заданном неориентированном графе найти кратчайший гамильтонов цикл (гамильтонов цикл – это цикл, однократно проходящий через все вершины графа)
3	Определить хроматическое число для заданного неориентированного графа (хроматическим числом неориентированного графа называется минимальное число цветов, в которое можно раскрасить вершины графа так, чтобы соединенные ребром вершины были раскрашены в разные цвета)
4	Найти такой способ перенумерации вершин (подстановку изоморфизма), который дает минимальное число расхождений между матрицами смежности рассматриваемых графов (графы называются изоморфными, если из одного можно получить другой путем перенумерации его вершин)
5	Найти такое разбиение графа на N подграфов, чтобы сумма ребер, связывающих полученные подграфы, была минимальна (для неориентированного графа заданы веса ребер)
6	Для заданного графа оценить мощность максимального независимого множества (максимальным независимым множеством называется такое максимальное по включению подмножество вершин графа, в котором ни одна пара вершин не соединена ребром)
7	Для заданного неориентированного графа найти мощность максимального по включению полносвязного подграфа
8	Для заданного графа определить его диаметр (диаметр графа – это максимальное расстояние между вершинами графа)
9	Для заданного ориентированного графа оценить мощность компоненты сильной связности (компонентой сильной связности ориентированного графа называется

	максимальное по включению множество взаимно достижимых вершин)
10	Найти минимальное остовное дерево заданного неориентированного графа (минимальным остовным деревом графа G называется дерево с минимальной суммарной длиной ребер, образованное из ребер графа G)

2.2. Построить графики зависимости вероятности получения решения с заданным отклонением от оптимума (или условного оптимума, в зависимости от задачи) от величины допустимого отклонения. Если в состав метода входят настроечные параметры, произвести метаоптимизацию, результаты которой представить в виде графиков зависимостей средневыборочных величин от значений настроечных параметров. Для указанных выше средневыборочных оценок произвести построение графиков их зависимостей от размерности задачи N и плотности графа d .

Произвести оценку времени отыскания решения, построить графики зависимости средневыборочного качества решений от числа итераций.

3. Вопросы к лабораторной работе

- 1) Какое решение называется оптимальным?
- 2) Какие методы гарантируют получение оптимальных решений?
- 3) Дайте характеристику метода полного перебора
- 4) Какие алгоритмы называются генетическими?
- 5) Дайте характеристику обобщенного генетического алгоритма
- 6) Раскрыть процедуру сравнения качества решений

Лабораторная работа №4

Тема: Метод динамического программирования в задачах дискретной оптимизации

Цель работы: знакомство с оптимизационными задачами, изучение различных методов дискретной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Принцип динамического программирования

Метод динамического программирования – один из основных математических методов оптимизации. При его реализации одновременный выбор значений большого числа переменных решаемой экстремальной задачи заменяется поочередным определением каждой из них в зависимости от возможных обстоятельств. Процедуру выбора значений переменных будем трактовать как многоэтапный процесс управления некоторой системой. Далее объект Ω именуем дискретной управляемой системой, если множество его возможных состояний конечно и управления в нем осуществляются в дискретном времени, пошагово; каждое управление заключается в применении одного из конечного числа возможных воздействий; результатом любого воздействия является изменение состояния системы. С каждым изменением состояния системы связываем платеж (доход или расход); полагаем, что величина платежа на любом шаге зависит от текущего состояния системы и применяемого управляющего воздействия. Начальное состояние системы считается заданным. Определено также множество финальных состояний (достигнув финального состояния, система прекращает функционировать). Каждая последовательность управлений, переводящая систему из начального в одно из финальных состояний, фактически определяет некоторую полную траекторию движения системы. Требуется найти полную траекторию, оптимальную по значению суммарного платежа, т.е. обеспечивающую максимальный суммарный доход или минимальный суммарный расход. Возможно и другое требование – найти полную траекторию с минимальным значением максимального из пошаговых расходов (максимальным значением минимального из пошаговых доходов).

Формально дискретную управляемую систему определяем как совокупность:

$$\Omega = \{D; x_0; F; V(x), f(x, v), s(x, v)\},$$

где D – множество состояний системы (множество D конечно); x_0 – начальное состояние; F – множество финальных состояний, $x_0 \in D$, $x_0 \notin F$, $F \subset D$; $V(x)$ – конечное множество возможных в состоянии x управлений, $x \in D \setminus F$; $f(x, v)$ – функция переходов (из состояния x под воздействием управления v система переходит в состояние $f(x, v)$), $x \in D \setminus F$, $v \in V(x)$, $f(x, v) \in D$; $s(x, v)$ – функция платежа, здесь $x \in D \setminus F$, $v \in V(x)$; значения функции платежа считаются неотрицательными.

Состояние x системы Ω называем промежуточным, если оно не является ни начальным, ни финальным. Конечную последовательность $T = \{x_0, x_1, x_2, \dots, x_n\}$ состояний системы Ω именуем траекторией системы, если выполняются условия:

$$x^t = f(x^{t-1}, v^t), \text{ где } v^t \in V(x^{t-1}), t = 1, 2, \dots, n.$$

Состояние x^0 – начальное состояние траектории T , а состояние x^n – ее конечное состояние. Состояния x^2, x^3, \dots, x^{n-1} из T являются промежуточными состояниями данной

траектории.

Траекторию называем полной, если начальным ее состоянием является начальное состояние системы Ω , а конечным – одно из финальных состояний этой системы. Таким образом, для полной траектории $T = \{x^0, x^1, x^2, \dots, x^n\}$ имеет место $x^0 = x_0$ и $x^n \in F$.

Траекторию $T = \{x^0, x^1, x^2, \dots, x^n\}$ называем заключительной x -траекторией, если $x^0 = x$ и $x^n \in F$. Заключительная x -траектория, имея своим начальным состоянием произвольное состояние x системы Ω , заканчивается в одном из финальных состояний системы.

Траекторию $T = \{x^0, x^1, x^2, \dots, x^n\}$ называем начальной x -траекторией, если $x^0 = x_0$ и $x^n = x$. Начальная x -траектория, имея своим конечным состоянием x , начинается от начального состояния системы.

Состояние x системы Ω именуем достижимым, если имеет место $P(x^0, x)$. Отметим, что начальное состояние x_0 достижимо по определению.

Состояние x системы Ω именуем продуктивным, если для некоторого финального состояния f имеет место $P(x, f)$. Финальные состояния системы продуктивны по определению.

Пусть $T = \{x^0, x^1, x^2, \dots, x^n\}$ – произвольная траектория системы Ω , причем:

$$x^t = f(x^{t-1}, v^t), \text{ где } v^t \in V(x^{t-1}), t = 1, 2, \dots, n.$$

Стоимость $C(T)$ траектории T определяем следующим образом:

$$C(T) = \sum_{t=1}^n s(x^{t-1}, v^t).$$

Таким образом, стоимость траектории – это сумма пошаговых платежей, имеющих место при реализации траектории. Центральной является задача построения полной траектории, оптимальной по значению суммарной стоимости, т.е. обеспечивающей максимальный суммарный доход или минимальный суммарный расход. Обе экстремальные задачи по способу решения идентичны. Для определенности проблему формулируем следующим образом.

Задача А. Для системы Ω найти полную траекторию T минимальной стоимости. Дополнительно введем в рассмотрение совокупность частных задач $A(x)$, где x принадлежит $D \setminus F$.

Задача $A(x)$. Для системы Ω и ее состояния x найти заключительную x -траекторию минимальной стоимости.

Решения сформулированных экстремальных задач A и $A(x)$ именуем оптимальной полной траекторией и оптимальной заключительной x -траекторией соответственно.

В основе метода динамического программирования лежит принцип оптимальности, сформулированный американским математиком Ричардом Беллманом следующим образом: «Оптимальное поведение обладает тем свойством, что каковы бы ни были первоначальное состояние и решение в начальный момент, последующие решения должны составлять оптимальное поведение относительно состояния, являющегося результатом первого решения».

Следующая теорема выражает принцип Беллмана в применении к задачам построения оптимальных траекторий.

Теорема. Если траектория $T = \{x^0, x^1, x^2, \dots, x^n\}$ оптимальна, то любая ее заключительная часть $T_k = \{x^k, x^{k+1}, \dots, x^n\}$ также оптимальна.

Стоимость оптимальной в задаче A полной траектории обозначим символом B_A . Стоимость заключительной x -траектории, оптимальной в задаче $A(x)$, обозначим $B(x)$. Функция $B(x)$ называется функцией Беллмана. Отметим, что $B_A = B(x_0)$.

Очевидны следующие равенства:

$$B(x) = 0, x \in F. \quad (6.1)$$

Пусть x - произвольное нефинальное состояние системы. В этом состоянии можно реализовать любое принадлежащее множеству $V(x)$ управление. Предположим, что выбрано управление $v, v \in V(x)$. Данное управление влечет платеж $s(x, v)$, а следующим состоянием системы оказывается $f(x, v)$. Если далее реализуется оптимальная заключительная $f(x, v)$ - траектория, то суммарная величина последующих платежей оказывается равной $B(f(x, v))$. Из принципа Беллмана получаем:

$$B(x) = \min_{v \in V(x)} \{s(x, v) + B(f(x, v))\}, x \in D \setminus F. \quad (6.2)$$

Вычисление значений функции Беллмана по соотношениям (6.1) - (6.2) выполняется поэтапно в следующем порядке.

На первом этапе фиксируются значения $B(x) = 0$ для всех $x \in F$.

Далее на каждом следующем этапе вычисление очередного значения функции Беллмана выполняется для произвольного состояния x такого, что $B(x)$ неизвестно, но значения $B(y)$ для всех непосредственно следующих за x состояний y уже найдены (состояние y системы Ω относим к числу непосредственно следующих за состоянием x , если пара $\{x, y\}$ является траекторией системы Ω).

Последним в процессе счета определяется значение $B(0)$.

Для синтеза оптимальной полной траектории системы при выполнении, определяемой соотношениями (6.1)-(6.2), процедуры счета следует дополнительно составлять список оптимальных переходов (СОП). Каждая запись этого списка имеет вид $[x_i, x_j]$, где x_i - произвольное нефинальное состояние, а $x_j = f(x_i, v^*)$ - состояние, в которое система переходит из x_i под воздействием управления v^* , возможного в состоянии x_i и обращающего в минимум правую часть (6.2); если таких управлений несколько, фиксируется любое из них (нашей целью считаем построение любой из оптимальных траекторий, но не всех таких траекторий). Записи СОП взаимно однозначно соответствуют нефинальным состояниям системы, при этом запись $[x_i, f(x_i, v)]$ считается соответствующей состоянию x_i . Составленный СОП полагаем упорядоченным по возрастанию индексов первых компонент входящих в него записей. Чтобы построить оптимальную траекторию, из СОП извлекаем последовательность записей W , первым элементом которой является запись, соответствующая состоянию x^0 ; каждая следующая запись последовательности W имеет своей первой компонентой вторую компоненту предыдущей записи этой последовательности. В последней записи последовательности W второй компонентой является некоторое финальное состояние системы Ω . Из составленного СОП последовательность W извлекается однозначно.

Пусть $W = \{[x_0, y_1], [y_1, y_2], [y_2, y_3], \dots, [y_{m-1}, y_m]\}$, здесь $y_m \in F$. Тогда $T = \{x_0, y_1, y_2, y_3, \dots, y_m\}$ - искомая оптимальная траектория.

Уравнения (6.1) - (6.2) - рекуррентные соотношения динамического программирования для решения задачи A . Основанный на этих соотношениях метод решения задачи A называется методом динамического программирования.

Пусть N - число состояний системы Ω . Тогда верхняя оценка числа вычисляемых значений функции Беллмана в процессе решения задачи A равна N . Число элементарных операций, выполняемых при определении по формуле (6.2) каждого конкретного значения этой функции не превышает CN , где C - некоторая не зависящая от N константа. Таким образом, верхней оценкой числа элементарных операций, выполняемых при решении задачи A методом динамического программирования является CN^2 , где N - число состояний системы Ω , а C - не зависящая от N константа. Для запоминания в процессе

вычислений всех найденных значений функции Беллмана необходим объем памяти, пропорциональный N ; объем памяти такого же порядка нужен для создания списка оптимальных переходов. В итоге получаем, что объем памяти, нужной для решения задачи A методом динамического программирования, линейно зависит от N .

Следует отметить, что в связи с конечностью числа состояний системы Ω число ее траекторий конечно и задача A в принципе может быть решена путем перебора конечного числа вариантов. Метод динамического программирования позволяет определенным образом упорядочить и существенно сократить перебор.

ПРИМЕР 1. Имеется дискретно управляемая система ΩI , определяющий ее граф $G(\Omega I)$ представлен на рисунке 1.

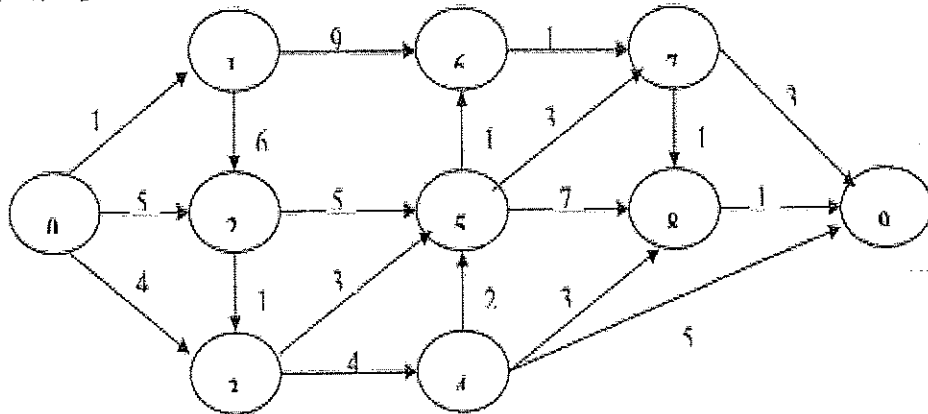


Рисунок 1 – Граф $G(\Omega I)$

Начальным является состояние 0, множество финальных состояний одноэлементно: $F=\{9\}$. Требуется найти полную траекторию минимальной стоимости.

Для состояний системы ΩI (вершин графа $G(\Omega I)$) вычисляем значения функции Беллмана. На основании (6.1) фиксируем $B(9)=0$. Далее, пользуясь (6.2), последовательно получаем:

$$B(8)=1; \quad B(7)=\min[3, 1+B(8)]=2; \quad B(6)=1+B(7)=3;$$

$$B(5)=\min[1+B(6), 3+B(7), 7+B(8)]=\min[4, 6, 8]=4;$$

$$B(4)=\min[2+B(5), 3+B(8), 5+B(9)]=\min[6, 4, 5]=4;$$

$$B(3)=\min[3+B(5), 4+B(4)]=\min[7, 8]=7;$$

$$B(2)=\min[1+B(3), 5+B(5)]=\min[8, 9]=8;$$

$$B(1)=\min[9+B(6), 6+B(2)]=\min[12, 14]=12;$$

$$B(0)=\min[1+B(1), 5+B(2), 4+B(3)]=\min[13, 13, 11]=11.$$

Таким образом, минимальная из стоимостей полных траекторий в рассматриваемой задаче равна 11. Получаемый результат представлен на рисунке 2.

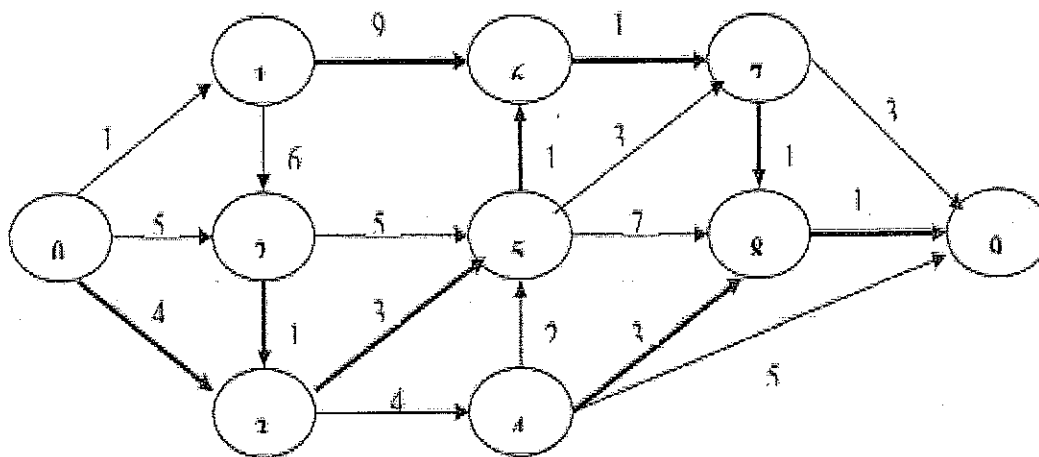


Рисунок 2 – Граф $G(\Omega)$ с выделенными дугами

При подсчетах значений функции Беллмана в графе $G(\Omega)$ для каждой нефинальной вершины специально выделяем (рисуем жирной линией) дугу, соответствующую управлению, в ней принимаемому (т.е. обращающему в минимум правую часть (6.2)).

Как показывают выделенные дуги, на первом такте процесса управления следует выполнить переход из вершины 0 в вершину 3, на втором такте - из вершины 3 в вершину 5, на третьем такте - из вершины 5 в вершину 6, на четвертом такте - из вершины 6 в вершину 7, на пятом такте - из вершины 7 в вершину 8, на шестом, заключительном такте - из вершины 8 в вершину 9.

Таким образом, оптимальной по критерию суммарной стоимости полной траекторией является $T = \{0, 3, 5, 6, 7, 8, 9\}$.

Далее введем совокупность частных задач следующего вида.

Задача $A^*(x)$. Для системы Ω и ее состояния x найти начальную x -траекторию минимальной стоимости.

Оптимальное решение задачи $A^*(x)$ именуем оптимальной начальной x -траекторией системы Ω .

Имеет место следующий факт.

Теорема 2. Если траектория $T = \{x^0, x^1, x^2, \dots, x^n\}$ оптимальна, то любая ее начальная часть также оптимальна.

Стоимость начальной x -траектории, оптимальной в задаче $A^*(x)$, обозначим $B^*(x)$. Очевидно, что:

$$B^*(x_0) = 0. \quad (6.3)$$

В системе Ω произвольное состояние x непосредственно предшествует состоянию y , если пара $\{x, y\}$ является траекторией системы Ω . Управление, переводящее систему Ω из состояния x непосредственно в состояние y , обозначим $v[x, y]$. Множество состояний системы, непосредственно предшествующих состоянию y , обозначим $\Gamma(y)$. Основываясь на теореме 2, запишем соотношение, позволяющее организовать процесс вычислений значений функции $B^*(y)$ для всех состояний y , отличных от начального:

$$B^*(y) = \min_{x \in \Gamma(y)} \{B^*(x) + s(x, v[x, y])\}, \quad y \in D \setminus x_0; \quad (6.4)$$

Отметим, что:

$$B_A = \min_{x \in F} B^*(x). \quad (6.5)$$

Вычисление значений функции $B^*(y)$ на основе рекуррентных соотношений (6.3) -

(6.4) выполняется поэтапно в следующем порядке. На первом этапе фиксируется значение $V^*(x_0)=0$. Далее на каждом следующем этапе вычисление очередного значения функции V^* выполняется для произвольного состояния y такого, что $V^*(y)$ неизвестно, но значения $V^*(x)$ для всех состояний x , непосредственно предшествующих состоянию y , уже найдены.

Метод решения задачи A , основанный на соотношениях (6.3) – (6.5), – *прямой метод Беллмана*. В отличие от него, метод, основанный на соотношениях (6.1) – (6.2), носит название *обратного метода Беллмана*. Функции $V^*(x)$ и $V(x)$ будем именовать функциями Беллмана для прямого и обратного счета соответственно. Реализация как обратного, так и прямого метода Беллмана для решения задачи A требует квадратично зависящего от числа состояний системы Ω количества элементарных операций. Иногда прямой и обратный метод Беллмана, подразумевая, что задача решается с использованием соотношений динамического программирования, называют *методами прямого и обратного счета* соответственно.

ПРИМЕР 2. Имеется дискретно управляемая система система $\Omega 1$, определяющий ее граф $G(\Omega 1)$ представленный на рисунке 3.

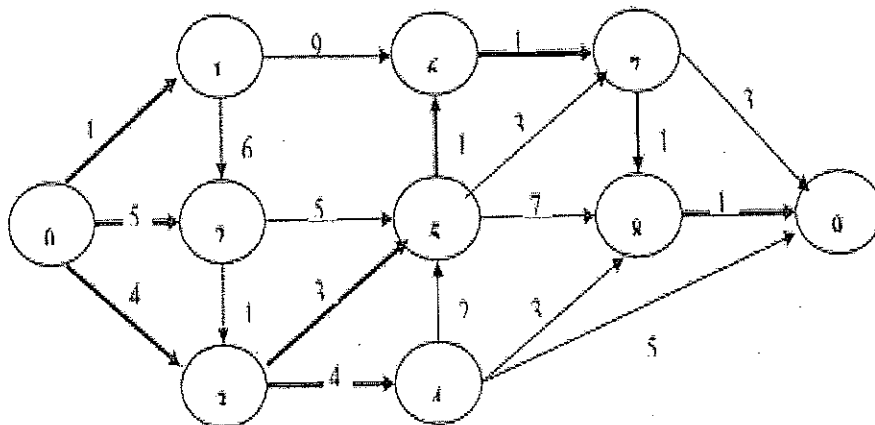


Рисунок 3. – Граф $G(\Omega 1)$ с выделенными дугами.

Начальным является состояние 0, множество финальных состояний одноэлементно: $F=\{9\}$.

Методом прямого счета требуется найти полную траекторию минимальной стоимости.

Для состояний системы $\Omega 1$ (вершин графа $G(\Omega 1)$) вычисляем значения прямой функции Беллмана. На основании (6.3) фиксируем $V^*(0)=0$.

Далее, пользуясь (6.4), последовательно получаем:

$$V^*(1)=1;$$

$$V^*(2)=\min[5, V^*(1)+6]=5;$$

$$V^*(3)=\min[4, V^*(2)+1]=4;$$

$$V^*(4)=V^*(3)+4=8;$$

$$V^*(5)=\min[V^*(2)+5, V^*(3)+3, V^*(4)+2]=7;$$

$$V^*(6)=\min[V^*(1)+9, V^*(5)+1]=8;$$

$$V^*(7)=\min[V^*(6)+1, V^*(5)+3]=9;$$

$$V^*(8)=\min[V^*(5)+7, V^*(7)+1]=10;$$

$$V^*(9)=\min[V^*(7)+3, V^*(8)+1, V^*(4)+5]=11.$$

При вычислении значений функции Беллмана в графе $G(\Omega 1)$ для каждой вершины y , отличной от начальной, специально выделяем (рисуем жирной линией) дугу (x, y) , где x

- вершина для которой значение правой части (6.4) при подсчете $V^*(y)$ оказывается минимальным (этот способ выделения дуг с учетом используемого метода счета называем *прямым*; способ выделения, использованный в решении предыдущего примера, именуем *обратным*). Полученный результат представлен на рисунке 3.

Как показывают выделенные дуги, на последнем такте процесса управления следует выполнить переход из вершины 8 в вершину 9, при этом переход в вершину 8 реализуется из вершины 7, в вершину 7 - из вершины 6, в вершину 6 - из вершины 5, в вершину 5 - из вершины 3, в вершину 3 - из вершины 0. Оптимальной является траектория $\{0, 3, 5, 6, 7, 8, 9\}$.

Полученные в результате решения примеров 1 и 2 оптимальные в системе ΩI траектории совпали.

Отметим следующее обстоятельство. Пусть в процессе вычислений по соотношениям (6.3) - (6.4) для некоторого финального состояния x^* оказалось найденным значение $V^*(x^*)$; полную траекторию стоимости $V^*(x^*)$, имеющую своим конечным состоянием x^* , обозначим T^* . Если каждая из остальных траекторий имеет в своем составе как минимум одно состояние из некоторого подмножества состояний Q , причем все значения $V^*(x)$ при $x \in Q$ уже вычислены и оказались не меньшими $V^*(x^*)$, то T^* - полная оптимальная траектория. Сделанное утверждение нередко дает возможность ограничить объем требуемых вычислений.

Подмножество промежуточных состояний M системы Ω назовем *разрезающим*, если каждая полная траектория T этой системы имеет в своем составе состояния из M .

Разрезающее подмножество M^* назовем *минимальным разрезающим*, если среди его собственных подмножеств нет разрезающих.

Минимальная, из стоимостей полных траекторий системы Ω равна:

$$\min_{x \in M^*} (V^*(x) + V(x)), \quad (6.6)$$

здесь M^* - любое минимальное разрезающее подмножество состояний системы Ω .

Основанный на данном факте метод решения задачи синтеза полной траектории минимальной стоимости называется *методом встречного счета*.

Для системы ΩI , определяемой представленным на рисунке 2 графом $G(\Omega I)$, минимальным разрезающим является, в частности, подмножество вершин-состояний $M^* = \{4, 5, 6\}$. При решении задачи синтеза оптимальной по стоимости полной траектории методом встречного счета выполняем следующие действия.

Прямым счетом последовательно определяем:

$$V^*(1)=1; V^*(2)=5; V^*(3)=4; V^*(4)=8; V^*(5)=7; V^*(6)=8.$$

Обратным счетом находим:

$$V(8)=1; V(7)=2; V(6)=3; V(5)=4; V(4)=4.$$

Для принадлежащих подмножеству M^* состояний находим:

$$V^*(4)+V(4)=12; V^*(5)+V(5)=11; V^*(6)+V(6)=11.$$

Получаем, что минимальная из стоимостей полных траекторий системы ΩI равна 11.

Оптимальные полные траектории системы ΩI (их, в принципе, может быть несколько) проходят через вершины 5 и 6. При выполненном подсчете значений функции $V^*(x)$, см. решение примера 2, определено, что оптимальной начальной 5-траекторией системы ΩI является $T^*(5) = \{0, 3, 5\}$, а оптимальной начальной 6-траекторией этой системы является $T^*(6) = \{0, 3, 5, 6\}$. При подсчете значений функции $V(x)$, см. решение примера 1, установлено, что оптимальной заключительной 5-траекторией системы ΩI

является $T(5) = \{5, 6, 7, 8, 9\}$, а оптимальной заключительной b -траекторией этой системы является $T(6) = \{6, 7, 8, 9\}$. Получаем, что единственной оптимальной полной траекторией системы ΩI является $T = \{0, 3, 5, 6, 7, 8, 9\}$.

1.2. Решение дискретных оптимизационных задач методом динамического программирования

Множество вариантов конечно, но число их, как правило, уже при относительно малых размерностях решаемых задач решение оказывается столь большим, что полный перебор практически неосуществим. Использование схемы динамического программирования дает возможность определенным образом упорядочить перебор, существенно сократив при этом число подлежащих рассмотрению вариантов.

Задача оптимального распределения капиталовложений

Пусть имеются денежная сумма (капитал) C и возможные сферы вложений S_1, S_2, \dots, S_n . Считается, что в каждую сферу S_i может быть вложен целочисленный вклад, не превосходящий константы C_i . Для каждой сферы S_i полагается известной монотонно возрастающая в нестрогом смысле функция $f_i(u)$, определяющая величину дохода, получаемого при вложении в данную сферу вклада u .

Требуется определить обеспечивающее максимальный суммарный доход распределение суммы C или ее части по сферам вложений.

Математическая модель задачи следующая:

$$\sum_{i=1}^n f_i(u_i) \rightarrow \max, \quad (6.7)$$

при условиях:

$$\sum_{i=1}^n u_i \leq C;$$

$$u_t \in \{0, 1, \dots, C_t\}, t = \overline{1, n}. \quad (6.8-6.9)$$

Сформулированную задачу можно трактовать как задачу поиска полной траектории, обеспечивающей максимальный суммарный доход соответствующим образом построенной дискретной управляемой системе.

Исходные данные задачи (ЗОРК) определяют систему $\Omega(\text{ЗОРК})$, управления в которой осуществляется в дискретном времени $t=0, 1, 2, \dots, n-1$ (в момент t определяется вклад в сферу $St+1$).

Состояния системы - пары (t, U) , где t - момент дискретного времени, а U - сумма, которая к данному моменту еще не распределена, здесь $t \in \{0, 1, \dots, n\}$, $U \in \{0, 1, \dots, C\}$. Начальное состояние системы $(0, C)$; финальными являются состояния вида (n, U) . В любом нефинальном состоянии (t, U) выбор реализуемого в этом состоянии управления u_{t+1} осуществляется из совокупности $V(t, U) = \{0, 1, \dots, \min(C_{t+1}, U)\}$; выбранная величина u_{t+1} - размер вклада в сферу $St+1$. Под воздействием управления $ut+1$ система $\Omega(\text{ЗОРК})$ из состояния (t, U) переходит в состояние $(t+1, U - u_{t+1})$, получаемый при этом доход равен $f_{t+1}(u_{t+1})$. Каждая последовательность управлений, переводящая систему из начального состояния в одно из финальных состояний, определяет некоторое допустимое решение ЗОРК (распределение (u_1, u_2, \dots, u_n) капиталовложений).

Исходную ЗОРК трактуем как задачу отыскания траектории системы $\Omega(\text{ЗОРК})$, обеспечивающей максимальный суммарный доход. Функция Беллмана для системы $\Omega(\text{ЗОРК})$ определяется соотношениями:

$$B(n, U) = 0, \quad U \in \{0, 1, \dots, C\}; \quad (6.10)$$

$$B(t, U) = \max_{u_{t+1} \in \{0, 1, \dots, \min(C_{t+1}, U)\}} \{f_{t+1}(u_{t+1}) + B(t+1, U - u_{t+1})\}, \quad (6.11)$$

здесь $t \in \{0, 1, \dots, n-1\}$, $U \in \{0, 1, \dots, C\}$.

При реализации обратного метода Беллмана процесс вычислений реализуется в соответствии с записанными соотношениями. Зная, что $B(n, U) = 0$ при всех U , по формуле (6.11) находим все значения $B(n-1, U)$ при $U \in \{0, 1, \dots, S\}$; здесь оказывается справедливым тождество $B(n-1, U) = f_n(\min(C_n, U))$. Далее, имея вычисленными значения $B(n-1, U)$, по той же формуле (6.11) определяем все значения $B(n-2, U)$, где $U \in \{0, 1, \dots, S\}$, и т.д. до тех пор, пока не окажется найденной величина $B(0, C)$ - оптимальное значение критерия в решаемой ЗОРК: В процессе вычислений, с целью обеспечения возможности определить оптимальное распределение капиталовложений, после отыскания каждого следующего значения $B(t, U)$, где $t \in \{0, 1, \dots, n-1\}$ и $U \in \{0, 1, \dots, C\}$, для пары (t, U) отдельно фиксируем значение u_{t+1} , на котором достигается максимум правой части (6.11).

Прямой метод Беллмана для решения ЗОРК заключается в последовательном вычислении значений функции $B^*(t, U)$ в порядке возрастания значений первого аргумента. При этом:

$$B^*(0, U) = 0, U \in \{0, 1, \dots, C\}; \quad (6.12)$$

$$B^*(t+1, U) = \max_{u_{t+1} \in \{0, 1, \dots, \min(C_{t+1}, U)\}} \{B^*(t, U - u_{t+1}) + f_{t+1}(u_{t+1})\}, \quad (6.13)$$

здесь $t \in \{0, 1, \dots, n-1\}$, $U \in \{0, 1, \dots, C\}$.

Основанные на формуле (6.13) вычисления заканчиваются отысканием величин $B^*(n, U)$, где $U \in \{1, 2, \dots, C\}$; максимальная из найденных величин - оптимальное значение критерия в решаемой ЗОРК.

ПРИМЕР 3. Имеются сферы капиталовложений S_1, S_2, S_3, S_4 ; в каждую из них может быть внесен целочисленный вклад, не превосходящий 5. Величина подлежащего распределению капитала равна 6. Функции, определяющие доходы, получаемые при вложении в различные сферы вклада u , заданы таблицей 1.

Таблица 1.

Значения функций дохода $f_i(u)$.

	$F_1(u)$	$F_2(u)$	$F_3(u)$	$F_4(u)$
$u = 0$	0	0	0	0
$u = 1$	0,2	0	0,1	0
$u = 2$	0,2	0	0,2	0
$u = 3$	0,3	0,4	0,3	0,3
$u = 4$	0,3	0,4	0,4	0,4
$u = 5$	0,5	0,4	0,5	0,5

Требуется найти оптимальное распределение капитала между сферами вложений.

Решение данного примера можно выполнить как прямым, так и обратным методом Беллмана. Будем основываться на рекуррентных соотношениях (6.10) - (6.11), т.е. применим обратный метод. Отыскиваемые в процессе счета значения функции $B(t, U)$ вносим в таблицу 2. В каждую заполняемую клетку (t, U) этой таблицы (параметр t определяет столбец, параметр U - строку) одновременно с $B(t, U)$ вносится и записанное в квадратных скобках значение u_{t+1} , на котором достигается максимум правой части (6.11) при подсчете $B(t, U)$; если таких значений несколько, вносится любое из них (нашей целью является отыскание, какого-либо из оптимальных решений, а не полной их совокупности).

В связи с тем, что $B(4, U)$ тождественно равно нулю, процесс вычислений начинаем с определения значений $B(3, U)$ при всех возможных значениях U . Из равенства (6.11)

следует, что $B(3, U) = f_4(U)$ при $U = 0, 1, 2, 3, 4, 5$ и $B(3, 6) = f_4(5)$. Так заполняется последний столбец таблицы 2.

Таблица 2.

Значения функции $B(t, U)$.

	t=0	t=1	t=2	t=3
U=0		0 [0]	0 [0]	0 [0]
U=1		0,1 [0]	0,1 [1]	0 [0]
U=2		0,2 [0]	0,2 [2]	0 [0]
U=3		0,4 [3]	0,3 [3]	0,3 [3]
U=4		0,5 [3]	0,4 [4]	0,4 [4]
U=5		0,6 [3]	0,5 [5]	0,5 [5]
U=6	0,8 [1]	0,7 [3]	0,6 [1]	0,5 [5]

То же равенство (6.11) дает возможность последовательного (справа налево) заполнения остальных столбцов. Заметим, что для решения задачи в первом столбце достаточно найти только один, нижний элемент. В рассматриваемом примере оказалось, что $B(0, 6) = 0,8$. Это оптимальное значение критерия (максимально возможный доход). В клетке (0,6) в квадратных скобках записано число 1. Значит, в оптимальном распределении капиталовложений взнос в первую сферу должен равняться единице. В таком случае суммарный взнос в остальные сферы не превосходит 5. В клетке (1,5) в квадратных скобках записано число 3. Значит, в синтезируемом оптимальном распределении капиталовложений взнос во вторую сферу равняется трем. Получаем, что суммарный взнос в третью и четвертую сферы не превосходит 2. В клетке (2,2) в квадратных скобках записано число 2. В оптимальном распределении взнос в третью сферу должен равняться двум. В итоге получаем, что оптимальное распределение делит сумму $C = 6$ между первой, второй и третьей сферами капиталовложений; вносимые в них вклады равны соответственно 1, 3, 2. При этом первая сфера приносит доход 0,2; вторая сфера дает доход 0,4 и третья сфера - доход 0,2. Суммарный доход действительно равен 0,8.

Классическая задача коммивояжера

Имеется полный взвешенный ориентированный граф без петель G с множеством вершин $N = \{1, 2, \dots, n\}$; веса всех дуг неотрицательны. Требуется найти гамильтонов цикл минимального веса.

Исходную информацию по $ЗК$ считаем представленной в виде $(n \times n)$ - матрицы $S = \{s_{ij}\}$, где s_{ij} – вес дуги (i, j) графа G , $i = 1, n$, $j = 1, n$, $i \neq j$; все элементы главной диагонали матрицы S являются нулями.

В типовой интерпретации вершины $1, 2, \dots, n$ графа G – это города. Дуги отображают возможные элементарные переходы. Коммивояжеру, изначально находящемуся в городе 1 , необходимо обойти все остальные города, побывав в каждом из них ровно по одному разу, и затем вернуться в город 1 . Веса дуг графа трактуются как длины соответствующих элементарных переходов. Требуется найти имеющий минимальную длину допустимый (т.е. удовлетворяющий наложенным требованиям) маршрут коммивояжера.

С учетом других возможных интерпретаций, на матрицу S требование симметричности не налагается, не считается обязательным и выполнение неравенства треугольника.

Один из основных алгоритмов решения $ЗК$ основан на принципе динамического программирования. При изложении этого алгоритма будем придерживаться терминологии, соответствующей приведенной типовой интерпретации задачи.

Пусть i – произвольный город ($i \in N$), а V – любое подмножество городов, не содержащее города 1 и города i . Через $M(i, V)$ обозначим совокупность путей, каждый из которых начинается в городе i , завершается в городе 1 и проходит в качестве промежуточных только через города множества V , заходя в каждый из них ровно по одному разу. Через $B(i, V)$ обозначим длину кратчайшего пути множества $M(i, V)$.

Для решаемой задачи $B(i, V)$ – функция Беллмана. Как очевидно, $B(1, \{2, 3, \dots, n\})$ – искомая минимальная длина простого (без самопересечений) замкнутого пути, проходящего через все города.

Если V – одноэлементное множество, $V = \{j\}$, где $j \neq 1$ и $j \neq i$, то совокупность $M(i, V)$ состоит из единственного пути $\pi = (i, j, 1)$. Поэтому:

$$B(i, \{j\}) = s_{ij} + s_{j1}, \quad i \in N, \quad j \in \{2, 3, \dots, n\}, \quad j \neq i. \quad (6.18)$$

Предположим, что значения функции $B(i, V)$ для всех $i \in N \setminus \{1\}$ и всех возможных k -элементных ($k < n-1$) множеств V уже вычислены. Тогда значение $B(i, V')$, где V' – произвольное $(k+1)$ -элементное подмножество совокупности $N \setminus \{1, i\}$, вычисляется по формуле:

$$B(i, V') = \min_{j \in V'} (s_{ij} + B(j, V' \setminus \{j\})). \quad (6.19)$$

Уравнения (6.18)-(6.19) – рекуррентные соотношения динамического программирования для решения задачи коммивояжера, они реализуют обратный метод Беллмана.

ПРИМЕР 4. Решить задачу коммивояжера, определяемую матрицей:

$$S = \begin{pmatrix} 0 & 5 & 1 & 4 \\ 2 & 0 & 3 & 5 \\ 1 & 4 & 0 & 2 \\ 3 & 5 & 4 & 0 \end{pmatrix}$$

Сначала, пользуясь формулой (6.18), определяем значения $B(i, \{j\})$:

$$\begin{aligned} B(2, \{3\}) &= 4; & B(2, \{4\}) &= 8; & B(3, \{2\}) &= 6; \\ B(3, \{4\}) &= 5; & B(4, \{2\}) &= 7; & B(4, \{3\}) &= 5. \end{aligned}$$

Далее по формуле (6.19) последовательно получаем (в левой части каждого из ниже записанных равенств подчеркнуты те значения параметра j , на которых при подсчете реализуется указанный в правой части (6.19) минимум):

$$\begin{aligned} B(2, \{3, 4\}) &= \min(3 + 5, 5 + 5) = 8; \\ B(3, \{2, 4\}) &= \min(4 + 8, 2 + 7) = 9; \\ B(4, \{2, 3\}) &= \min(5 + 4, 4 + 6) = 9; \\ B(1, \{2, 3, 4\}) &= \min(5 + 8, 1 + 9, 4 + 9) = 10. \end{aligned}$$

Итак, оптимальное значение критерия в рассматриваемом примере равно 10. Выполненные подчеркивания позволяют определить оптимальный маршрут. Он следующий:

$$1 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1.$$

Для записи соотношений, по которым реализуется прямой метод Беллмана, введем новые обозначения. Пусть $M'(V, i)$ – совокупность путей, каждый из которых начинается в городе 1 , проходит в качестве промежуточных только через города подмножества V , заходя в каждый из них ровно по одному разу, и завершается в городе i ; здесь, как и ранее, i – произвольный город ($i \in N$), а V – любое подмножество N , не содержащее городов 1 и i . Длину кратчайшего пути множества $M'(V, i)$ обозначим $B^*(V, i)$. Как очевидно, $B^*(\{2, 3, \dots, n\}, 1)$ – искомая минимальная длина простого (без самопересечений) замкнутого пути, проходящего через все города. Если V – одноэлементное множество, $V = \{j\}$, где $j \neq 1$ и $j \neq i$, то совокупность $M'(V, i)$ состоит из единственного пути $\pi = (1, j, i)$. Поэтому:

$$B^*(\{j\}, i) = s_{1j} + s_{ji}, \quad i \in N, \quad j \in \{2, 3, \dots, n\}, \quad j \neq i. \quad (6.20)$$

Предположим, что значения функции $B^*(V, i)$ для всех $i \in N$ и всех возможных k -элементных ($k < n-1$) множеств V уже вычислены. Тогда значение $B^*(V', i)$, где V' – произвольное $(k+1)$ -элементное подмножество совокупности $N \setminus \{i\}$, вычисляется по формуле:

$$B^*(V', i) = \min_{j \in V'} (B^*(V' \setminus \{j\}, j) + s_{ij}). \quad (6.21)$$

Уравнения (6.20)-(6.21) – рекуррентные соотношения динамического программирования для решения классической задачи коммивояжера, они реализуют прямой метод Беллмана. Заметим, что при решении задачи коммивояжера обратным методом состояния моделирующей системы – пары вида (i, V) , где i – город, в котором сейчас находится коммивояжер, а V – множество городов, которые ему еще предстоит обойти перед тем, как вернуться в исходный город 1. При решении задачи коммивояжера прямым методом состояния моделирующей системы – пары вида (V, i) , где i – город, в котором коммивояжер находится сейчас, а V – множество городов, которые коммивояжер уже обошел после того, как вышел из города 1. По сути, (i, V) и $(\{2, 3, \dots, n\} \setminus (V \cup \{i\}), i)$ – это разные обозначения одной и той же ситуации. Данный факт используется при решении рассматриваемой задачи методом встречного счета.

ПРИМЕР 1.5. Методом встречного счета решить задачу коммивояжера, определяемую матрицей:

$$S = \begin{pmatrix} 0 & 5 & 1 & 4 \\ 2 & 0 & 3 & 5 \\ 1 & 4 & 0 & 2 \\ 3 & 5 & 4 & 0 \end{pmatrix}$$

В качестве разрезающего выберем множество, включающее следующие состояния:

- $(2, \{3\})$ /совпадает с $(\{4\}, 2)$ /;
- $(2, \{4\})$ /совпадает с $(\{3\}, 2)$ /;
- $(3, \{2\})$ /совпадает с $(\{4\}, 3)$ /;
- $(3, \{4\})$ /совпадает с $(\{2\}, 3)$ /;
- $(4, \{2\})$ /совпадает с $(\{3\}, 4)$ /;
- $(4, \{3\})$ /совпадает с $(\{2\}, 4)$ /.

Реализуя метод обратного счета, определяем:

$$B(2, \{3\}) = 4; B(2, \{4\}) = 8; B(3, \{2\}) = 6; B(3, \{4\}) = 5; B(4, \{2\}) = 7; B(4, \{3\}) = 5.$$

Реализуя метод прямого счета, получаем:

$$B^*(\{4\}, 2) = 9; B^*(\{3\}, 2) = 5; B^*(\{4\}, 3) = 8; B^*(\{2\}, 3) = 8; B^*(\{3\}, 4) = 3; B^*(\{2\}, 4) = 10.$$

Далее имеем:

$$\begin{aligned} B(2, \{3\}) + B^*(\{4\}, 2) &= 13; \\ B(2, \{4\}) + B^*(\{3\}, 2) &= 13; \\ B(3, \{2\}) + B^*(\{4\}, 3) &= 14; \\ B(3, \{4\}) + B^*(\{2\}, 3) &= 13; \\ B(4, \{2\}) + B^*(\{3\}, 4) &= 10; \\ B(4, \{3\}) + B^*(\{2\}, 4) &= 15. \end{aligned}$$

Минимальное, равное 10, значение правая часть принимает в пятом равенстве. Отсюда следует:

- а) оптимальное значение критерия в рассматриваемой задаче равно 10;
- б) оптимальная траектория системы проходит через состояние $(\{3\}, 4)$ или, что то же самое, через состояние $(4, \{2\})$. Последнее дает возможность установить оптимальный

маршрут коммивояжера:

$$1 \rightarrow 3 \rightarrow 4 \rightarrow 2 \rightarrow 1.$$

Применение метода встречного счета в сравнении с прямым или обратным счетом тем более выгодно, чем больше размер решаемой задачи.

Классическая задача о ранце (КЗР).

Имеются ранец и предметы $П_1, П_2, \dots, П_n$; каждый предмет $П_i$ характеризуется двумя целыми положительными показателями: c_i – стоимость предмета; v_i – вес предмета, $i = 1, n$. Предметы недробимы, т.е. каждый из них можно поместить в ранец только целиком. Требуется найти совокупность предметов, которые следует поместить в ранец с тем, чтобы суммарная стоимость предметов в ранце оказалась максимальной при условии, что суммарный вес предметов в ранце не может превысить заданной натуральной константы W (предполагается, что суммарный вес всех имеющихся предметов больше W). КЗР формулируется как следующая задача булева линейного программирования (БЛП):

$$\sum_{i=1}^n c_i x_i \rightarrow \max \quad (6.29)$$

$$\sum_{i=1}^n v_i x_i \leq W;$$

$$x_i \in \{0, 1\}, i = \overline{1, n}. \quad (1.30-1.31)$$

Задачу (1.29)-(1.31) будем именовать задачей Z .

По начальным данным задачи Z формулируем совокупность частных задач о ранце $Z(k, p)$, где $k \in \{1, 2, \dots, n\}$; $p \in \{1, 2, \dots, W\}$.

Задача $Z(k, p)$ записывается следующим образом:

$$\sum_{i=1}^k c_i x_i \rightarrow \max$$

при условиях:

$$\sum_{i=1}^k v_i x_i \leq p;$$

$$x_i \in \{0, 1\}, i = \overline{1, k}.$$

(в частной задаче $Z(k, p)$ имеющимися в наличии считаются только первые k предметов из совокупности $\{П_1, П_2, \dots, П_n\}$, суммарный вес предметов в ранце не может превышать константы p). Оптимальное значение критерия в частной задаче $Z(k, p)$ обозначим $B^*(k, p)$. Как очевидно, задача $Z(n, W)$ совпадает с исходной задачей Z ; поэтому $B^*(n, W)$ – оптимальное значение критерия в задаче Z .

Легко видеть, что:

$$0 \text{ при } p \in \{0, 1, 2, \dots, v_1 - 1\};$$

$$B^*(1, p) =$$

$$c_1 \text{ при } p \in \{v_1, v_1 + 1, \dots, W\}. \quad (6.32)$$

Пусть для некоторого k , принадлежащего множеству $\{1, 2, \dots, n-1\}$ и всех $p \in \{1, 2, \dots, W\}$ значения функции $B^*(k, p)$ уже вычислены. При $p \in \{0, 1, 2, \dots, v_{k+1} - 1\}$ ситуация, возникающая при нахождении $B^*(k+1, p)$, фактически не отличается от ситуации, имевшей место при подсчете $B^*(k, p)$, ибо дополнительно появившийся при переходе от

задачи $Z(k, p)$ к задаче $Z(k+1, p)$ предмет $pk+1$ в ранец поместить невозможно. В случае, когда $p \in \{vk+1, vk+1+1, \dots, W\}$, при переходе от задачи $Z(k, p)$ к задаче $Z(k+1, p)$ мы получаем дополнительную возможность поместить в ранец предмет $pk+1$. Этой возможностью можно: а) не воспользоваться; б) воспользоваться. Очевидно, что в случае а) мы получим уже вычисленную величину $B^*(k, p)$. В случае б) в ранец кладется предмет $pk+1$ стоимостью $ck+1$ и так как вес этого предмета равен $vk+1$, то далее мы оказываемся в условиях задачи $Z(k, p - vk+1)$. В итоге получаем:

$$B^*(k, p), \text{ если } p \in \{0, 1, 2, \dots, vk+1-1\};$$

$$B^*(k+1, p) =$$

$$\max \{ B^*(k, p), c_{k+1} + B^*(k, p - vk+1) \} \text{ при } p \in \{vk+1, vk+1+1, \dots, W\}; \quad (6.33)$$

Равенства (6.32) - (6.33) – рекуррентные соотношения динамического программирования для решения задачи о ранце. Пользуясь ими, мы находим сначала величины $B^*(1, p)$, $p \in \{1, 2, \dots, W\}$, затем величины $B^*(2, p)$, $p \in \{1, 2, \dots, W\}$, и т.д., вплоть до отыскания $B^*(n, W)$ – оптимального значения критерия в исходной задаче Z . Процедура вычисления значений функции $B^*(k, p)$ целесообразно реализовывать как процесс последовательного заполнения строк таблицы стоимостей. В этой таблице строки соответствуют значениям параметра k (по возрастанию), а столбцы – значениям параметра p (также по возрастанию). В каждую клетку, расположенную в пересечении произвольной строки k и произвольного столбца p , вносится значение $B^*(k, p)$. Первая строка заполняется согласно формуле (6.32). Далее заполнение каждой $(k+1)$ -ой строки, $k = 1, 2, \dots, n-1$, осуществляется в соответствии с формулой (1.33). Одновременно целесообразно фиксировать множества предметов $M(k, p)$, которые обеспечивают значения $B^*(k, p)$ критериев рассматриваемых частных задач. Как очевидно, $M(1, p) = \emptyset$, если $p < v_1$; $M(1, p) = \{1\}$ в противном случае. Далее можно считать $M(k, p)$, если $B^*(k+1, p) = B^*(k, p)$; $M(k+1, p) = M(k, p - vk+1) \cup \{k+1\}$ в противном случае. В заполненной таблице стоимостей содержимое клетки (n, W) – оптимальное значение критерия в задаче Z . Для обеспечения этого значения в ранец следует поместить предметы совокупности $M(n, W)$. Изложенный алгоритм решения КЗР далее будем называть алгоритмом Акзр.

ПРИМЕР 6. Решить КЗР:

$$5x_1 + 2x_2 + 4x_3 + 7x_4 + 5x_5 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 \leq 8;$$

$$x_i \in \{0, 1\}, i=1, 5.$$

Задача решается путем последовательного, сверху вниз, заполнения строк таблицы стоимостей (таблица 4).

Таблица 4.

Таблица стоимостей

	1	2	3	4	5	6	7	8
1	0	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)	5 (1)
2	2 (2)	5 (1)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)	7 (1,2)
3	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	11 (1,2,3)	11 (1,2,3)	11 (1,2,3)
4	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	12 (1,4)	14 (1,2,4)	16 (1,3,4)
5	2 (2)	5 (1)	7 (1,2)	9 (1,3)	11 (1,2,3)	12 (1,4)	14 (1,2,4)	16 (1,3,4)

В каждой клетке (k, p) данной таблицы вслед за полученным значением $B^*(k, p)$ в скобках малыми цифрами перечисляются номера предметов, которые согласно оптимальному

решению частной задачи $Z(k, p)$ следует положить в ранец. Если задача $Z(k, p)$ имеет несколько оптимальных решений, то указывается только одно из них.

Заполнение первой строки таблицы выполняется в соответствии с формулой (1.32); здесь считается, что в нашем распоряжении имеется только предмет $П1$, его стоимость равна 5 и вес равен 2. Предмет $П1$ можно положить в ранец, если дозволённый вес предметов в ранце не меньше чем 2. Поэтому $B^*(1, 1) = 0$ и $B^*(1, p) = 5$, если $p \geq 2$.

Заполнение второй и каждой из нижеследующих строк выполняется в соответствии с формулой (1.33). Рассмотрим в качестве иллюстрации, причем на содержательном уровне, как заполняется клетка (4, 8). Данная клетка соответствует задаче $Z(4, 8)$. В сравнении с задачей $Z(3, 8)$, которой соответствует клетка (3, 8), в $Z(4, 8)$ появляется новая возможность – в ранец можно положить предмет $П4$.

Согласно формуле (1.33), выбирается лучший из двух вариантов.

Первый вариант предполагает, что возможностью положить в ранец предмет $П4$ мы не воспользовались. Тогда имеет место абсолютно та же ситуация, что в задаче $Z(3, 8)$, и мы можем обеспечить суммарную стоимость предметов в ранце равную $B^*(3, 8)$, т.е. числу 11, записанному в клетке, расположенной непосредственно над заполняемой. Итак, первый вариант обеспечивает суммарную стоимость предметов в ранце, равную 11; реализация данного варианта означает, что в ранец кладутся предметы $П1$, $П2$ и $П3$.

Второй вариант предусматривает, что предмет $П4$, его стоимость 7, кладется в ранец. Вес предмета $П4$ равен 4. Поэтому от заполняемой клетки в четвертой строке перемещаемся на 4 клетки влево, оказываемся в столбце, соответствующем оставшемуся резерву веса ранца после того, как в него положен предмет $П4$; непосредственно над полученной клеткой, а именно в клетке (3, 4) записано число 9, это максимально возможная суммарная стоимость предметов, взятых из совокупности $\{П1, П2, П3\}$, при условии, что их суммарный вес не превышает четырех (в этой ситуации как показывает заполнение клетки (3, 4), из указанной совокупности следует взять первый и третий предмет).

Таким образом, второй вариант обеспечивает суммарную стоимость предметов в ранце $7 + 9 = 16$; реализация данного варианта означает, что в ранец кладутся предметы $П1$, $П3$ и $П4$. Второй вариант дает лучший результат, заполнение клетки (4, 8) осуществляется по этому варианту.

Задача отыскания кратчайших путей (ЗОКП).

Считается заданным конечный взвешенный ориентированный граф G с множеством вершин $N = \{1, 2, \dots, n\}$, веса всех дуг неотрицательны. Веса дуг трактуем как их длины. Последовательность i_1, i_2, \dots, i_k вершин графа G определяет путь из вершины i_1 в вершину i_k , если для каждого $t = 1, 2, \dots, k-1$ в данном графе имеется дуга (i_t, i_{t+1}) ; указанные дуги образуют путь i_1, i_2, \dots, i_k . Сумма длин дуг, образующих путь, называется длиной этого пути. Для каждой вершины x графа требуется найти путь минимальной длины (кратчайший путь) из вершины 1 в вершину x . Вес дуги (i, j) графа G обозначаем $c(i, j)$. Длину кратчайшего пути из 1 в x будем называть расстоянием от вершины 1 до вершины x . Расстояние от вершины 1 до вершины x будем обозначать $s(x)$. Ясно, что $s(1) = 0$.

Пусть H – множество вершин, длины кратчайших путей из вершины 1 в которые уже известны. В начальной ситуации это множество одноэлементно: $H = \{1\}$. Через $s(1, M)$ обозначим минимальную из длин кратчайших путей от вершины 1 до вершин множества M , $M \subseteq \{2, 3, \dots, n\}$. Как очевидно, для множества вершин H , содержащего вершину 1, имеет место:

$$s(I, N \setminus H) = \min_{i \in H, j \in N \setminus H} (s(i) + c(i,j)). \quad (6.40)$$

Пусть минимум правой части (6.40) реализуется на паре (i^0, j^0) . Тогда кратчайший путь из вершины I в вершину j_0 получаем добавлением к кратчайшему пути из вершины I в вершину i^0 дуги (i^0, j^0) , длина $s(j^0)$ этого пути равна $s(i^0) + c(i^0, j^0)$.

Формула (6.40) – рекуррентное соотношение динамического программирования для решения ЗОКП. Пользуясь этой формулой, на первом шаге определяем ближайшую к вершине I вершину i_1 из совокупности $\{2, 3, \dots, n\}$. На втором шаге определяем ближайшую к вершине I вершину i_2 из совокупности $\{2, 3, \dots, n\} \setminus \{i_1\}$. На третьем шаге определяем ближайшую к вершине I вершину i^3 совокупности $\{2, 3, \dots, n\} \setminus \{i^1, i^2\}$, и т.д. В процессе счета строится дерево D кратчайших путей из вершины I в остальные вершины. Корнем D является вершина I ; если на произвольном шаге алгоритма минимум правой части (6.40) реализуется на паре (i^0, j^0) , к конструируемому дереву добавляется ребро (i^0, j^0) . Ход выполняемых вычислений можно оформить в виде процесса заполнения таблицы, строки которой соответствуют вершинам графа G (i -ая строка – вершине $i, i=1, n$). Таблица заполняется по столбцам, слева направо (каждый столбец для определенности заполняется сверху вниз). При этом первый (левый) столбец таблицы имеет имя " I ". При заполнении столбца " I " в клетку, стоящую в пересечении с j -ой строкой, $j \in \{2, 3, \dots, n\}$, вносится $c(I, j)$, если дуга (I, j) в графе G имеется, и $+\infty$, если такой дуги в графе нет. После заполнения этого столбца, в нем символом * отмечается наименьший элемент (в случае, когда таких элементов несколько, отмечается любой из них). Если отмеченный символом * элемент равен u и находится в строке номер v , то длина кратчайшего пути от вершины I до вершины v найдена, $s(v) = u$.

Первый шаг работы алгоритма реализован. Каждый следующий шаг алгоритма предусматривает заполнение очередных двух столбцов таблицы. Результатом каждого следующего шага является определение расстояния от вершины I до еще одной вершины. После того, как очередное расстояние $s(x)$ определилось, соответствующая строка x таблицы считается отмеченной, дальнейшее заполнение клеток отмеченных строк не производится. В момент, когда строка становится отмеченной, во всех незаполненных ее клетках ставим специальный символ z . Изначально отмеченной считается первая строка таблицы, ибо значение $s(I)$ известно с самого начала: $s(I) = 0$.

На втором шаге алгоритма второй столбец таблицы, а на последующих шагах – каждый следующий четный столбец, приобретает название последней вершины x графа G , для которой реализуемая вычислительная процедура в результате выполнения предшествующего шага определила расстояние от I до x .

В заголовках первого, второго и далее каждого следующего четного столбца рядом с его именем " x " указывается значение $s(x)$. При заполнении второго и каждого следующего четного столбца " x " в каждую клетку, находящуюся в пересечении с неотмеченной строкой j вносится сумма $s(x) + c(x, j)$; если в графе G дуга (x, j) отсутствует, в клетку вносится символ $+\infty$.

Третий и далее каждый следующий нечетный столбец имеет наименование " \min ". В каждую клетку такого столбца, стоящую в пересечении с j -ой неотмеченной строкой, вносится минимальное из расстояний, записанных в двух соседних слева клетках данной строки.

После заполнения каждого очередного столбца " \min ", в нем отмечается символом * наименьший элемент (если таких элементов несколько, отмечается любой из них). Если отмеченный символом * элемент последнего заполненного столбца " \min " равен u и

находится в строке номер v , то, согласно (6.40), считаем найденной длину кратчайшего пути от вершины l до вершины v , $s(v)=u$. С этого момента v -ая строка таблицы переходит в число отмеченных строк, дальнейшее заполнение ее клеток производиться не будет. Следующий подлежащий заполнению столбец четный, ему назначается имя " v ", в заголовке этого столбца указывается также найденное значение $s(v)$. Далее заполняется очередной столбец "min". В результате заполнения каждой следующей пары столбцов определяется расстояние от вершины l до еще одной вершины графа, таким образом завершается реализация следующего шага алгоритма. Изначально известно, что $s(l) = 0$. Поэтому общее число заполняемых пар столбцов $n-1$. Заполнение каждого очередного столбца требует выполнения линейно зависящего от n числа элементарных операций. Таким образом, реализация изложенного алгоритма требует выполнения квадратично зависящего от n числа элементарных операций.

ПРИМЕР 7. Взвешенный ориентированный граф G задан матрицей:

$$M = \begin{pmatrix} 0 & 9 & 7 & \infty & 2 & 1 & 9 \\ \infty & 0 & 4 & \infty & 3 & 8 & 5 \\ 6 & 3 & 0 & 1 & \infty & \infty & 6 \\ \infty & 1 & \infty & 0 & 4 & 5 & 6 \\ \infty & 9 & 1 & 3 & 0 & 7 & \infty \\ \infty & 5 & 9 & 9 & \infty & 0 & 2 \\ 9 & 8 & 7 & 6 & \infty & 5 & 0 \end{pmatrix}$$

Числовой элемент m_{ij} этой матрицы равен весу дуги (i,j) графа G ; если дуга в графе отсутствует, то $m_{ij} = \infty$. Веса дуг трактуются как их длины.

Требуется найти расстояния от вершины l до остальных вершин графа.

Ход выполняемых в процессе решения вычислений отображаем как последовательное заполнение столбцов таблицы 5.

Таблица 5.

Таблица определения расстояний

	"1" 0	"6" 1	min	"5" 2	min	"3" 3	Min	"7" 3	min	"4" 4	min
1	z	z	z	z	z	z	Z	z	z	z	z
2	9	9	9	11	9	6	6	11	6	5	5*
3	7	8	7	3	3*	z	Z	z	z	z	z
4	∞	7	7	5	5	4	4	9	4*	z	z
5	2	∞	2*	z	z	z	z	z	z	z	z
6	1*	z	z	z	z	z	z	z	z	z	z
7	9	3	3	∞	3	9	3*	z	z	z	z

В итоге получаем, что в рассматриваемом графе G расстояния от вершины l до вершин 2, 3, 4, 5, 6 и 7 равны соответственно 5, 3, 4, 2, 1 и 3. Отметим, что ранее для определяющего дискретную управляемую систему Ω графа $G(\Omega)$ был изложен другой, основанный на рекуррентных соотношениях (6.3)-(6.5) и использующий ацикличность графа, алгоритм поиска кратчайших путей.

Действительно, если в графе $G(\Omega)$ вес каждой дуги (x, y) трактуется как длина соответствующего одношагового перехода, то значение функции Беллмана $B^*(y)$ — длина кратчайшего пути из начальной вершины в вершину y .

Выше отмечалось, что метод динамического программирования относят к числу схем инвариантного погружения. Суть указанных схем в том, что каждая подлежащая решению задача заменяется упорядоченным семейством однотипных задач, последней из которых является решаемая. Далее в порядке возрастания значений одного или нескольких параметров последовательно решаются задачи введенного семейства; знание результатов решения предыдущих задач существенно облегчает решение следующих; в итоге решенной оказывается исходная задача. Удачная замена исходной задачи семейством порождаемых ею задач – главное условие эффективности такого подхода.

Классическая задача о назначениях (КЗН).

Имеются множество исполнителей $I = \{1, 2, \dots, n\}$, множество работ $R = \{r_1, r_2, \dots, r_n\}$ и $(n \times n)$ -матрица численных оценок (например, производительностей) $A = \{a_{ij}\}$, где a_{ij} – оценка закрепления исполнителя i за работой r_j , $i=1, n, j=1, n$.

Назначения – взаимно однозначные отображения множества $\{1, 2, \dots, n\}$ в себя; назначение π исполнителю i предписывает работу $r_{\pi(i)}$; численной оценкой такого закрепления является $a_{i\pi(i)}$ – элемент матрицы A , здесь $i=1, 2, \dots, n$. Каждое назначение π оцениваем по критерию:

$$K(\pi) = \sum_{i=1}^n a_{i\pi(i)}$$

в указанной интерпретации значение этого критерия – суммарная производительность исполнителей при назначении π .

Требуется найти назначение, при котором суммарная производительность исполнителей максимальна.

КЗН формулируется в виде:

$$\max_{\pi} \sum_{i=1}^n a_{i\pi(i)}$$

Рассмотрим вопрос о решении записанной КЗН, обозначим ее символом Z , методом динамического программирования. С этой целью введем в рассмотрение совокупность частных задач $Z(i, W_i)$, формулируемых по исходным данным задачи Z ; здесь $i \in \{1, 2, \dots, n\}$, а W_i – i -элементные подмножества совокупности $\{1, 2, \dots, n\}$.

В задаче $Z(i, W_i)$ между исполнителями $\{1, 2, \dots, i\}$ следует распределить совокупность работ, индексы которых перечислены в W_i ; критерий задачи прежний – суммарная производительность имеющихся исполнителей. Оптимальное значение критерия в задаче $Z(i, W_i)$ обозначим $B^*(i, W_i)$. В таком случае $B^*(n, \{1, 2, \dots, n\})$ – оптимальное значение критерия в исходной задаче Z .

Как очевидно:

$$B^*(1, \{j\}) = a_{1j} \text{ для всех } j \in \{1, 2, \dots, n\}. \quad (6.44)$$

Если $i > 1$, имеем:

$$B^*(i, W_i) = \max_{j \in W_i} (a_{ij} + B^*(i-1, W_i \setminus \{j\})), \quad i=2, 3, \dots, n. \quad (6.45)$$

здесь W_i – произвольное i -элементное подмножество из совокупности $\{1, 2, \dots, n\}$.

Записанные равенства (6.44)-(6.45) являются соотношениями динамического программирования для решения КЗН. Легко видеть, что оценкой числа элементарных операций, выполняемых основанным на соотношениях (6.44) - (6.45) алгоритмом, является функция, которая экспоненциально зависит от n . В отличие от всех остальных рассматриваемых в данном разделе задач, для КЗН имеются алгоритмы, обладающие

существенно лучшими в сравнении с процедурой счета по соотношениям динамического программирования, характеристиками быстродействия.

2. Задание на лабораторную работу

1) Имеются сферы капиталовложений S_1, S_2, S_3, S_4 ; в каждую из них может быть внесен целочисленный вклад, не превосходящий 5. Величина подлежащего распределению капитала равна 12. Функции, определяющие доходы, получаемые при вложении в имеющиеся сферы вклада u , заданы приведенной таблицей. Прямым методом Беллмана требуется найти оптимальное распределение капитала между сферами вложений.

Таблица значений функций дохода $f_i(u)$.

	$f_1(u)$	$f_2(u)$	$f_3(u)$	$f_4(u)$
$u = 0$	0	0	0	0
$u = 1$	0,1	0,1	0,2	0
$u = 2$	0,3	0,1	0,2	0
$u = 3$	0,3	0,4	0,3	0,3
$u = 4$	0,4	0,4	0,3	0,4
$u = 5$	0,5	0,4	0,3	0,4

2) Автомобиль, начиная от базы и заканчивая на базе замкнутый маршрут, должен доставить с базы в точки T_1, T_2, \dots, T_m некоторые грузы. Известно, что вес направляемого в точку T_i груза равен $c_i, i = 1, m$. Выполняя тот же маршрут, автомобиль должен собрать в точках R_1, R_2, \dots, R_n грузы, адресованные базе. Известно, что вес груза, направляемого из точки R_j на базу, равен $w_j, j = 1, n$. Через каждую из перечисленных точек маршрут должен проходить однократно. Матрица расстояний S размера $(m+n+1) \times (m+n+1)$ задана. Требуется записать рекуррентные соотношения динамического программирования для определения маршрута, минимального по количеству выполняемых тонно-километров.

3) Записать рекуррентные соотношения динамического программирования для классической задачи о ранце, дополненной условием: из каждой пары предметов $(P_1, P_n), (P_2, P_{n-1}), \dots, (P_k, P_{n-k+1})$ в ранец можно поместить только один предмет, здесь $k \leq n/2$.

4) Решить задачу о ранце:

$$7x_1 + 4x_2 + 4x_3 + 5x_4 + 2x_5 + 3x_6 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 + 2x_6 \leq 10;$$

$$x_i \in \{0, 1\}, i = \overline{1, 5}.$$

5) Решить следующую модификацию классической задачи о ранце:

$$5x_1 + 2x_2 + 4x_3 + 7x_4 + 5x_5 \rightarrow \max$$

при условиях:

$$2x_1 + x_2 + 2x_3 + 4x_4 + 3x_5 \leq 12;$$

$$x_i \in \{0, 1\}, i = 1, 2, 5;$$

$$x_i \in \{0, 1, 2\}, i = 3, 4.$$

6) Задача о путешественнике. На местности имеется сеть дорог, связывающих несколько населенных пунктов. Путешественник находится в пункте a_0 , из которого, двигаясь по одной из трех дорог, можно попасть в пункты a_1, a_2, a_3 . Из каждого пункта

опять выходят ровно три дороги, ведущие в a_4, a_5, a_6 . Из них – в a_7, a_8, a_9 и так далее, вплоть до конечных пунктов $b_1=a_3N-2, b_2=a_3N-1, b_3=a_3N$. Длины всех дорог заданы. Найти наиболее короткий путь из a_0 в один из конечных пунктов. Решить задачу при $N=5$. Оцените количество операций сложения и сравнения при ее решении по методу Беллмана, а также при полном переборе всех путей.

7) Задача о распределении инвестиций. Нужно распределить между N предприятиями сумму a , выделенную для их инвестирования. Известно, что вложение средств в размере y в k -ое предприятие обеспечивает прибыль в размере $d_k(y)$. Целью распределения является получение максимального суммарного дохода. Решить задачу при $N=4, a=300$ при условии, что суммы инвестиций всегда кратны 50, а функции $d_k(y)$ для $y=50j$ ($j=0, 1, \dots, 6$) принимают значения, заданные в таблице.

Таблица

		Значения функции $d_k(y)$					
y	0	50	100	150	200	250	300
$d_1(y)$	0	50	120	140	150	200	250
$d_2(y)$	0	60	130	140	130	160	200
$d_3(y)$	0	30	60	100	130	200	250
$d_4(y)$	0	40	100	110	120	160	220

8) Имеется множество недробимых камней $\{K_1, K_2, \dots, K_8\}$. Веса камней (в порядке возрастания индексов) следующие: 7, 5, 8, 4, 3, 6, 3, 1. Требуется разбить совокупность $\{K_1, K_2, \dots, K_8\}$ на два подмножества так, чтобы суммарный вес камней одного подмножества минимально отличался от суммарного веса камней другого подмножества.

9) Задача о замене оборудования. Частное предприятие планирует в течение N лет заниматься выпуском изделий, используя некоторое оборудование. В начале, можно либо купить новое оборудование возраста $x_0 = 0$ лет и стоимостью p , либо подержанное оборудование возраста $x_0 > 0$ лет по его ликвидной стоимости $\phi(x_0)$. Показатели эксплуатации оборудования включают: $f(t)$ – стоимость произведенных за год изделий на оборудовании возраста t лет; $r(t)$ – затраты на эксплуатацию в течение года оборудования возраста t лет.

В процессе эксплуатации оборудование можно менять, продавая старое по ликвидной стоимости $\phi(t)$ и покупая новое стоимостью p . В конце N -го года оборудование продается по ликвидной стоимости. Определить оптимальный возраст оборудования x_0 при начальной покупке и оптимальный график его замены. Выполнить расчеты при $N = 8, x_0 \in \{0, 1, 2\}$,

$$f(t) = 30 - t/2, r(t) = 13 + t/2, p = 17, \phi(t) = \begin{cases} 6 & \text{при } 0 \leq t \leq 6 \\ 2 & \text{при } 7 \leq t \leq 10 \end{cases}$$

10) Задача о загрузке судна. Судно, имеющее грузоподъемность a , загружается предметами N типов. Один предмет i -го типа имеет стоимость y_i и вес z_i . Требуется найти вариант загрузки судна, при котором стоимость взятых на борт предметов максимальна.

Решить задачу для $N = 3, a = 200, y_1 = 25, y_2 = 40, y_3 = 80, z_1 = 40, z_2 = 50, z_3 = 70$.

3. Вопросы к лабораторной работе

- 1) Каковы особенности применения генетических алгоритмов
- 2) Опишите операторы скрещивания в задаче коммивояжера.
- 3) Опишите операторы мутации в решении задачи коммивояжера
- 4) Раскрыть принцип динамического программирования.
- 5) Описать прямой метод Беллмана
- 6) Описать обратный метод Беллмана
- 7) Основные рекуррентные соотношения динамического программирования.

- 8) Каким образом будет выглядеть формула метода динамического программирования для решения задачи о ранце?
- 9) В чем заключается задача коммивояжера?
- 10) Приведите детерминированный алгоритм решения задачи о назначениях.
- 11) Что является в задаче о назначениях геном, генотипом и особью?
- 12) Как осуществляется генерация начальной популяции в задаче оптимального распределения капиталовложений?

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие / А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ
МЕТОДЫ МНОГОКРИТЕРИЛЬНОЙ ОПТИМИЗАЦИИ

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук,
доцент К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

МЕТОДЫ МНОГОКРИТЕРИЛЬНОЙ ОПТИМИЗАЦИИ: методические
указания для выполнения лабораторных работ по дисциплине «МЕТОДЫ
ОПТИМИЗАЦИИ В БИЗНЕС-АНАЛИТИКЕ» для студентов направления
09.04.03 «Прикладная информатика» направленность «Искусственный
интеллект и бизнес-аналитика» / Составитель: Т. В. Хоменко. – Ульяновск:
УлГТУ, 2021. – 30с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и
бизнес-аналитика». Цель данного практикума – ориентировать студентов
на содержание и порядок выполнения лабораторных работ во время
прохождения студентами курсов «Искусственный интеллект и бизнес-
аналитика». Даются задания для лабораторных занятий, а также
приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание

ВВЕДЕНИЕ	4
Лабораторная работа №5	5
Тема: Многокритериальная оптимизация. Парето-оптимальность.....	5
1. Краткие теоретические сведения.....	5
2. Задание на лабораторную работу	18
3. Вопросы к лабораторной работе.....	19
Лабораторная работа №6	21
Тема: Методы свертки критериев для решения задач многокритериальной оптимизации.....	21
1. Краткие теоретические сведения.....	21
2. Задание на лабораторную работу	27
3. Вопросы к лабораторной работе.....	29
Список литературы	30

ВВЕДЕНИЕ

В методических указаниях описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.).

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Лабораторная работа №5

Тема: Многокритериальная оптимизация. Парето-оптимальность.

Цель работы: знакомство с оптимизационными задачами, изучение различных методов векторной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Задачи многокритериальной оптимизации и методы их решения

Рассмотрим постановку задачи оптимизации с векторным критерием эффективности, т.е. задачу многокритериальной оптимизации.

Пусть система зависит от n варьируемых параметров (независимых переменных):

$$\bar{x} = [x_1, x_2, \dots, x_n]^T, \text{ где } \bar{x} - \text{точка } n\text{-мерного пространства.}$$

Предположим, что цель задачи состоит в максимизации (или минимизации) критериев: $f_1(\bar{x}), f_2(\bar{x}), \dots, f_r(\bar{x})$:

$$f_1(\bar{x}) \rightarrow \underset{(\min)}{\max}, f_2(\bar{x}) \rightarrow \underset{(\min)}{\max}, \dots, f_r(\bar{x}) \rightarrow \underset{(\min)}{\max}, \quad (7.1)$$

где $f_i(\bar{x}), i = \overline{1, r}$ – характеристики системы, связанные с оценками ее качества монотонной зависимостью.

Также задаются параметрические ограничения:

$$x_i^* \leq x_i \leq x_i^{**}, i = \overline{1, n} \quad (7.2)$$

и функциональные ограничения:

$$C_i^* \leq \Phi_i(\bar{x}) \leq C_i^{**}, i = \overline{1, k}. \quad (7.3)$$

здесь x_i^*, x_i^{**} – заранее известные границы изменения параметров x_i ; C_i^*, C_i^{**} – заданные константы.

Могут быть заданы также и критериальные ограничения:

$$f_i^* \leq f_i(\bar{x}) \leq f_i^{**}, i = \overline{1, r}; \quad (7.4)$$

где f_i^*, f_i^{**} – границы изменения критерия $f_i(x)$.

Ограничения (7.2) выделяют в n -мерном пространстве параллелепипед с

объемом $\prod_{i=1}^n (x_i^{**} - x_i^*)$. Ограничения (7.3) и (7.4) выделяют в параллелепипеде подмножество, которое в общем случае может быть не односвязным (рисунок 1).

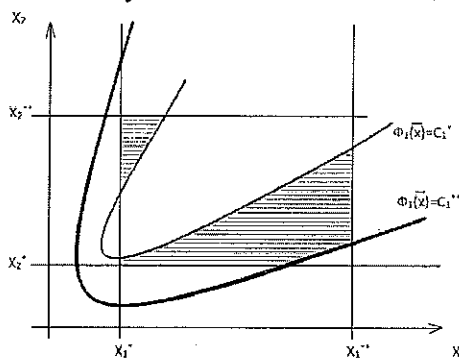


Рисунок 1 – Подмножество в параллелепипеде

Это множество $Gx \in Rn$ называют множеством альтернатив или допустимым множеством.

Задача оптимального проектирования сводится к решению многокритериальной задачи (7.1)–(7.4).

1.2. Основные подходы к решению задач многокритериальной оптимизации

Многокритериальные задачи очень разнообразны по содержанию, по объему и качеству информации. Возникает вопрос: а что же следует считать наилучшей альтернативой в задаче с несколькими целевыми функциями, которые противоречивы и достигают минимума (максимума) в различных точках множества альтернатив?

Основываясь на различных определениях наилучшей альтернативы, можно провести некоторую классификацию известных процедур решения многокритериальных задач.

Априорные процедуры многокритериальной оптимизации

Идея всех априорных процедур – это сведение задачи многокритериальной оптимизации к задаче однокритериальной оптимизации, потому что методы для решения задач однокритериальной оптимизации хорошо разработаны.

Традиционная схема рассуждений в значительно упрощенном виде выглядит следующим образом.

Каждый критерий $f_i(x)$, $i=1, 2, \dots, r$ характеризует некоторое локальное качество на множестве альтернатив Gx , например, вес, надежность, стоимость, быстродействие и т.д. Наилучшая альтернатива характеризуется наиболее удачным сочетанием всех этих локальных качеств, т.е. имеет максимальное значение «глобального» качества.

Таким образом, для выбора наилучшей альтернативы достаточно понять, каким образом глобальное качество зависит от локальных качеств, после чего многокритериальная задача будет сведена к задаче скалярной оптимизации. Понятно, что вид глобальной функции качества существенно зависит от типа решаемой задачи. Способы построения глобальной функции качества, т.е. преодоление неопределенности цели.

Пусть задача многокритериальной оптимизации имеет вид:

$$\begin{aligned} \max f_i(\bar{x}), i = \overline{1, r} \quad (r \geq 2) \\ \bar{x} \in G_x \end{aligned} \quad (7.5)$$

1) Выделение главного критерия (простейший способ преодоления неопределенности цели)

В задачах проектирования задается некоторая система нормативов $f_1^*, f_2^*, \dots, f_r^*$. Это значит, например, что параметры будущей конструкции должны быть таковы, чтобы максимизировать функции $f_i(\bar{x})$ при условиях:

$$f_i(\bar{x}) \geq f_i^*, \quad i = \overline{1, r}, \quad (7.6)$$

где f_i^* – некоторая система контрольных показателей.

Предположим, что, помимо того, среди критериев f_i мы выделили некоторый главный, например, $f_1(x)$. Тогда мы приходим к однокритериальной задаче:

Найти:

$$f_1(x) \rightarrow \max \quad (7.7)$$

при условии:

$$\bar{x} \in G_x, f_i(\bar{x}) \geq f_i^*, i = \overline{2, r}. \quad (7.8)$$

Данная схема сведения задачи (7.5) к однокритериальной задаче (7.7)–(7.8) является самой простой и наиболее употребительной в инженерной практике. Недостатками этого способа являются трудности выбора главного критерия и определения критериальных ограничений (назначения допустимых границ используемых показателей f_i^*). Также к недостаткам нужно отнести возможность варианта, что введенные ограничения окажутся несовместными.

Если задача (7.7)–(7.8) разрешима, то ее решение принимается в качестве решения задачи (7.5), либо на основе полученного решения (например, при помощи двойственных оценок ограничений) устанавливаются новые пределы f_i^* ($f_i \geq f_i^*$), и задача решается снова.

Если задача (7.7)–(7.8) решения не имеет, то приходится назначать новые пределы $f_i^* \leq f_i$ и задачу перерешать с f_i^* вместо f_i .

Задача проектировщика в данном подходе сводится только к назначению допустимых границ используемых показателей.

2) Линейная свертка

Вместо r частных критериев $f_i(\bar{x})$ предлагается рассматривать один критерий вида:

$$F(\bar{x}) = \sum_{i=1}^r c_i \cdot f_i(\bar{x}), \quad (7.9)$$

где $c_i \geq 0$ и $\sum_{i=1}^r c_i = 1$.

Такой способ свертки вводит, по существу, отношение эквивалентности различных критериев, так как величины c_i показывают, насколько изменяется целевая функция $F(\bar{x})$ при изменении критерия $f_i(\bar{x})$ на единицу.

Коэффициенты c_i – результаты экспертизы, они отражают представление лица, принимающего решение о содержании компромисса, который он вынужден принять.

Если же критерии $f_i(\bar{x})$ не выражаются в одних и тех же единицах измерения, то для использования формулы (7.9) их приводят к безразмерному виду. Известны разные способы. Например, путем деления каждого критерия на его максимальное значение на допустимом множестве G_x :

$$F(\bar{x}) = \sum_{i=1}^r c_i \frac{f_i(\bar{x})}{f_i^M},$$

где f_i^M – оптимальные значения функций $f_i(\bar{x})$ $i=1, 2, \dots, r$ на G_x .

Более сложный способ – введение функции:

$$F(\bar{x}) = \sum_{i=1}^r c_i \cdot \frac{f_i(\bar{x}) - f_i^m}{f_i^M - f_i^m},$$

где f_i^M – максимальное, а f_i^m – минимальное значения критерия $f_i(\bar{x})$ на G_x .

При этом вклад «локального» качества $f_i(\bar{x})$ в «глобальное» качество $F(\bar{x})$ зависит от того, насколько сильно меняется локальное качество на допустимом множестве G_x .

3) Принцип справедливого компромисса

Основным недостатком предыдущего способа является возможность

компенсации недопустимо малых значений некоторых критериев достаточно большими значениями других. Действительно, если x характеризует проект самолета, а критерии f_1 и f_2 представляют собой его надежность и скорость, соответственно, то очень высокой надежностью может обладать самолет, который никогда не сможет взлететь, и, следовательно, будет иметь нулевую скорость. Несмотря на явную бесполезность такого устройства, значение функции:

$$F(\bar{x}) = c_1 f_1(\bar{x}) + c_2 f_2(\bar{x})$$

может быть довольно большим. В таких случаях необходимо использовать другие принципы. Например,

$$F(\bar{x}) = \prod_{i=1}^r f_i(\bar{x}),$$

Этот принцип получил название принципа справедливого компромисса.

4) Принцип гарантированного результата

Предположим, что мы ввели некоторую систему контрольных показателей f_i^* , относительно которых критерии $f_i(\bar{x})$ должны удовлетворять условиям:

$$f_i(\bar{x}) \geq f_i^*, \quad i = 1, 2, \dots, r.$$

В таких случаях целевую функцию удобно представлять в виде:

$$F(\bar{x}) = \min_{1 \leq i \leq r} \frac{f_i(\bar{x})}{f_i^*}$$

и искать вектор \bar{x} , который обеспечивает максимальное значение $F(\bar{x})$.

Смысл здесь достаточно прост. При данном значении вектора \bar{x} : величина $F(\bar{x})$ дает нам значение наихудшего из показателей $f_i(\bar{x})$. Максимизация $F(\bar{x})$ обеспечивает наибольшее значение для наихудшего из показателей $f_i(\bar{x})$, т.е. гарантированный результат.

5) Метод идеальной точки в пространстве критериев

Предположим, что мы решили систему однокритериальных задач:

$$f_i(\bar{x}) \rightarrow \max, \quad i = 1, 2, \dots, r,$$

и нашли идеальную точку: $\bar{\xi} = (\xi_1, \dots, \xi_r)$, где:

$$\xi_i = \max_{x \in X} f_i(\bar{x}).$$

Точка $\bar{\xi}$ является недостижимой в пространстве критериев. Введем положительно определенную матрицу $R = \{r_{ij}\}$. Тогда скалярная величина:

$$H = \sqrt{\sum_{i,j} (f_i(\bar{x}) - \xi_i) r_{ij} (f_j(\bar{x}) - \xi_j)} \quad (7.10)$$

определяет некоторое расстояние от точки, соответствующей данному вектору x , до идеальной точки. В частном случае, когда R – единичная матрица:

$$H = \sqrt{\sum_i (f_i(\bar{x}) - \xi_i)^2}$$

есть евклидово расстояние до точки $\bar{\xi}$

В качестве нового скалярного критерия мы можем принять функцию (7.10). Её минимизация дает определенную полезную информацию: показывает наши предельные возможности достижения идеальной точки.

б) Метод ранжирования критериев

Предположим, что локальные критерии упорядочены по важности:

$$f_1(\bar{x}) > f_2(\bar{x}) > \dots > f_r(\bar{x}),$$

здесь $(f_i(\bar{x}) > f_{i+1}(\bar{x}))$ означает, что $f_i(\bar{x})$ важнее $f_{i+1}(\bar{x})$.

Тогда решение задачи сводится к тому, чтобы:

найти:

$$\max_{\bar{x} \in G_x} f_1(\bar{x}) = f_1^*;$$

найти:

$$\max_{\bar{x} \in G_x} f_2(\bar{x}) = f_2^* \quad \text{при} \quad f_1(\bar{x}) = f_1^*;$$

...

найти:

$$\max_{\bar{x} \in G_x} f_n(\bar{x}) = f_n^* \quad \text{при} \quad f_i(\bar{x}) = f_i^*, \quad i = 1, 2, \dots, r-1$$

т.е. к многоэтапному алгоритму с оптимизацией каждого критерия при фиксированных предыдущих.

На практике для получения приемлемых решений по менее важным критериям приходится делать уступки по другим, наиболее важным. Данный подход реализуется в алгоритме последовательных уступок, который сводится к последовательному решению задач однокритериальной нелинейной оптимизации:

найти:

$$f_1^* = \max_{\bar{x} \in G_x} f_1(\bar{x}),$$

найти:

$$f_2^* = \max_{\bar{x} \in G_x} f_2(\bar{x}) \quad \text{при} \quad f_1(\bar{x}) \geq f_1^* - \Delta_1$$

и т.д.;

найти:

$$f_r^* = \max_{\bar{x} \in G_x} f_r(\bar{x}) \quad \text{при} \quad f_i(\bar{x}) \geq f_i^* - \Delta_i, \quad i = 1, 2, \dots, r-1$$

где Δ_i – величина уступки по i -му критерию.

Другими словами, решается скалярная задача оптимизации по критерию $f_i(\bar{x})$. Если $\max_{\bar{x} \in G_x} f_i(\bar{x}) = f_i^*$, то по критерию $f_i(\bar{x})$ устанавливается уступка Δ_i , на которую согласно ЛПР, чтобы упростить решение по следующим критериям (в первую очередь, по $f_2(\bar{x})$), и т. д.

Следовательно, процедуры рассмотренного типа основаны на предположении возможности априорного (до решения задачи) определения наилучшего соотношения между требованиями, предъявляемыми различными критериями.

Построение множества оптимальных по Парето альтернатив

Наличие нескольких критериев задает на множестве альтернатив частичное упорядочение. Подозрительными на наилучшее соответствие цели проектирования являются максимальные по этому упорядочению альтернативы.

Предположим, что мы должны сделать выбор на допустимом множестве

альтернатив G_x , так, чтобы:

$$f_1(\bar{x}) \rightarrow \max, \dots, f_r(\bar{x}) \rightarrow \max$$

Допустим, мы сделали некоторый выбор. Обозначим его \bar{x} и предположим, что существует некоторый другой выбор \bar{x}^* , такой, что для всех критериев $f_i(\bar{x})$ имеет место неравенство:

$$f_i(\bar{x}^*) \geq f_i(\bar{x}), \quad i = 1, 2, \dots, r, \quad (7.11)$$

причем, хотя бы одно неравенство – строгое.

Пусть решается задача:

$$f_1(\bar{x}) \rightarrow \max, f_2(\bar{x}) \rightarrow \max, \quad \bar{x} \in G_x.$$

На рисунке 2 показано, что каждой точке $\bar{x} \in G_x$ соотношения $f_1 = f_1(\bar{x}), f_2 = f_2(\bar{x})$ $T = (f_1, f_2) \in E_B$ плоскости критериев и определяют отображение G_x на G_F .

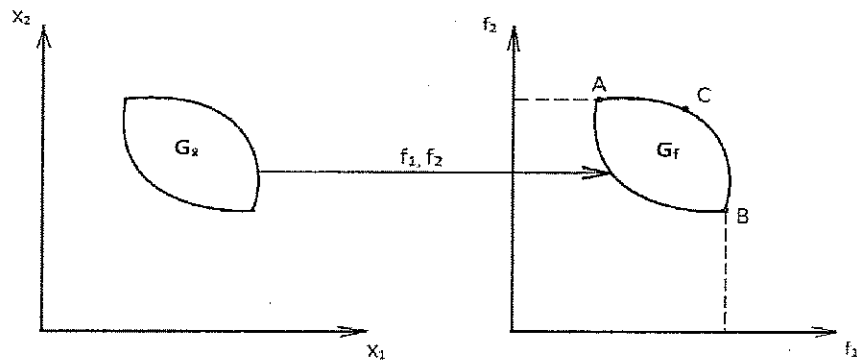


Рисунок 2 – Отображение G_x на G_F

Множество G_F носит название множества достижимости или множества предельных возможностей, или пространства критериев. Изучение структуры этого множества может оказаться весьма полезным при исследовании различных задач проектирования. Заметим, что образ множества Парето в множестве G_F представляет собой лишь часть множества G_F (на рисунке 2 – дуга ACB). Приближенное построение образа множества Парето сводится к последовательному решению ряда задач математического программирования.

Опишем две из возможных схем расчета.

1) Пусть f_1 и f_2 ; $f_1=C_1, f_2=C_2$ – некоторое желательное значение критериев. Значения C_1 и C_2 следует выбрать так, чтобы они принадлежали множеству достижимости. Далее решаются две оптимизационные задачи:

$$\begin{array}{ll} \text{I. } f_1(\bar{x}) \rightarrow \max & \text{II. } f_2(\bar{x}) \rightarrow \max \\ \bar{x} \in G_x, f_2(\bar{x}) = C_2 & \bar{x} \in G_x, f_1(\bar{x}) = C_1. \end{array}$$

Решив эти задачи, мы определим точки A и B в множестве достижимости. Проведя через них прямую, получим простейшую аппроксимацию образа множества Парето в множестве G_F .

Для уточнения аппроксимации решаются следующие задачи:

$$\begin{array}{ll} \text{III. } f_1(\bar{x}) \rightarrow \max & \text{IV. } f_2(\bar{x}) \rightarrow \max \\ \bar{x} \in G_x, f_2(\bar{x}) = C_4 & \bar{x} \in G_x, f_1(\bar{x}) = C_3 \end{array}$$

Находим еще две точки – C и D . Через точки A, C, D, B (рисунок 3).

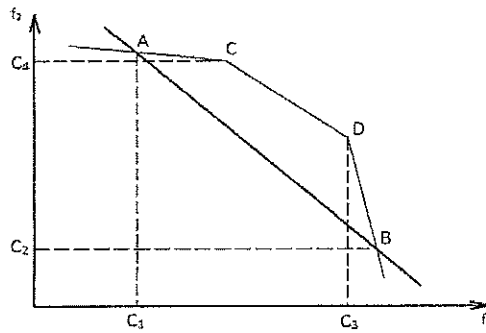


Рисунок 3 – Приближение образа множества Парето

Проведем ломаную, которая и будет следующим приближением образа множества Парето. Данная процедура может быть продолжена.

2) Пусть $\lambda_i, i=1, 2, \dots, r$ – неотрицательные числа, такие, что:

$$\sum_{i=1}^r \lambda_i = 1$$

Составим новый критерий:

$$F(\bar{x}) = \sum_{i=1}^r \lambda_i \cdot f_i(\bar{x})$$

и решим следующую задачу:

$$F(\bar{x}) \rightarrow \max, \quad \bar{x} \in G_x.$$

Оказывается, что решение этой задачи определяет такой вектор \bar{x}^* , что точка с координатами $(f_1(\bar{x}^*), f_2(\bar{x}^*), \dots, f_r(\bar{x}^*))$ принадлежит образу множества Парето. Поэтому, варьируя значения коэффициентов λ_i можно получить любое количество точек образа множества Парето. Следовательно, как и в предыдущем случае, можно построить ломаную, являющуюся приближением образа множества Парето в пространстве F .

Соответствующие описанные способы можно распространить и на случай большего числа критериев.

Окончательный выбор оптимального решения из множества Парето остается за ЛПР.

Процедуры апостериорного типа

В основе апостериорных процедур лежит предположение, что формальная модель многокритериальной задачи не содержит информации, достаточной для однозначного выбора наилучшей альтернативы. Отметим, что решения, принимаемые с помощью апостериорных процедур, имеют принципиально субъективный характер. Учет предпочтений ЛПР в этом случае является одним из наиболее эффективных методов снятия имеющейся неопределенности.

Адаптивные процедуры

Существенно уменьшить, а в ряде случаев и упростить требуемую от ЛПР информацию можно, если попытаться определить наилучшую или наиболее предпочтительную альтернативу, не восстанавливая функцию полезности. Процедуры, реализующие эту идею, используют подход, основанный на выявлении предпочтений одновременно с исследованием допустимого множества альтернатив.

Одна из идей, используемая здесь, сводится к тому, чтобы выбрать любой известный алгоритм математического программирования, который хорошо зарекомендовал себя при решении скалярных оптимизационных задач, и распределить функции между человеком и ЭВМ таким образом, чтобы в результате их

взаимодействия обеспечивалось получение всей необходимой для работы алгоритма информации.

Метод Соболя-Статникова

Область альтернатив (заданных ограничениями (7.2), (7.3), (7.4)) покрывается равномерно распределенными точками. В каждой точке вычисляются значения целевых функций, и результаты систематизируются в виде таблиц испытаний, в которых критерии упорядочены. Решение задачи сводится к анализу таблиц испытаний. Проектировщик (ЛПР) просматривает таблицы испытаний каждого критерия, обращая внимание на следующие факты:

- в каких пределах заключены показатели $f_j(\bar{x})$ для каждого j ;
- как много точек $f_j(\bar{x})$ отвечают «хорошим» проектным решениям;
- каков «средний» уровень показателей в данной таблице и как сильно разбросаны вокруг него остальные решения.

В зависимости от ситуации проектировщик назначает $2r$ чисел f_i^* и f_i^{**} , $i=1, r$, являющихся, соответственно, нижними и верхними границами, фигурирующими в тех неравенствах (7.4), которые образуют условия приемлемости проектного решения. После этого в автоматическом режиме проверяется совместимость критериальных ограничений, т.е. определяет множество решений, одновременно удовлетворяющих ограничениям:

$$f_i^* \leq f_i(\bar{x}) \leq f_i^{**}.$$

Если такие точки существуют, то это говорит о том, что задача имеет решение. Если нет, то можно ослабить критериальные ограничения или же увеличить число точек в области поиска и после этого еще раз применить диалоговую процедуру определения границ. После конечного числа шагов получится допустимое множество решений D , удовлетворяющее ограничениям (7.2), (7.3), (7.4).

Во многих случаях проектировщик (ЛПР) предпочитает сам назначить «наилучшее» решение путем тщательного изучения полученного допустимого множества решений. В этом случае есть возможность облегчить еще работу по изучению точек допустимого множества. Можно исключить заведомо «плохие» решения, т.е. выделить множество Парето G в множестве D .

Таким образом, при поиске «наилучшего» решения ЛПР может ограничиться изучением множества G . Более того, множество G может быть использовано при назначении критериальных ограничений на каждом шаге диалоговой процедуры. После конечного числа шагов итерационной процедуры количество точек в множестве G можно сократить до минимума и тем самым облегчить принятие окончательного решения.

Другой способ использования множества D – это попытка сформулировать решающее правило, т.е. критерий, и искать его оптимум в допустимой области. При этом решения, полученные в результате осуществления описанной диалоговой процедуры, можно использовать в качестве начального приближения.

Данный подход можно успешно использовать тогда, когда отсутствуют явные выражения целевых функций и ограничений задачи.

ПРИМЕР 1. Инвестор рассматривает 10 вариантов бизнес-проекта. Для него важны 3 критерия. Самым важным для инвестора является годовая прибыль проекта (желательно, чтобы она была максимальной). Кроме того, он хочет, чтобы необходимый объем первоначальных инвестиций был как можно меньше. Третьим по

степени важности является желание создать социально-ориентированное предприятие, что будет характеризоваться высоким общим объемом годовой заработной платы. Варианты проектов с соответствующими показателями приведены в таблице 1.

Таблица 1

Варианты проектов с соответствующими показателями

Вариант проекта	Годовая прибыль, млн. руб.	Начальные инвестиции, млн. руб.	Общая годовая зарплата, млн. руб.
1	1000	2200	500
2	1300	1800	300
3	700	1800	900
4	1600	2500	500
5	900	1600	1000
6	800	1700	1000
7	1000	2000	600
8	1100	1800	250
9	1600	2500	900
10	1300	3000	700

Задание:

1. Определите Парето-оптимальные варианты проектов. То есть проекты, для которых нельзя найти альтернативные варианты не худшие по всем критериям.

2. Из Парето-оптимальных проектов выберите лучший:

- методом пропорции,
- методом идеальной точки.

Как было сказано выше, главным для инвестора является прибыль. Но инвестор не хочет вкладывать более 2300 млн. руб. и настаивает, чтобы совокупная заработная плата на предприятии была не ниже 500 млн. руб.

- методом последовательных уступок.

После определения оптимальной прибыли, допускается снижение прибыли не более чем на 600 млн.руб. И после определения минимальных инвестиций для таких условий допускается их увеличение не более, чем на 15%.

3. Выбрать графически Парето-эффективные варианты в предположении, что для инвестора актуальны только два критерия:

- Прибыль и Инвестиции;
- Прибыль и Зарплата;
- Инвестиции и Зарплата.

4. Для каждой пары критериев, рассмотренных выше, убедиться, что графически отброшенные варианты действительно могут быть улучшены одним или несколькими из оставленных Парето-эффективных.

5. В одном из случаев останется единственный Парето-оптимальный вариант решения. Какой вывод из этого следует?

Шаг 1. Определим Парето-оптимальные решения.

Для этого будем сравнивать попарно варианты проекта всем критериям.

Следует учесть, что по прибыли и зарплате чем больше, тем лучше; по инвестициям чем меньше, тем лучше.

Сравним 1 и 2 варианты:

- по прибыли: 1 вариант хуже чем 2 ($1000 < 1300$);
- по инвестициям: 1 вариант хуже чем 2 ($2200 > 1800$);
- по зарплате: 1 вариант лучше чем 2 ($500 > 300$).

Так как есть разные варианты сравнения, то на основе данного сравнения исключить 1 или 2 альтернативу нельзя.

Сравним 1 и 3 варианты:

- по прибыли: 1 вариант лучше чем 3 ($1000 > 700$);
- по инвестициям: 1 вариант хуже чем 3 ($2200 > 1800$).

Так как есть разные варианты сравнения, то на основе данного сравнения исключить 1 или 3 альтернативу нельзя.

Сравним 1 и 4 варианты:

- по прибыли: 1 вариант хуже чем 4 ($1000 < 1600$);
- по инвестициям: 1 вариант лучше чем 4 ($2200 > 2500$).

Так как есть разные варианты сравнения, то на основе данного сравнения исключить 1 или 4 альтернативу нельзя.

Сравним 1 и 5 варианты:

- по прибыли: 1 вариант лучше чем 5 ($1000 > 900$);
- по инвестициям: 1 вариант хуже чем 5 ($2200 > 1600$).

Так как есть разные варианты сравнения, то исключить на основе данного сравнения какую-либо из альтернатив 1 или 5 нельзя.

Сравним 1 и 6 варианты:

- по прибыли: 1 вариант лучше чем 6 ($1000 > 800$);
- по инвестициям: 1 вариант хуже чем 6 ($2200 > 1700$).

Так как есть разные варианты сравнения, то исключить на основе данного сравнения какую-либо из альтернатив 1 или 6 нельзя.

Сравним 1 и 7 варианты:

- по прибыли: 1 вариант равен 7 ($1000 = 1000$);
- по инвестициям: 1 вариант хуже чем 7 ($2200 > 2000$);
- по зарплате: 1 вариант хуже чем 7 ($500 < 600$).

По всем вариантам сравнения 7 проект лучше или не хуже чем 1-й. Значит 1-й проект можно исключить из рассмотрения.

Дальше сравниваем с остальными 2-й проект.

Можно показать (проверьте самостоятельно), что 2-й проект не может быть сравнен с проектами 3, 4, 5, 6, 7, 9, 10.

При сравнении 2 и 8 проектов видно, что:

- по прибыли: 2 вариант лучше 8 ($1300 > 1100$);
- по инвестициям: 2 вариант одинаков с 8 ($1800 = 1800$);
- по зарплате: 2 вариант лучше 8 ($300 > 250$).

Значит, на основе данного сравнения можно исключить из рассмотрения 8-й вариант проекта.

Так как после всех сравнений 2-й вариант проекта не исключается, то он будет парето-оптимальным. Путем дальнейших сравнений исключаются все варианты кроме 2, 5, 7, 9 (определите, относительно каких проектов исключаются остальные).

Таким образом, Парето-оптимальными (или Парето-эффективными) стратегиями являются стратегии, представленные в таблице 2.

Таблица 2

Парето-оптимальные стратегии

Вариант проекта	Годовая прибыль, млн. руб.	Начальные инвестиции, млн. руб.	Общая годовая зарплата, млн. руб.
2	1300	1800	300
5	900	1600	1000

7	1000	2000	600
9	1600	2500	900

Применим все шесть критериев для нашего случая:

$$F \rightarrow \max; g \rightarrow \min; Z \rightarrow \max.$$

Тут через F обозначена прибыль, через g – инвестиции, через Z – зарплата:

Веса определяем из условия 2.2 задания: $vF=10$; $vg = 7$; $vZ = 3$.

Из условия 2.5 определяем ограничения, которые будут использоваться в методе главного критерия: $g_{\max}=2300$; $Z_{\min}=500$.

Из условия 2.6 определяем уступки, которые будут использованы в методе последовательных уступок: для прибыли задана абсолютная уступка $= \Delta F 600$; для инвестиций задана относительная уступка $\delta g=15\%=0,15$.

Шаг 2. Выбор лучшего Из Парето-оптимальных проектов

– Метод пропорции (отношения, пропорциональности)

Умножаем прибыль на зарплату и делим на инвестиции:

$$P_2 = \frac{1300 \cdot 300}{1800} = 216,66$$

$$P_3 = \frac{900 \cdot 1000}{1600} = 562,5$$

$$P_7 = \frac{1000 \cdot 600}{2000} = 300$$

$$P_9 = \frac{1600 \cdot 900}{2500} = 576$$

Очевидно, наибольшее значение достигается в проекте номер 9.

Заметим, что в этом примере метод пропорции представляется нелогичным. Произведение прибыли на зарплату не имеет никакого смысла. Более логичным был бы специфический для данного примера критерий отношения суммы прибыли и зарплаты к инвестициям.

– Метод идеальной точки (минимального отличия от идеала)

Вначале определи идеалы критериев из таблицы.

Наибольшая прибыль достигается в 9 проекте и равна 1600 млн.руб. Значит $F^*=1600$.

Наименьшие инвестиции требуются в 5 проекте и равны 1600 млн.руб. Значит $g^*=1600$.

Наибольшая зарплата в 5 проекте, она равна 1000 млн.руб. Значит $Z^*=1000$. Для каждого проекта составляем сумму квадратов разностей показателей с их идеалами:

$$I_2 = (1300-1600)^2 + (1800-1600)^2 + (300-1000)^2 = 620000$$

$$I_3 = (900-1600)^2 + (1600-1600)^2 + (1000-1000)^2 = 490000$$

$$I_7 = (1000-1600)^2 + (2000-1600)^2 + (600-1000)^2 = 680000$$

$$I_9 = (1600-1600)^2 + (2500-1600)^2 + (900-1000)^2 = 820000$$

Очевидно, наименьшее значение достигается в проекте номер 5. Именно его и выбираем по методу идеальной точки.

По методу идеальной точки выбираем проект 2.

– Метод последовательных уступок

1) Определяем оптимальное значение главного критерия – прибыли.

Максимальная прибыль достигается в 9 проекте и равна 1600 млн.руб. То есть, $F^*=1600$.

2) Сделаем первую уступку. Будем искать наименьшие инвестиции из всех проектов, удовлетворяющих условию уступки:

$$g \rightarrow \min$$

$$F \geq 1600 - 1000 = 600$$

То есть нас интересуют все проекты, в которых прибыль не ниже 1000 млн.руб. Такими проектами являются 2, 7 и 9 проекты. Наименьшие инвестиции в них равны 1800 млн.руб. (во 2 проекте). (Заметим, что данный условный идеал инвестиций отличается от абсолютного идеала 1600 млн.руб., отброшенного из-за слишком малой прибыли соответствующего проекта).

3) Сделаем вторую уступку. Будем искать максимум зарплаты при двух ограничениях:

$$Z \rightarrow \max$$

$$\begin{cases} F \geq 1600 - 600 = 1000 \\ g \leq 1800 \cdot (1 + 0,15) = 2070 \end{cases}$$

То есть нам надо найти проект с максимальной зарплатой из всех проектов, при которых прибыль не ниже 1000 млн.руб. и инвестиции не выше 2070 млн.руб.

Этим условиям удовлетворяют проекты 2 и 7. Наибольшая совокупная зарплата из этих двух проектов реализуется в 7 проекте и равна 600 млн.руб.

Значит, этот проект и будет выбран по методу последовательных уступок.

По методу последовательных уступок выбираем проект 7.

Замечание. В этой задаче ранее использованное ограничение можно было не записывать, так как неудовлетворяющий ему проект уже был отброшен. В ряде задач (особенно для непрерывных функций) необходимо всегда записывать все ограничения.

Шаг 3. Выбор Парето-оптимальные критерии графически в случае двух критериев

— важными являются только два критерия «Прибыль и Инвестиции»

Таблица 3

Вариант проекта	Годовая прибыль, млн. руб.	Начальные инвестиции, млн. руб.	
1	1000	2200	
2	1300	1800	
3	700	1800	
4	1600	2500	
5	900	1600	
6	800	1700	
7	1000	2000	
8	1100	1800	
9	1600	2500	
10	1300	3000	

На рисунке 2 представлена точечная диаграмма по данным таблицы 3.

По оси X отложены прибыли, они должны быть максимальными. По оси Y отложены инвестиции, они должны быть минимальными. То есть, точки тем лучше, чем они правее или ниже. Очевидно, «неулучшаемых», Парето-оптимальных точек в этом случае будет три. Они выделены на диаграмме большими ромбами.

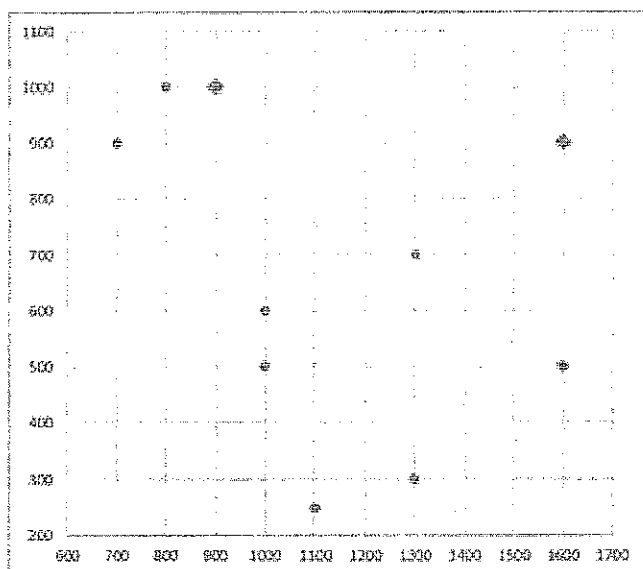


Рисунок 1 – Точечная диаграмма данных таблицы 3

Этим точкам соответствуют проекты 2, 5 и 9. Они и будут Парето-эффективными проектами.

Покажем, что оставшиеся проекты могут быть улучшены хотя бы одним из выбранных.

Рассмотрим проект 1. В нем прибыль меньше, чем в проекте 2, а инвестиции больше, чем в проекте 2. То есть с точки зрения обоих критериев лучше выбрать проект 2, чем проект 1.

Парето-оптимальными проектами для критериев «Прибыль и Инвестиции» являются проекты 2, 5, 9.

– важными являются только два критерия «Прибыль и Зарплата»

Таблица 4

Вариант проекта	Годовая прибыль, млн. руб.	Общая годовая зарплата, млн. руб.
1	1000	500
2	1300	300
3	700	900
4	1600	500
5	900	1000
6	800	1000
7	1000	600
8	1100	250
9	1600	900
10	1300	700

На рисунке 1 представлена точечная диаграмма по данным таблицы 4.

По оси X отложены прибыли, они должны быть максимальными. По оси Y отложены зарплаты, они тоже должны быть максимальными. То есть, точки тем лучше, чем они правее или выше. Очевидно, «неулучшаемых», Парето-оптимальных точек в этом случае будет две. Они выделены на диаграмме большими ромбами.

Этим точкам соответствуют проекты 5 и 9. Они и будут Парето-эффективными проектами.

Парето-оптимальными проектами для критериев «Прибыль и Зарплата» являются проекты 5 и 9.

– важными являются только два критерия «Инвестиции и Зарплата»

Таблица 5

Вариант проекта	Начальные инвестиции, млн. руб.	Общая годовая зарплата, млн. руб.
1	2200	500
2	1800	300
3	1800	900
4	2500	500
5	1600	1000
6	1700	1000
7	2000	600
8	1800	250
9	2500	900
10	3000	700

На рисунке 2 представлена точечная диаграмма по данным таблицы 5.

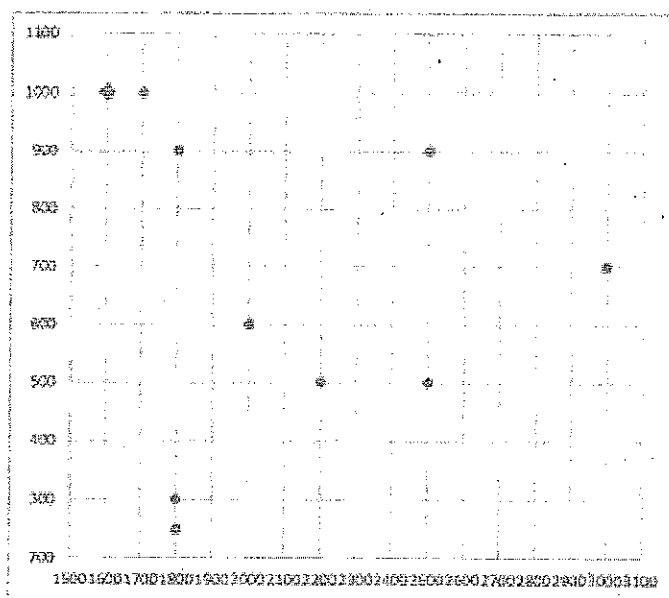


Рисунок 2 – Точечная диаграмма данных таблицы 5

По оси X отложены инвестиции, они должны быть минимальными. По оси Y отложены зарплаты, они должны быть максимальными. То есть, точки тем лучше, чем они левее или выше. Очевидно, в этом случае единственное «неулучшаемое» решение – Парето-оптимальная точка одна. Она выделена на диаграмме большим ромбом.

Этой точке соответствует проект 5. Это тот несчастный случай, когда оба критерия приводят к единому лучшему решению и именно его и надо выбирать.

Парето-оптимальным проектом для критериев «Инвестиции и Зарплата» оказался единственный проект 5. Значит, он будет однозначно лучшим для случая этих критериев.

Как видно из решений подпунктов пункта 3, при оставлении частных случаев от исходного (оставлении только двух критериев из трех), Парето-оптимальные множества решения оказываются подмножествами от исходного Парето-множества. В случае, когда Парето-оптимальным является единственный вариант решения, то он и должен быть выбран в любом случае.

2. Задание на лабораторную работу

- 2.1. Изучить предлагаемые методы
- 2.2. Решить задачи

№1. Планируется покупка автомобиля для коммерческих нужд. Цена и годовая прибыль от использования автомобиля в зависимости от его марки записаны в таблице ниже.

Марка	Цена	Прибыль
Ауди	1200	450
Матиз	350	150
Форд	750	350
Ниссан	1000	300
Хонда	1000	450
ВАЗ	480	150
Мазда	1000	400
Хундай	500	200
Шевроле	800	250
Рено	750	400
Опель	550	300

Определить Парето-оптимальные варианты автомобилей. Покажите, как этот выбор выглядит на графике.

Из Парето-оптимальных решений выбрать лучшее по разным методикам:

- определить наилучший вариант проекта по методу пропорции,
- определить наилучший вариант проекта по методу минимального отличия от идеала (идеальной точки).

№2. Решить задачу многокритериальной оптимизации, сформировав обобщенный критерий по методам:

- каждого критерия в отдельности;
- пропорциональности;
- минимального отличия от идеала;

$$F_1 = 5x_1 + 3x_2 + 7x_3 \rightarrow \max$$

$$F_2 = 30x_1 + 75x_2 + 40x_3 \rightarrow \min$$

$$F_3 = 0,2x_1 + 0,5x_2 + 0,4x_3 \rightarrow \max$$

$$\begin{cases} x_1 \geq 10; x_2 \geq 20; x_3 \geq 30 \\ 3x_1 + 7x_2 + 5x_3 \leq 1000 \end{cases}$$

Критерии расположены в порядке снижения значимости. Вес первого критерия 10, второго 5, третьего 3. Максимально допустимое значение второго критерия равно 6000, минимальное значение третьего равно 60.

2.3. Оформить отчет о выполнении задания с приведением условия задачи, алгоритмов и/или программ, указанных в задании методов поиска, результатов задачи.

Представить результаты исследований в таблицах и графиками. Сделать выводы по проведенным исследованиям. В отчете должны быть отражены все этапы выполнения работы и представлены выводы и анализ результатов. Расчеты рекомендуется проводить при помощи табличного процессора Excel, с приложением листингов в отчете по лабораторной работе.

3. Вопросы к лабораторной работе

- 1) Математическая постановка задач многокритериальной оптимизации
- 2) Методы решения задач многокритериальной оптимизации
- 3) Оптимальное проектирование и задачи многокритериальной оптимизации

- 4) Какое множество называется множеством Парето?
- 5) Что означает «оптимальность по Парето»?
- 6) Какие процедуры многокритериальной оптимизации называются априорными?
- 7) Какие процедуры многокритериальной оптимизации называются апостериорными?
- 8) Что означает адаптивный подход решения задач многокритериальной оптимизации?

Лабораторная работа №6

Тема: Методы сверток критериев для решения задач многокритериальной оптимизации

Цель работы: знакомство с оптимизационными задачами, изучение различных методов многокритериальной оптимизации и сравнение эффективности их применения.

1. Краткие теоретические сведения

Многокритериальная задача – задача, где имеется не один, а несколько критериев оценки качества решения. Например, задача распределения инвестиций по этапам выполнения проекта: критерий (1) «минимальные расходы на выполнение проекта» обычно дополняется критериями (2) «минимальный срок выполнения проекта» и (3) «максимум качества проектных решений».

Главная проблема постановки и решения многокритериальных задач – согласование критериев, т.е. выбор наилучшего соотношения между оценками по разным критериям. Ответ на этот вопрос обычно не определяется условиями задачи. Нужна дополнительная информация, которая может быть получена только от руководства проектом.

1.1. Перевод критериев в ограничения

Одним из первых подходов к принятию решений при двух критериях является метод «стоимость-эффективность». Метод «стоимость-эффективность» состоит из трех основных этапов:

- построения модели эффективности: $z_E(x_1, \dots, x_n) \rightarrow \max$;
- построения модели стоимости: $z_C(x_1, \dots, x_n) \rightarrow \min$;
- синтеза оценок стоимости и эффективности.

Модель стоимости представляет зависимость общей стоимости от количества ракет, а модель эффективности – зависимость вероятности поражения целей от количества ракет. Обе модели в данном случае объективны: они строятся на базе фактических данных, надежного статистического материала.

Однако выходные параметры этих моделей нельзя просто механически объединить.

Для синтеза оценки стоимости и эффективности используется субъективное суждение руководителя, формализовать которое трудно или даже невозможно.

В общем случае на этапе синтеза стоимости и эффективности рекомендуется использовать два основных подхода:

1) фиксированной эффективности при минимально возможной стоимости, т.е. выбирается самая «дешевая» альтернатива, обладающая заданной эффективностью:

$$z_E(x_1, \dots, x_n) = \text{const}_E, z_C(x_1, \dots, x_n) \rightarrow \min;$$

2) фиксированной стоимости и максимально возможной эффективности (случай бюджетных ограничений):

$$z_E(x_1, \dots, x_n) \rightarrow \max, z_C(x_1, \dots, x_n) \rightarrow \text{const}_C.$$

Смысл этих подходов – перевод одного из критериев оценки в ограничение.

Как только это сделано, мы получаем обычную однокритериальную задачу. Главный вопрос – как, на каком уровне установить ограничение одного из критериев. Ответ на этот вопрос в общем случае не вытекает из условий задачи. Ни требуемая эффективность, ни бюджетные ограничения не устанавливаются обычно достаточно жестко.

Итак, одним из способов решения многокритериальных задач, сводящим их к однокритериальным, является перевод всех критериев, кроме одного, в ограничения. Однако, выбор граничных значений, в общем, произволен, причем, чем больше критериев переводится, тем больше произвола.

ПРИМЕР 1. Выбор секретаря из пяти девушек: (1) Ольга, (2) Елена, (3) Светлана, (4) Галина, (5) Жанна, – по критериям: (1) «знание делопроизводства», (2) «внешность», (3) «знание английского языка», (4) «работа на компьютере», (5) «общение по телефону». Характеристика альтернатив с точки зрения критериев:

Альт.Крит.	(1)	(2)	(3)	(4)	(5)
(1)	хор. – 4	хор. – 4	отл. – 5	плохо – 1	удов. – 3
(2)	неуд. – 2	отл. – 5	плохо – 1	плохо – 1	хор. – 4
(3)	отл. – 5	удов. – 3	удов. – 3	хор. – 4	хор. – 4
(4)	хор. – 4	неуд. – 2	неуд. – 2	удов. – 3	отл. – 5
(5)	плохо – 1	хор. – 4	удов. – 3	хор. – 4	неуд. – 2

Числовые значения характеристик формируют матрицу рейтингов альтернатив по критериям.

В качестве главного критерия выбора секретаря выберем «знание делопроизводства» (1). Очевидно, что минимально допустимый уровень всех остальных критериев – «удовлетворительно», т.е. ≥ 3 . Как видно, при такой замене критериев ограничениями единственная приемлемая альтернатива (3) – Светлана. Если понизить требования до ≥ 2 , то приемлемой станет также альтернатива (4) – Галина, и по критерию (1) будет выбрана альтернатива (3) – Светлана.

Заметим, что если в качестве главного критерия выбрать «общение по телефону» (5), то при ограничениях ≥ 2 приемлемыми будут те же две альтернативы, но выбрана будет альтернатива (4) – Галина.

1.2. Обобщённый критерий

Одним из подходов к поиску компромиссного решения задачи векторной оптимизации является сведение её к скалярной (однокритериальной) оптимизации: частные критерии:

$$z_i(X), X = (x_1, \dots, x_n)$$

объединяются в обобщенный критерий:

$$Z(X) = \Phi[z_1(X), z_2(X), \dots, z_m(X)]$$

который затем оптимизируется.

Наиболее распространённым обобщённым критерием является аддитивный критерий – взвешенная сумма частных критериев.

1.2.1. Аддитивный критерий

Аддитивный критерий – это формальный математический приём, придающий задаче удобный для решения вид, однако в нем может происходить взаимная компенсация частных критериев: значительное уменьшение одного из них вплоть до нуля может быть покрыто возрастанием другого. Для ослабления этого недостатка вводятся

ограничения на минимальные значения частных критериев и их весовые коэффициенты.

Обобщённый критерий записывается в виде:

$$Z(X) = \sum_{i=1}^m w_i z_i(X),$$

где w_i – весовые коэффициенты, удовлетворяющие условиям:

$$0 \leq w_i \leq 1, \sum_{i=1}^m w_i = 1.$$

Величина w_i определяет важность i -го частного критерия, при этом более важному критерию приписывается больший вес.

Замечание. Как правило, частные критерии имеют различную размерность, поэтому обобщённый критерий формируется из их нормированных значений: отношений «натуральных» частных критериев к некоторой нормирующей величине.

Варианты выбора нормирующего делителя:

- директивные значения параметров, заданные заказчиком (указанные в ТЗ);
- максимальные значения критериев, достигаемых в области допустимых решений (области D);
- лучшие мировые достижения в данной области;
- разность между максимальным и минимальным значениями критерия в области D :

$$\bar{z}_i(X) = \frac{z_i^{\max} - z_i(X)}{z_i^{\max} - z_i^{\min}} \quad \text{или} \quad \bar{z}_i(X) = \frac{z_i(X) - z_i^{\min}}{z_i^{\max} - z_i^{\min}}.$$

Нормированные критерии будем обозначать через $\bar{z}_i(X)$, т.е. аддитивный критерий примет вид:

$$Z(X) = \sum_{i=1}^m w_i \bar{z}_i(X)$$

Замечание: Если некоторые критерии требуется максимизировать, а другие – минимизировать, следует изменить знак всех критериев одной из групп и решать задачу поиска минимума или максимума критерия $Z(X)$.

Пусть имеется два решения X_1 и X_2 . Для обоснования перехода от X_1 к X_2 необходимо вычислить сумму абсолютных изменений всех частных критериев:

$$\Delta Z = \sum_{i=1}^m w_i [\bar{z}_i(X_2) - \bar{z}_i(X_1)] = \sum_{i=1}^m w_i \bar{z}_i(X_2) - \sum_{i=1}^m w_i \bar{z}_i(X_1)$$

В случае $\Delta Z < 0$ решение X_2 признаётся лучшим, чем X_1 , если критерий $Z(X)$ минимизируется, а если максимизируется, то X_1 признаётся лучшим, чем X_2 . Тогда при поиске минимума $Z(X)$ оптимальному решению X_{opt} соответствует $\Delta Z \geq 0$ при переходе к любому другому решению X :

$$\sum_{i=1}^m w_i \bar{z}_i(X) \geq \sum_{i=1}^m w_i \bar{z}_i(X_{opt}),$$

а при поиске максимума $Z(X)$:

$$\sum_{i=1}^m w_i \bar{z}_i(X) \leq \sum_{i=1}^m w_i \bar{z}_i(X_{opt})$$

Таким образом, оптимальному решению соответствует минимум либо

максимум суммы нормированных частных критериев.

ПРИМЕР 2. Переносной автомат для забивания стальных дюбелей в бетонные стены состоит из корпуса, магазина с дюбелями, подающего спускового механизма с зарядами и ствола. Требуется определить длину ствола L и число дюбелей N при следующих исходных данных: масса m одного дюбеля с зарядом равна 50г, масса ствола 1,6кг/м, масса корпуса 2кг, – и *ограничения*: число дюбелей $N \geq 12$, скорость выброса дюбеля $V \geq 100$ м/с, масса автомата $M \leq 6$ кг.

При фиксированной величине заряда и заданной массе дюбеля скорость выброса связана с длиной ствола соотношением $V = kL^{0.5}$; где $k = 150$ м^{0.5}/с. В качестве частных критериев, которые необходимо максимизировать, выберем V (чем выше скорость выброса, тем надёжнее дюбеля проникают в бетонные стены) и N (чем больше число дюбелей в магазине, тем удобнее работать). По мнению экспертов, оба критерия имеют одинаковую важность.

Введём обозначения: $z_1(N, L) = k \cdot L^{0.5}$ – первый критерий (скорость); $z_2(N, L) = N$ – второй критерий (число дюбелей). Задача многокритериальной оптимизации формулируется следующим образом:

$$\max Z = \max(z_1, z_2)$$

при ограничениях:

$$\begin{aligned} N &\geq 12; \\ V &\geq 100; \\ 1.6L + 0.05N + 2 &\leq 6. \end{aligned}$$

Так как критерии имеют одинаковую важность, то $w_1 = w_2 = 0.5$ и можно их не учитывать. В качестве нормирующих делителей возьмём максимальные значения критериев, достигаемых в области допустимых решений: при $V_{min} = 100$ м/с ограничение на массу автомата примет вид: $1.6(V_{min}/k)^2 + 0.05N + 2 \leq 6$, откуда $N_{max} = 66$. Из того же ограничения при $N = 12$ получим $V_{max} = 219$ м/с.

Замечание. Постулат «низкая оценка по одному критерию может быть компенсирована высокой оценкой по другому» верен далеко не всегда.

Например, если качество оператора ввода текстов оценивать критериями: 1) *скорость ввода (символов в минуту)* и 2) *среднее количество ошибок на страницу текста*, – то увеличение количества ошибок не может быть компенсировано увеличением скорости ввода.

1.3. Мультипликативный критерий

Теоретической основой использования мультипликативного обобщенного критерия задачи многокритериальной оптимизации является *принцип справедливой относительной компенсации*: справедливым следует считать такой компромисс, когда суммарный уровень относительного снижения значений одного или нескольких критериев не превышает суммарного уровня относительного увеличения других критериев.

Математическая формулировка условия оптимальности на основе этого принципа имеет вид:

$$\sum_{i=1}^m \frac{\Delta z_i(X)}{z_i(X)} = 0$$

где $\Delta z_i(X)$, $z_i(X)$ – приращение величины i -го критерия и его первоначальное значение соответственно. Полагая $\Delta z_i(X) \ll z_i(X)$, можно записать:

$$\sum_{i=1}^m \frac{\Delta z_i(X)}{z_i(X)} = \sum_{i=1}^m d(\ln z_i(X)) = d \ln \prod_{i=1}^m z_i(X) = 0,$$

откуда следует, что принцип справедливой относительной компенсации приводит к мультипликативному обобщённому критерию оптимальности:

$$Z(X) = \prod_{i=1}^m z_i(X),$$

в который, в случае неравноценности частных критериев, вводятся весовые коэффициенты w_i :

$$Z(X) = \prod_{i=1}^m z_i^{w_i}(X)$$

Замечание. Мультипликативный критерий иногда представляется в виде отношения:

$$Z(X) = \frac{\prod_{i=1}^{m_1} z_i^+(X)}{\prod_{j=1}^{m_2} z_j^-(X)},$$

где в числителе перемножаются частные критерии, требующие максимизации и имеющие ограничения $z_i^+(X) \geq TT_i$, а в знаменателе – частные критерии, требующие минимизации и имеющие ограничения $z_i^-(X) \leq TT_i$; TT_i – значение технического требования, предъявленного к i -му критерию; $m_1 + m_2 = m$. Такая целевая функция в дальнейшем подвергается максимизации.

Достоинство мультипликативного критерия – при его использовании не требуется нормирование частных критериев.

Недостаток: недостаточная величина одного частного критерия компенсируется избыточной величиной другого.

1.4. Решение оптимизационной задачи методами многокритериального выбора

Аддитивный критерий оптимальности определяется по формуле:

$$F_i[a_{ij}] = \sum_{j=1}^n \lambda_j a_{ij}, \quad i = 1, 2, \dots, m \quad (8.16)$$

где a_{ij} – частные критерии; λ_j – весовые коэффициенты.

$$\sum_{j=1}^n \lambda_j = 1, \quad \lambda_j \geq 0. \quad (8.17)$$

Обобщенная целевая функция (8.16) может быть использована для свертывания частных критериев оптимальности, если:

- частные критерии количественно соизмеримы по важности;
- частные критерии являются однородными.

Если частные критерии не однородны, т.е. имеют различные единицы измерения, то в этом случае требуется нормализация критериев.

Под нормализацией критериев понимается такая последовательность процедур, с помощью которой все критерии приводятся к единому, безразмерному масштабу измерения. Рассмотрим некоторые способы нормализации.

Определим максимум и минимум каждого частного критерия, т.е.:

$$\begin{aligned} a_j^+ &= \max(a_{ij}), \quad i = \overline{1, m}, \\ a_j^- &= \min(a_{ij}), \quad i = \overline{1, m}. \end{aligned} \quad (8.18-8.19)$$

Выделим группу критериев $a_j, j=1, \dots, k$, которые максимизируются при решении задачи, и группу критериев $a_j, j=k+1, \dots, n$, которые минимизируются при решении задачи.

В соответствии с принципом максимальной эффективности нормализованные критерии определяются из соотношений (8.20), (8.21), (8.22), (8.23).

$$\begin{aligned} \hat{a}_{ij} &= \frac{a_{ij}}{a_j^+}, & j = \overline{1, k}, \\ \hat{a}_{ij} &= 1 - \frac{a_{ij}}{a_j^+}, & j = \overline{k+1, n} \end{aligned} \quad (8.20-8.21)$$

или

$$\begin{aligned} \hat{a}_{ij} &= \frac{(a_{ij} - a_j^-)}{(a_j^+ - a_j^-)}, & j = \overline{1, k}, \\ \hat{a}_{ij} &= \frac{(a_j^+ - a_{ij})}{(a_j^+ - a_j^-)}, & j = \overline{k+1, n}. \end{aligned} \quad (8.22-8.23)$$

Оптимальным будет тот вариант, который обеспечивает максимальное значение целевой функции:

$$F_i[a_{ij}] = \max \sum_{j=1}^n \lambda_j \hat{a}_{ij}, \quad i = \overline{1, m}. \quad (8.24)$$

В соответствии с принципом минимальной потери нормализованные критерии определяются соотношениями:

$$\begin{aligned} \hat{a}_{ij} &= 1 - \frac{a_{ij}}{a_j^+}, & j = \overline{1, k}, \\ \hat{a}_{ij} &= \frac{a_{ij}}{a_j^+}, & j = \overline{k+1, n}. \end{aligned} \quad (8.25-8.26)$$

или

$$\begin{aligned} \hat{a}_{ij} &= \frac{(a_j^+ - a_{ij})}{(a_j^+ - a_j^-)}, & j = \overline{1, k}, \\ \hat{a}_{ij} &= \frac{(a_{ij} - a_j^-)}{(a_j^+ - a_j^-)}, & j = \overline{k+1, n}. \end{aligned} \quad (8.27-8.28)$$

При этом оптимальным будет тот вариант, который обеспечивает минимальное значение целевой функции:

$$F_i[a_{ij}] = \min \sum_{j=1}^n \lambda_j \hat{a}_{ij}, \quad i = \overline{1, m}. \quad (8.29)$$

ПРИМЕР 1. Одной из фирм требуется выбрать оптимальную стратегию по обеспечению нового производства оборудованием. С помощью экспериментальных наблюдений были определены значения частных критериев функционирования соответствующего оборудования, выпускаемого тремя заводами-изготовителями. На основе экспертных оценок были также определены веса частных критериев.

Все данные приведены в таблице 1.

Таблица 1

варианты структуры ВС	Частные критерии			
	производительность, д.е.	стоимость, д.е.	энергоёмкость, у.е.	надёжность, у.е.
1	5	7	5	6
2	3	4	7	3
3	4	6	2	4
Весовые	0,4	0,2	0,1	0,3

коэффициенты			
--------------	--	--	--

Решение

- 1) Определим максимум каждого частного критерия: $a_{15}, a_{27}, a_{37}, a_{46}$.
- 2) При решении задачи максимизируются первый (производительность) и четвертый (надежность) критерии, а минимизируются второй (стоимость) и третий (энергоёмкость) критерии.
- 3) Исходя из принципа максимизации эффективности, нормализуем критерии, используя формулы (8.20), (8.21):

$$\hat{a}_{i1} = \frac{a_{i1}}{a_1^+}, \quad i = \overline{1,3}$$

$\hat{a}_{11} = \frac{a_{11}}{a_1^+} = \frac{5}{5} = 1$	$\hat{a}_{21} = \frac{a_{21}}{a_1^+} = \frac{3}{5} = 0,6$	$\hat{a}_{31} = \frac{a_{31}}{a_1^+} = \frac{4}{5} = 0,8$
---	---	---

$$\hat{a}_{i4} = \frac{a_{i4}}{a_4^+}, \quad i = \overline{1,3}$$

$\hat{a}_{14} = \frac{a_{14}}{a_4^+} = \frac{6}{6} = 1$	$\hat{a}_{24} = \frac{a_{24}}{a_4^+} = \frac{3}{6} = 0,5$	$\hat{a}_{34} = \frac{a_{34}}{a_4^+} = \frac{4}{6} = \frac{2}{3}$
---	---	---

$$\hat{a}_{i2} = 1 - \frac{a_{i2}}{a_2^+}, \quad i = \overline{1,3}$$

$$\hat{a}_{i3} = 1 - \frac{a_{i3}}{a_3^+}, \quad i = \overline{1,3}$$

$\hat{a}_{12} = 1 - \frac{a_{12}}{a_2^+} =$ $= 1 - \frac{7}{7} = 0$	$\hat{a}_{22} = 1 - \frac{a_{22}}{a_2^+} =$ $= 1 - \frac{4}{7} = \frac{3}{7}$	$\hat{a}_{32} = 1 - \frac{a_{32}}{a_2^+} =$ $= 1 - \frac{6}{7} = \frac{1}{7}$
$\hat{a}_{13} = 1 - \frac{a_{13}}{a_3^+} =$ $= 1 - \frac{5}{7} = \frac{2}{7}$	$\hat{a}_{23} = 1 - \frac{a_{23}}{a_3^+} =$ $= 1 - \frac{7}{7} = 0$	$\hat{a}_{33} = 1 - \frac{a_{33}}{a_3^+} =$ $= 1 - \frac{2}{7} = \frac{5}{7}$

- 4) Определим обобщенную функцию цели по каждому варианту, используя формулу (8.24).

$$F_1 = \lambda_1 \hat{a}_{11} + \lambda_2 \hat{a}_{12} + \lambda_3 \hat{a}_{13} + \lambda_4 \hat{a}_{14} =$$

$$= 0,4 \cdot 1 + 0,2 \cdot 0 + 0,1 \cdot \frac{2}{7} + 0,3 \cdot 1 \approx 0,729$$

$$F_2 = \lambda_1 \hat{a}_{21} + \lambda_2 \hat{a}_{22} + \lambda_3 \hat{a}_{23} + \lambda_4 \hat{a}_{24} =$$

$$= 0,4 \cdot 0,6 + 0,2 \cdot \frac{3}{7} + 0,1 \cdot 0 + 0,3 \cdot 0,5 \approx 0,476$$

$$F_3 = \lambda_1 \hat{a}_{31} + \lambda_2 \hat{a}_{32} + \lambda_3 \hat{a}_{33} + \lambda_4 \hat{a}_{34} =$$

$$= 0,4 \cdot 0,8 + 0,2 \cdot \frac{1}{7} + 0,1 \cdot \frac{5}{7} + 0,3 \cdot \frac{2}{3} \approx 0,603$$

- 5) Оптимальным является первый вариант оборудования, так как:

$$F_{\max} = F_1 = 0,729.$$

2. Задание на лабораторную работу

2.1. Изучить предлагаемые методы

2.2. Выполнить индивидуальное задание: сформулировать, записать исходные данные и решить аналитически многокритериальную задачу методом аддитивной оптимизации.

Варианты индивидуальных заданий

№1 Экономические показатели эффективности работы предприятий приведены в таблице «Показатели эффективности работы предприятий»

Показатели эффективности работы предприятий

№ предприятия	Показатели эффективности работы предприятий				
	Прибыль, д. е.	Себест. единицы продукции, д. е.	Доходы, д. е.	Фондо-отдача, у. е.	Производит. у. е.
1	30,0	40,0	20,0	0,2	300
2	25,0	20,0	30,0	0,3	200
3	40,0	45,0	54,0	0,1	250
4	28,0	30,0	35,0	0,4	160
5	15,0	12,0	20,0	0,25	280
6	50,0	30,0	40,0	0,21	120
Весов. коэф.	0,32	0,23	0,15	0,2	0,1

Определить и выбрать наиболее эффективно работающее предприятие

№2

Абсолютные показатели качества различных вариантов двигателей приведены в таблице «Показатели качества двигателей»

Показатели качества двигателей

Варианты двигателей	Показатели качества		
	мощность, л. с.	крутящий момент, кгс-м	масса, кг
1	180	67	850
2	176	70	1002
3	176	68	860
4	181	67	818
5	177	68	860
6	180	66	801
Весовые коэф.	0,4	0,24	0,36

Выбрать из предложенных вариантов оптимальный двигатель.

№3

Для пяти проектов технических систем определены относительные единичные показатели технического совершенства конструкции и весовые коэффициенты единичных показателей. Численные значения единичных показателей и весовых коэффициентов приведены в таблице «Показатели проектов технических систем»

Показатели проектов технических систем

Варианты технических систем	Относительные единичные показатели					
	сложности	веса	времени подготовки	автоматизации	мощности	унификации
I	1,0	0,88	1,0	1,0	0,72	0,614
II	0,72	1,2	0,8	0,78	0,81	0,42
III	0,658	0,358	0,765	0,782	0,525	0,915
IV	0,425	0,97	0,755	0,7	0,98	0,31
V	0,467	0,555	0,865	0,705	0,865	0,65
Весовые коэф.	0,157	0,124	0,21	0,195	0,174	0,14

Определить оптимальный проект технической системы.

№4

Для пяти проектов технических систем определены относительные параметры технического совершенства конструкции. Значения параметров и весовых коэффициентов приведены в таблице «Параметры проектов технических систем».

Вариант технической системы	Сложность конструкции	Масса и габарит	Время подготовки к работе	Уровень автоматизации	Мощность
I	0,542	0,881	1,0	0,79	0,82
II	0,623	1,0	0,91	0,88	0,81
III	0,558	0,558	0,965	0,82	0,65
IV	0,525	0,972	0,851	0,71	0,98
V	0,567	0,565	0,867	0,77	0,86
Весовые коэф.	0,177	0,124	0,24	0,295	0,164

Выбрать оптимальный проект технической системы

№5

Одной из фирм требуется выбрать оптимальную стратегию по техническому обеспечению процесса производства оборудованием. С помощью статистических данных и информации соответствующих заводов-изготовителей были определены локальные критерии эффективности функционирования вариантов необходимого оборудования. Исходные данные представлены в таблице «Локальные критерии эффективности оборудования».

Локальные критерии эффективности оборудования

Варианты оборудования	Локальные критерии эффективности оборудования			
	производительность, д. е.	стоимость оборудования, д. е.	объем памяти, у. е.	надежность, у. е.
I	100	5	5	8
II	150	6	8	5
III	120	4	6,5	6
IV	200	7	6	4
V	130	5	7	7
Весовые коэф.	0,20	0,25	0,23	0,32

Выбрать наиболее эффективный вариант поставки оборудования

3. Вопросы к лабораторной работе

- 1) Какие существуют методы свертывания критериев в многокритериальных задачах?
- 2) В чем заключается метод аддитивной оптимизации?
- 3) Что такое весовой коэффициент?
- 4) Как определяется обобщенная целевая функция в методе аддитивной оптимизации?
- 5) В чем заключается алгоритм нормализации критериев?

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие/ А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Т.В. Хоменко

МЕТОДЫ ОПТИМИЗАЦИИ
В БИЗНЕС-АНАЛИТИКЕ
ОПТИМИЗАЦИОННЫЕ МЕТОДЫ ОЦЕНКИ ПАРАМЕТРОВ

Методические указания для проведения лабораторных работ
для студентов направления
09.04.03 «Прикладная информатика»
направленность «Искусственный интеллект и бизнес-аналитика»

Ульяновск
УлГТУ
2021

УДК
ББК

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук,
доцент К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Составитель:

Хоменко, Татьяна Владимировна

ОПТИМИЗАЦИОННЫЕ МЕТОДЫ ОЦЕНКИ ПАРАМЕТРОВ: методические
указания для выполнения лабораторных работ по дисциплине «МЕТОДЫ
ОПТИМИЗАЦИИ В БИЗНЕС-АНАЛИТИКЕ» для студентов направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и бизнес-
аналитика» / Составитель: Т. В. Хоменко. – Ульяновск: УлГТУ, 2021. – 30с.

Составлены в соответствии с учебным планом направления 09.04.03
«Прикладная информатика» направленность «Искусственный интеллект и
бизнес-аналитика». Цель данного практикума – ориентировать студентов на
содержание и порядок выполнения лабораторных работ во время прохождения
студентами курсов «Искусственный интеллект и бизнес-аналитика». Даются
задания для лабораторных занятий, а также приводится обзор возможных
данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК

ББК

© Т.В. Хоменко, 2021.

© Оформление. УлГТУ, 2021.

Содержание

ВВЕДЕНИЕ	4
Лабораторная работа №7	5
Тема: Метод моментов Обобщенный метод моментов	5
1. Краткие теоретические сведения	5
2. Задание на лабораторную работу	11
3. Вопросы к лабораторной работе	11
Лабораторная работа №8	12
Тема: Метод анализа средних	12
1. Краткие теоретические сведения	12
2. Задание на лабораторную работу	17
3. Вопросы к лабораторной работе	19
Список литературы.....	20

ВВЕДЕНИЕ

В методических указаниях описаны лабораторные работы по дисциплине «Методы оптимизации в бизнес-аналитике», а также краткая теория, необходимая для их выполнения. Более полное изложение теории с примерами решения задач даются в учебных пособиях [1-4]

Для успешного освоения дисциплины и выполнения заданий лабораторных работ необходимо предварительное изучение таких дисциплин, как «Математика» (линейная алгебра, аналитическая геометрия, дифференциальное исчисление) и «Численные методы» (или ее аналогов – «Вычислительная математика», «Методы вычислений»). Также для выполнения лабораторных работ потребуются навыки работы с математическими пакетами (PTC MathCAD, MathWorks MATLAB, Wolfram Mathematica и т. п.) и/или программирования на языке высокого уровня (Pascal, C/C++, C# и т. д.).

В каждой лабораторной работе необходимо изучить теоретический материал, выполнить задание п.2 и ответить на вопросы п.3.

Лабораторная работа №7

Тема: Метод моментов Обобщенный метод моментов

Цель работы: знакомство с оптимизационными задачами, изучение различных методов оценки параметров и сравнение эффективности их применения.

1. Краткие теоретические сведения

1.1. Основные понятия

Пусть x_1, x_2, \dots, x_n – выборка из теоретического распределения $F(x|\theta)$, зависящего от неизвестного параметра θ .

Оценкой (статистикой) $\hat{\theta}_n$ неизвестного параметра θ называется любая борелевская функция $\hat{\theta}_n = \hat{\theta}_n(x_1, x_2, \dots, x_n)$.

Оценка $\hat{\theta}_n$ неизвестного параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру:

$$M \hat{\theta}_n = \theta,$$

В противном случае оценка называется *смещенной*.

Можно показать, что выборочное среднее \bar{x} является несмещенной оценкой теоретического математического ожидания, а выборочная дисперсия \tilde{S}^2 является смещенной оценкой теоретической дисперсии.

Исправленная выборочная дисперсия:

$$S^2 = \frac{n}{n-1} \tilde{S}^2$$

является несмещенной оценкой теоретической дисперсии.

Оценка $\hat{\theta}_n$ неизвестного параметра θ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру:

$$\lim_{n \rightarrow \infty} P \{ |\hat{\theta}_n - \theta| < \epsilon \} = 1.$$

Можно показать, что оценки \bar{x} , \tilde{S}^2 и S^2 являются состоятельными оценками теоретического математического ожидания и теоретической дисперсии соответственно.

Несмещенная оценка $\hat{\theta}_n$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию в классе несмещенных оценок, т.е. если $\hat{\theta}_n$ – произвольная несмещенная оценка параметра θ , а $\hat{\theta}_n$ – эффективная, то $D(\hat{\theta}_n) \leq D(\tilde{\theta}_n)$.

Точечная оценка предполагает нахождение единственной числовой величины, которая и принимается за значение искомого параметра. Существует несколько методов получения точечных оценок.

1.2. Метод моментов

Пусть имеется выборка $X_{(n)} = (X_1, \dots, X_n)$ из распределения $F(x, \theta_1, \dots, \theta_k)$, зависящего от неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$, которые нужно оценить.

Поскольку известен вид теоретической функции распределения, можем вычислить

первые k теоретических моментов (начальных или центральных). Эти моменты будут зависеть от k неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$:

$$\begin{cases} m_1 = m_1(\theta_1, \theta_2, \dots, \theta_k), \\ m_2 = m_2(\theta_1, \theta_2, \dots, \theta_k), \\ \dots \\ m_k = m_k(\theta_1, \theta_2, \dots, \theta_k). \end{cases}$$

Суть метода моментов заключается в следующем:

Выборочные моменты являются оценками соответствующих теоретических моментов, поэтому теоретические моменты m_1, m_2, \dots, m_k приравнивают к соответствующим выборочным $m_1^*, m_2^*, \dots, m_k^*$, а затем, решая систему относительно $\theta_1, \theta_2, \dots, \theta_k$, находят оценки неизвестных параметров.

Таким образом, в методе моментов оценки $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ неизвестных параметров $\theta_1, \theta_2, \dots, \theta_k$ определяются как решение системы k уравнений с k неизвестными:

$$\begin{cases} m_1^* = m_1(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k) \\ m_2^* = m_2(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k) \\ \dots \\ m_k^* = m_k(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k) \end{cases}$$

В таблице 1 приведены формулы для вычисления выборочных и соответствующих им теоретических моментов порядка k .

Таблица 1

Выборочные и теоретические моменты

Моменты	Теоретические	Выборочные
Начальные	$b_k = M\xi^k$	$b_k^* = \overline{x^k} = \frac{1}{n} \sum_i x_i^k$
Центральные	$c_k = M(\xi - \bar{\xi})^k$	$c_k^* = \tilde{S}^k = \frac{1}{n} \sum_i (x_i - \bar{x})^k$

Таким образом, если имеется выборка $X_{[n]} = (X_1, \dots, X_n)$ из генеральной совокупности с функцией распределения: $F(x, \theta_1, \dots, \theta_k)$, такая, что все моменты конечны:

$$a_r(\theta_1, \dots, \theta_k) = \int_{-\infty}^{\infty} x^r dF(x, \theta_1, \dots, \theta_k), \quad r = 1, \dots, k,$$

и что система уравнений:

$$a_r(\theta_1, \dots, \theta_k) = a_r^*, \quad r = 1, \dots, k$$

однозначно разрешима относительно:

$$\theta_1, \theta_2, \dots, \theta_k, \quad \text{где} \quad a_r^* = \frac{1}{n} \sum_{i=1}^n X_i^r$$

тогда оценки:

$$\hat{\theta}_r = g_r(a_1^*, \dots, a_k^*), \quad r = 1, \dots, k,$$

получаемые как решение системы уравнений называются *оценками по методу моментов*.

ПРИМЕР 1. Пусть задана выборка X , полученная из показательного распределения с параметром. Найти точечную оценку параметра методом моментов.

Для показательного распределения плотность распределения вероятностей имеет вид:

$$p(x) = \lambda e^{-\lambda x}, x \geq 0$$

Для данного закона распределения определим теоретический момент:

$$b_1 = \int_{-\infty}^{+\infty} xp(x)dx = \int_0^{+\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

По заданной выборке определим значение выборочного момента:

$$b_1^* = \bar{x} = \frac{1}{30} \sum_{i=1}^{30} x_i \approx 0,05.$$

Составим уравнение:

$$\frac{1}{\lambda} = \bar{x}$$

Таким образом, получим оценку:

$$\hat{\lambda} = 20.$$

ПРИМЕР 2. Пусть задана выборка X , полученная из равномерного распределения на отрезке $[a, b]$. Методом моментов найти точечные оценки параметров a и b .

Для равномерного распределения плотность распределения вероятностей имеет вид:

$$p(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

Равномерное распределение имеет два параметра, поэтому необходимо составить два уравнения. Определим теоретические моменты:

$$b_1 = \int_{-\infty}^{+\infty} xp(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2};$$

$$b_2 = \int_{-\infty}^{+\infty} x^2 p(x)dx = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}$$

По выборке определим значения выборочных моментов:

$$b_1^* = \bar{x} = \frac{1}{30} \sum_{i=1}^{30} x_i \approx 0,05;$$

$$b_2^* = \overline{x^2} = \frac{1}{30} \sum_{i=1}^{30} x_i^2 \approx 1,26.$$

Составим систему уравнений:

$$\begin{cases} \frac{a+b}{2} = 0,05, \\ \frac{a^2 + ab + b^2}{3} = 1,26. \end{cases}$$

Решив ее относительно неизвестных параметров a и b , получим:

$$a_1 = -1,89, \quad b_1 = 1,99;$$

$$a_2 = 1,99, \quad b_2 = -1,89$$

Затем, вместо истинного момента берем выборочный. Поскольку $a < b$, то имеем:

$$\hat{a} = -1,89,$$

$$\hat{b} = 1,99.$$

Можно сказать, что мы берем в качестве оценки такое (случайное) значение параметра, при котором истинный момент совпадает с выборочным.

1.3. Обобщённый метода моментов

Пусть имеется случайная выборка z_1, \dots, z_n и решается уравнение вида:

$$\frac{1}{n} \sum_{i=1}^n m(z_i, \hat{\theta}) = 0.$$

Случай 1. Пусть, неизвестных столько же, сколько уравнений и $\hat{\theta}$ - оценка метода моментов. Тогда, точечная оценка предполагает нахождение единственной числовой величины:

– рассматривая только первое из условий, оценка метода моментов будет равна:

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n z_i.$$

– рассматривая первое и второе из условий уравнения, оценка метода моментов будет иметь неявный вид:

$$\sum_{i=1}^n (z_i - \hat{\theta}_2)^3 = 0.$$

– рассматривая только первое или второе из условий, оценка метода моментов будет соответственно равна:

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^n x_i, \quad \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^n y_i, \quad \text{где } \hat{\theta}_1 \neq \hat{\theta}_2.$$

Случай 1. Пусть, неизвестных больше, чем уравнений и $\hat{\theta}$ - оценка метода моментов. Тогда, возникает проблема нахождения точечной оценки.

Решением проблемы служит преобразование условий на моменты в оптимизационную задачу:

$$E[m(z, \theta)] = 0 \quad \Rightarrow \quad \theta = \arg \min_{q \in \Theta} \underbrace{E[m(z, q)]'}_{1 \times l} \underbrace{W}_{l \times l} \underbrace{E[m(z, q)]}_{l \times 1},$$

где W - любая симметричная положительно определённая матрица.

Теперь можно построить оценку обобщённого метода моментов, которая гарантировано существует:

$$\hat{\theta} = \arg \min_{q \in \Theta} \left(\frac{1}{n} \sum_{i=1}^n m(z_i, q) \right)' W_n \left(\frac{1}{n} \sum_{i=1}^n m(z_i, q) \right)$$

где W_n - симметричная положительно определённая матрица, состоятельно оценивающая W :

$$W_n \xrightarrow{P} W.$$

Замечание. В типичных случаях W_n зависит от данных, $\hat{\theta}$ зависит от выбора W и W_n .

1.4. Метод максимального правдоподобия

Метод получения оценок, который чаще всего приводит к наилучшим оценкам, называется метод максимального правдоподобия (предложен Р. Фишером). Этот метод применяется в случаи, когда известен закон распределения, но не известен параметр Θ , которым определяется функция плотности $f(x, \Theta)$ распределения для данного закона распределения. Метод максимального правдоподобия заключается в следующем:

В качестве оценок параметра Θ принимается то значение $\tilde{\Theta}$, при котором функция правдоподобия достигает максимума:

$$L = \prod_{i=1}^n f(x_i, \tilde{\Theta}) \rightarrow \max$$

Для нахождения оценок необходимо исследовать функцию правдоподобия на экстремум. С целью упрощения вычислительной процедуры находится экстремум логарифма функции правдоподобия $\ln L(x, \tilde{\Theta})$ (Экстремум этой функции достигается при тех же значениях $\tilde{\Theta}$). Нахождение оценок сводится к решению нелинейного уравнения(уравнений) относительно искомого параметра (параметров) вида[^]

$$\frac{d \ln L(x, \tilde{\Theta})}{d \tilde{\Theta}} = 0.$$

Таким образом, если имеется выборка $X_{[n]} = (X_1, \dots, X_n)$ из генеральной совокупности с функцией распределения $F(x, \theta_1, \dots, \theta_k)$, которая имеет плотность распределения $f(x, \theta_1, \dots, \theta_k)$ и функцию правдоподобия выборки вида:

$$L(X_1, \dots, X_n, \theta_1, \dots, \theta_k) = \prod_{i=1}^n f(X_i, \theta_1, \dots, \theta_k),$$

тогда, оценками по методу максимального правдоподобия $\hat{\theta}_i = \hat{\theta}_i(X_{[n]})$, $i = 1, \dots, k$, называются оценки, которые обращают в максимум функцию правдоподобия:

$$L(X_{[n]}, \hat{\theta}_1, \dots, \hat{\theta}_k) = \max_{\theta_1, \dots, \theta_k} L(X_{[n]}, \theta_1, \dots, \theta_k)$$

Аналогично определяются оценки по методу максимального правдоподобия в дискретном случае. При нахождении оценок максимального правдоподобия иногда бывает удобно логарифмировать функцию правдоподобия, после чего вычислять производные по параметрам и приравнять их к нулю, получая так называемое *уравнение правдоподобия*, однако следует помнить, что это не более, чем прием решения задачи максимизации функции правдоподобия, который далеко не всегда приводит к успеху.

ПРИМЕР 3. Рассмотрим параметрическое семейство определяемое плотностью следующего вида:

$$f(x, \theta) = \begin{cases} \theta^{-1}, & x \in [0, \theta] \\ 0, & x \notin [0, \theta] \end{cases}$$

Нельзя найти оценку по методу максимального правдоподобия в этом примере, решая уравнение правдоподобия. Однако нетрудно построить график функции правдоподобия, как функции параметра θ , для фиксированной выборки. График функции имеет разрыв. Максимум достигается в точке $\hat{\theta} = \max\{X_1, \dots, X_n\}$.

1.5. Обобщенный метод максимального правдоподобия

Пусть повторная выборка $\{x_i\}_{1 \leq i \leq n}$ имеет плотность распределения $f(x; \theta)$, $\theta \in \Theta$. Выберем произвольное положительное число r и пусть $L_x(\theta)$ - функция правдоподобия, построенная по выборке $x = (x_1, \dots, x_n)$. В этом случае статистика:

$$\delta_r(x) = \max_{\theta} \int_{\theta-r}^{\theta+r} L_x(\theta) d\theta,$$

называется *оценкой максимальной вероятности* параметра θ .

Устремляя r к 0 ($r \rightarrow 0$) в пределе получим оценку максимального правдоподобия.

ПРИМЕР 4. Пусть X_i подчиняется показательному распределению с параметром сдвига $\theta \in \mathbb{R}$, т.е.:

$$f(x; \theta) = \exp(-(x - \theta))$$

для $x \geq \theta$, и нуль в противном случае.

Тогда:

$$L_x(\theta) = h(X^{(1)} - \theta) \exp\left(-\sum_{i=1}^n X_i + n\theta\right),$$

$$h(x) = 1, x \geq 0; h(x) = 0, x < 0.$$

При фиксированном $r > 0$ получим:

$$\delta_r(x) = X^{(1)} - r.$$

Квадратичный риск этой оценки равен:

$$E_{\theta}((\delta_r(X) - \theta)^2) = E_{\theta}((X^{(1)} - r - \theta)^2) = r^2 - \frac{2r}{n} + \frac{2}{n^2}.$$

Минимизируя его по r , получаем оценку:

$$\delta_{1/n}(X) = X^{(1)} - 1/n.$$

2. Задание на лабораторную работу

2.1. Изучить теоретический материал

2.2. В соответствии с вариантом задания, определенным преподавателем, найти точечную оценку параметра: методом моментов, методом максимального правдоподобия, для выборки X (таблица 2), полученной из распределения, вид которого определен соответственно варианту задания

Таблица 2

№п/п	Y_1	Y_2	Y_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
1	9,26	204,2	13,26	0,23	0,78	0,40	1,37	1,23	0,23	1,45	26006	167,69	47750	6,40	17,72
2	9,38	209,6	10,16	0,24	0,75	0,26	1,49	1,04	0,39	1,30	23935	186,10	50391	7,80	18,39
3	12,11	222,6	13,72	0,19	0,68	0,40	1,44	1,80	0,43	1,37	22589	220,45	43149	9,76	26,46
4	10,81	236,7	12,85	0,17	0,70	0,50	1,42	0,43	0,18	1,65	21220	169,30	41089	7,90	22,37
5	9,35	62,0	10,63	0,23	0,62	0,40	1,35	0,88	0,15	1,91	7394	39,53	14257	5,35	28,13
6	9,87	53,1	9,12	0,43	0,76	0,19	1,39	0,57	0,34	1,68	11586	40,41	22661	9,90	17,55
7	8,17	172,1	25,83	0,31	0,73	0,25	1,16	1,72	0,38	1,94	26609	102,96	52509	4,50	21,92
8	9,12	56,5	23,39	0,26	0,71	0,44	1,27	1,70	0,09	1,89	7801	37,02	14903	4,88	19,52
9	5,88	52,6	14,68	0,49	0,69	0,17	1,16	0,84	0,14	1,94	11587	45,74	25587	3,46	23,99
10	6,30	46,6	10,65	0,36	0,73	0,39	1,25	0,60	0,21	2,06	9475	40,07	16821	3,60	21,76
11	6,22	53,2	13,99	0,37	0,68	0,33	1,13	0,82	0,42	1,96	10811	45,44	19459	3,56	25,68
12	5,49	31,0	9,68	0,43	0,74	0,25	1,10	0,84	0,05	1,02	6371	41,08	12973	5,65	18,13
13	6,50	146,4	10,03	0,35	0,66	0,32	1,15	0,67	0,29	1,85	26761	136,14	50907	4,28	25,74
14	6,61	18,1	9,13	0,38	0,72	0,02	1,23	1,04	0,48	0,88	4210	42,39	6920	8,85	21,21
15	4,32	13,6	5,37	0,42	0,68	0,06	1,39	0,66	0,41	0,62	3557	37,39	5736	8,52	22,97
16	7,37	89,8	9,86	0,30	0,77	0,15	1,38	0,86	0,62	1,09	14148	101,78	26705	7,19	16,38
17	7,02	62,5	12,62	0,32	0,78	0,08	1,35	0,79	0,56	1,60	9872	47,55	20068	4,82	13,21
18	8,25	46,3	5,02	0,25	0,78	0,20	1,42	0,34	1,76	1,53	5975	32,61	11487	5,46	14,48
19	8,15	103,5	21,18	0,31	0,81	0,20	1,37	1,60	1,31	1,40	16662	103,25	32029	6,20	13,38
20	8,72	73,3	25,17	0,26	0,79	0,30	1,41	1,46	0,45	2,22	9166	38,95	18946	4,25	13,69
21	6,64	76,6	19,40	0,37	0,77	0,24	1,35	1,27	0,50	1,32	15118	81,32	28025	5,38	16,66
22	8,10	73,01	21,0	0,29	0,78	0,10	1,48	1,58	0,77	1,48	11429	67,26	20968	5,88	15,06
23	5,52	32,3	6,57	0,34	0,72	0,11	1,24	0,68	1,20	0,68	6462	59,92	11049	9,27	20,09
24	9,37	199,6	14,19	0,23	0,79	0,47	1,40	0,86	0,21	2,30	24628	107,34	45893	4,36	15,98
25	13,17	598,1	15,81	0,17	0,77	0,53	1,45	1,98	0,25	1,37	49727	512,60	99400	10,31	18,27
26	6,67	71,2	5,23	0,29	0,80	0,34	1,40	0,33	0,15	1,51	11470	53,81	20719	4,69	14,42
27	5,68	90,8	7,99	0,41	0,71	0,20	1,28	0,45	0,66	1,43	19448	80,83	36813	4,16	22,76
28	5,22	82,1	17,50	0,41	0,79	0,24	1,33	0,74	0,74	1,82	18965	59,42	33936	3,13	15,41
29	10,02	76,2	17,16	0,22	0,76	0,54	1,22	0,03	0,32	2,62	9185	36,96	17016	4,02	19,35
30	8,16	119,5	14,54	0,29	0,78	0,40	1,28	0,99	0,89	1,75	17478	91,43	34873	5,23	16,83

2.3. Оформить отчёт о выполнении задания с приведением условия задачи, алгоритмов и/или программ, указанных в задании методов поиска, результатов задачи.

Представить результаты исследований в таблицах и графиками. Сделать выводы по проведенным исследованиям. В отчете должны быть отражены все этапы выполнения работы и представлены выводы и анализ результатов. Расчеты рекомендуется проводить при помощи табличного процессора Excel, с приложением листингов в отчете по лабораторной работе.

3. Вопросы к лабораторной работе

- 1) Дайте определение статистической оценки неизвестного параметра теоретического распределения
- 2) В чем заключается суть метода моментов?
- 3) В чем заключается суть метода максимального правдоподобия?
- 4) Раскрыть обобщённый метода моментов
- 5) Раскрыть обобщённый метода максимального правдоподобия

Лабораторная работа №8

Тема: Метод анализа средних

Цель работы: знакомство с оптимизационными задачами, изучение различных методов оптимизации и сравнение эффективности их применения для конкретных целевых функций.

1. Краткие теоретические сведения

1.1. Основные понятия

Одной из наиболее практически значимых задач в бизнес-аналитике является построение плавных временных зависимостей функции $x(t)$, если эта функция имеет случайную составляющую.

Дискретную зависимость $x_i(t_i)$ называют *временным рядом* или, в общем случае, *дискретным случайным процессом*.

По сравнению с экспериментальными зависимостями $y = f(x)$, полученных в результате однофакторного эксперимента, особенности экспериментальных временных рядов заключаются в следующем:

- на практике обычно приходится анализировать временные ряды с достаточно большим количеством отсчетов (не менее нескольких десятков),
- отсчеты, как правило, производятся через равные промежутки времени (равноотстоящие узлы зависимости при $t_{i+1} - t_i = const$),
- зависимость $x_i(t_i)$ заведомо немонотонна и, чаще всего, ограничена, поэтому, если выбрать конечный интервал Δx , то для любого x с увеличением длины временного ряда количество значений x_i , попадающих в интервал $[x, x + \Delta x]$ будет увеличиваться.

Приближающую функцию, аппроксимирующую временной ряд по всей его длине, как правило, невозможно описать аналитически. На рисунке 1 представлен временной ряд на графике.

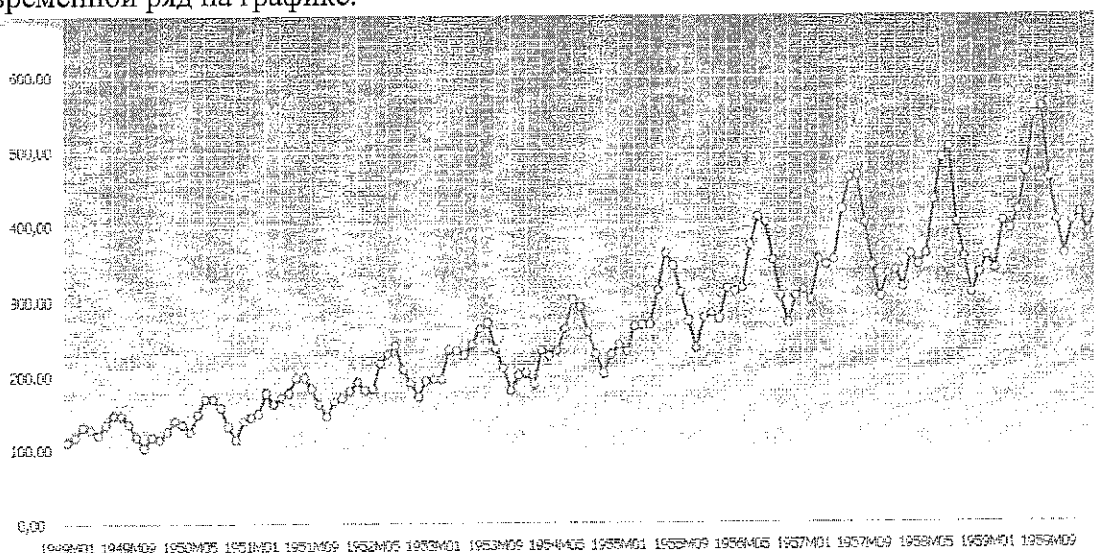


Рисунок 1 – Временной ряд на графике

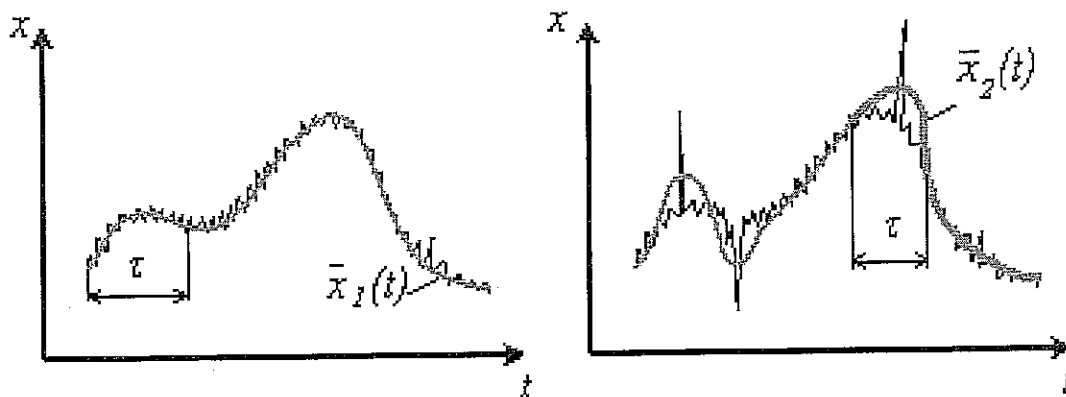
Особенности временных рядов хорошо поясняет простейший пример.

ПРИМЕР 1. Проведем однофакторный эксперимент, регистрируя значения входных x_i и выходных y_i параметров одновременно, в моменты времени t_i , через равные промежутки $t_{i+1} - t_i = const$. Тогда функциональная зависимость $y_i = f(x_i)$, описывающая результаты эксперимента, задается параметрически, через два временных ряда $x_i(t_i)$, $y_i(t_i)$.

В теории обработки временных рядов существует множество способов их сглаживания: фильтрация с использованием преобразования Фурье, кусочная аппроксимация многочленами и другие. Наиболее используемыми методами являются два простых, но принципиально разных вида сглаживания, которые, в какой-то мере, обобщают особенности этой процедуры:

- 1) метод скользящего среднего;
- 2) медианное сглаживание.

Рассмотрим две случайные зависимости, показанные на рисунке 2, которые имеют одинаковую регулярную составляющую $\bar{x}(t)$. Однако, при этом, а) зависимости отличаются случайными составляющими, б) кроме мелких случайных флуктуаций имеет редкие выбросы достаточно большой амплитуды



а) б)
Рисунок 2 – Случайные зависимости

Целью сглаживания является получение плавной зависимости $\bar{x}(t)$.

1.2. Метод скользящего среднего

Метод скользящего среднего предполагает выбор окна усреднения τ , и для каждого t_i рассчитывается среднее значения на этом интервале:

$$\bar{x}_1(t_i) = \frac{1}{\tau} \int_{t_i - \frac{\tau}{2}}^{t_i + \frac{\tau}{2}} x(t) dt \quad (8.1)$$

Этот метод позволяет сгладить случайную составляющую зависимости, то есть избавиться от высокочастотных флуктуаций. При этом фильтрация высоких частот зависит от длины интервала усреднения τ .

Таким образом, метод скользящей средней состоит в замене фактических уровней временного ряда расчетными, имеющими значительно меньшую колеблемость, чем исходные данные. При этом средняя рассчитывается по группам данных за определенный интервал времени, причем каждая последующая группа образуется со сдвигом на один год (месяц). В результате подобной операции первоначальные колебания временного ряда сглаживаются, поэтому и операция называется сглаживанием рядов динамики (основная тенденция развития выражается при этом уже в виде некоторой плавной линии).

Если применить метод скользящего среднего к зависимости (рисунок 2.б), имеющей значительные, но редкие выбросы, то полученная сглаженная зависимость $\bar{x}_2(t)$ резко отличается от $\bar{x}_1(t)$. Выбросы, за счет их высокой амплитуды, сильно влияют на среднее значение, как только попадают в интервал усреднения, то есть каждый выброс искажает $\bar{x}(t)$ на интервале длиной τ . В случае, когда выбросы имеют редкий и случайный характер, для выделения регулярной составляющей метод скользящего среднего неприемлем.

Метод скользящей средней рассматривается как один из методов статистического прогнозирования. Необходимость применения скользящей средней вызывается следующими обстоятельствами: имеющиеся данные ряда не позволяют обнаруживать какую-либо тенденцию развития (тренд) того или иного процесса (из-за случайных и периодических колебаний исходных данных). В таких случаях для лучшего выявления тенденции прибегают к методу скользящей средней.

Экстраполяция по скользящей средней – может применяться для целей краткосрочного прогнозирования.

ПРИМЕР 2. Количество сбоев в информационной системе торговой компании составило: в январе - 60, в феврале - 85, в марте - 80, в апреле - 92, в мае – 88, в июне - 96, тогда прогноз сбоев на июль (для 5-ти месячного периода) составит:

$$(85 + 80 + 92 + 88 + 96) / 5 = 88,2.$$

Если реальное количество сбоев в июле составило 94, то их прогноз на август уже будет равен:

$$(80 + 92 + 88 + 96 + 94) / 5 = 90 \text{ и так далее.}$$

Число значений n для подсчета скользящей средней (в нашем примере $n = 5$) выбирается в зависимости от того, насколько важны старые значения исследуемого показателя в сравнении с новыми.

Так, если мы будем использовать для подсчета 3-х месячный период, тогда прогноз количества сбоев на июль составит:

$$(92 + 88 + 96) / 3 = 92,3.$$

В случае с 5-ти месячной средней старые значения имеют удельный вес $4/5$, а текущие – $1/5$.

В случае с 3-х месячной средней старые значения “весят” $2/3$, а текущие – $1/3$, т.е. скользящая средняя уже в большей степени зависит от текущего уровня и несколько слабее – от предшествующего.

Тренд – изменение, определяющее общее направление развития, основную тенденцию рядов динамики.

Для выявления тренда ряда динамики можно использовать метод скользящей

средней, в котором вместо фактического уровня берётся средняя, которая рассчитывается из нескольких уровней. Эта средняя будет скользящей, поскольку период усреднения постоянно меняется, вычитая один уровень и прибавляя другой.

$$Z_1 = (z_1 + z_2 + z_3 + z_4) / 4,$$

$$Z_2 = (z_2 + z_3 + z_4 + z_5) / 4,$$

а в общем виде

$$Z_m = (z_m + z_{m+1} + \dots + z_{m+n-1}) / n,$$

где n – количество выбранных значимых элементов временного ряда (выбранный период усреднения) для расчета скользящей средней,

m – уровень расчета скользящей средней,

$m = N - n$, $n, m < N$, N – длина временного ряда (количество его значений).

Скользящие средние не спрогнозируют изменения в тренде, а лишь просигналят об уже появившемся тренде.

ПРИМЕР 3. Имеется следующий ряд динамики:

t_i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	2	4	3	4	6	8	5	5	9	11	7	8

Определить значения всех скользящих средних, взяв период усреднения $n = 3$. Построить в одной системе координат график временного ряда и график его тренда по полученным скользящим средним. Расчеты по формулам и построение графиков осуществить посредством использования приложения MS Excel.

Вычислим значения скользящих средних.

$$Z_1 = (x_1 + x_2 + x_3) / 3 = (2+4+3)/3 = 3 \text{ – первая скользящая средняя;}$$

$$Z_2 = (x_2 + x_3 + x_4) / 3 = (4+3+4)/3 = 3,7 \text{ – вторая скользящая средняя;}$$

$$Z_3 = (x_3 + x_4 + x_5) / 3 = (3+4+6)/3 = 4,3;$$

$$Z_4 = (x_4 + x_5 + x_6) / 3 = (4+6+8)/3 = 6;$$

$$Z_5 = (x_5 + x_6 + x_7) / 3 = (6+8+5)/3 = 6,5;$$

$$Z_6 = (x_6 + x_7 + x_8) / 3 = (8+5+5)/3 = 6;$$

$$Z_7 = (x_7 + x_8 + x_9) / 3 = (5+5+9)/3 = 6,3;$$

$$Z_8 = (x_8 + x_9 + x_{10}) / 3 = (5+9+11)/3 = 8,3;$$

$$Z_9 = (x_9 + x_{10} + x_{11}) / 3 = (9+11+7)/3 = 9 \text{ – девятая скользящая средняя;}$$

$$Z_{10} = (x_{10} + x_{11} + x_{12}) / 3 = (11+7+8)/3 = 8,7 \text{ – десятая скользящая средняя.}$$

Формируем два ряда данных:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	t_i	1	2	3	4	5	6	7	8	9	10	11	12
2	x_i	2	4	3	4	6	8	5	5	9	11	7	8
3	Z_i	3	3,7	4,3	6	6,5	6	6,3	8,3	9	8,7		

Выбрав диапазон данных, построим графики $x = x(t)$ и $Z = z(t)$ (рисунок 3).



Рисунок 3 – Графики рядов

1.3. Метод медианного сглаживания

При наличии редких выбросов (рисунок 3.б) удобнее применять метод медианного сглаживания, в котором на «скользящем» интервале τ для получения $\bar{x}(t)$ используется не среднее значение функции, а медиана.

Пусть на интервал τ попало четыре значения временного ряда $y_i(t_i)$, и одно из значений y_i сильно отличается от других (рисунок 4, точка 3).

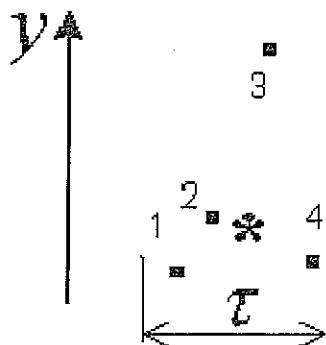


Рисунок 4 – интервал временного ряда

Построение медианы по y предполагает, что медианный центр, обозначенный звездочкой, будет находиться в области точек 1, 2 и 4, так как выше и ниже его должны находиться по две точки. Таким образом, выброс в точке 3 на сглаженной зависимости будет устранен.

Необходимо иметь в виду, что зависимости $\bar{x}_1(t)$ и $\bar{x}_2(t)$, полученные методом скользящего среднего и методом медианного сглаживания отличаются. Их отличие возрастает при увеличении частоты возникновения аномальных выбросов. Если выбросы возникают в анализируемой временной зависимости достаточно часто, то они могут рассматриваться как неотъемлемая характеристика флуктуационной составляющей и их устранение при сглаживании искажает адекватное описание случайного ряда.

1.4. Экспоненциальное сглаживание

Экспоненциальное сглаживание – простой и эффективный метод усреднения по времени временной ряд для оценки его основных статистик. Модель экспоненциальной скользящей средней (ЕМА-Exponential Moving Average) имеет вид линейного разностного уравнения первого порядка.

$$y(n) = \alpha X(n) + (1 - \alpha)y(n - 1);$$

$0 \leq \alpha \leq 1$ - параметр сглаживания,
 $X(n)$ -текущий отсчет ряда,
 $y(n)$ -модель тренда,

$$y(n) = \alpha \sum_k^{K-1} (1 - \alpha)^k x(n - k) + (1 - \alpha)^K y(n - K), n \geq K$$

$y(n - K)$ - начальное значение ЕМА, $K \rightarrow \infty$

$$y(n) = \sum_{k=0}^{\infty} \mathbb{E}_{\text{МА}}(k) x(n - k)$$

$$\mathbb{E}_{\text{МА}}(n) = \begin{cases} \alpha(1 - \alpha)^n, n \geq 0 \\ 0, n < 0 \end{cases}$$

$$\min_{y(n)} \{ \alpha(x(n) - y(n))^2 + (1 - \alpha)(y(n) - y(n - 1))^2 \}$$

Смысл формулы – в каждом классе с заданным квадратом ошибки выбрать наиболее гладкую. Первая разность характеризует гладкость модели. Максимально гладкая модель – уникальная возможность выбрать тренд.

2. Задание на лабораторную работу

2.1. Изучить предлагаемые методы анализа рядов.

2.2. В соответствии с вариантом задания, определенным преподавателем, сделайте прогноз курса валют на завтра.

В качестве базы расчета используйте данные Центрального банка РФ за предыдущие две недели. Значения аппроксимирующей функции определите по формулам, приведенным ниже.

– Метод среднего абсолютного отклонения

- 1) $F_t = X_{t-1} + A_{t-1}, t = 3, \dots, T + 1$
- 2) $A_t = \frac{C_t}{t-1}$
- 3) $C_t = C_{t-1} + X_t - X_{t-1}$
- 4) $C_1 = 0$
- 5) $X_{T+1}^{\text{прогноз}} = F_{T+1}$

– Метод скользящей средней

$$MA_t = \frac{\sum_{k=t-N}^{t-1} X_k}{N}; t = N + 1, \dots, T + 1$$

$$X_{T+1} = MA_{T+1}$$

$$\bar{Y}_{N_1} = \frac{Y_1 + Y_2 + \dots + Y_N}{N}$$

$$\bar{Y}_{N_2} = \frac{Y_1 + Y_2 + \dots + Y_{N+1}}{N}$$

.....

$$\bar{Y}_{N_i} = \frac{Y_1 + Y_2 + \dots + Y_{N+i-1}}{N}$$

где N – интервал усреднения, порядок скользящей средней,

Y_i – уровни временного ряда,

\bar{Y}_{N_i} – скользящая средняя N порядка

Замечание. При выборе интервала сглаживания необходимо иметь в виду, что обычно вычисленное среднее \bar{Y}_{N_i} относится к середине интервала сглаживания, который определяется по формуле: $m = (N + 1)/2$

При N четном среднее относится к серединам промежутков между двумя уровнями эмпирического ряда, что затрудняет сопоставление фактических уровней с их «сглаженными» значениями. Тогда, для сопоставления прибегают к центрированию сглаженного ряда, при этом центрированная скользящая средняя определяется как средняя двух рядом стоящих значений скользящей средней:

$$\bar{Y}_1^* = \frac{\bar{Y}_1 + \bar{Y}_2}{2}$$

При использовании скользящей средней для прогнозных расчетов значение скользящей средней выносят на шаг вперед или даже на несколько шагов вперед.

– Метод экспоненциального сглаживания

$$S_t = \alpha \cdot x_{t-1} + (1 - \alpha)S_{t-1}, t = 3, \dots, T + 1$$

$$S_2 = x_1$$

$$x_{T+1}^{\text{прогноз}} = S_{T+1}$$

Замечание. Прогноз вычисляется как среднее взвешенное всех элементов временного ряда. Последний элемент «взвешивается» с множителем α из интервала $(0, 1)$, предпоследний с множителем $(1 - \alpha)\alpha$ и т.д., первый с множителем $(1 - \alpha)^{T-1}$, где T – количество элементов ряда. Сумма всех весовых коэффициентов равна 1. Коэффициент α подбирается так, чтобы средняя квадратическая ошибка была минимальной. При построении прогноза учитывается, что первые уровни ряда менее значимы, для тенденции, а последние – более значимы, им придается больший вес, а первым уровням меньший.

T	x_t	S_t
1	x_1	-
2	x_2	$S_2 = x_1$
3	x_3	$S_3 = \alpha x_2 + (1 - \alpha)x_1$
4	x_4	$S_4 = \alpha x_3 + (1 - \alpha)\alpha x_2 + (1 - \alpha)^2 x_1$
5	x_5	$S_5 = \alpha x_4 + (1 - \alpha)\alpha x_3 + (1 - \alpha)^2 \alpha x_2 + (1 - \alpha)^3 x_1$
6	x_6	$S_6 = \alpha x_5 + (1 - \alpha)\alpha x_4 + (1 - \alpha)^2 \alpha x_3 + (1 - \alpha)^3 \alpha x_2 + (1 - \alpha)^4 \alpha x_1$
...
N	x_n	$S_{10} = \alpha x_9 + (1 - \alpha)\alpha x_8 + \dots + (1 - \alpha)^8 x_1$
		$S_{11} = \alpha x_{10} + (1 - \alpha)\alpha x_9 + (1 - \alpha)^2 \alpha x_8 + \dots + (1 - \alpha)^8 \alpha x_2 + (1 - \alpha)^9 \alpha x_1$

Проверить качество аппроксимации, рассчитав следующие показатели:

- показатель средней относительной погрешности,
- показатель средней квадратической погрешности,
- коэффициент несовпадения.

Варианты индивидуальных заданий

№ варианта	Валюта
1	Австралийский доллар
2	Азербайджанский манат
3	Армянский драм
4	Белорусский рубль
5	Болгарский лев
6	Бразильский реал
7	Венгерский форинт
8	Вон Республики Корея
9	Датская крона
10	Доллар США
11	Евро
12	Индийская рупия
13	Казахский тенге
14	Канадский доллар
15	Киргизский сом
16	Китайский юань
17	Польский злотый
18	Украинская гривна
19	Фунт стерлингов Соединенного королевства
20	Чешская крона
21	Японская иена

3. Вопросы к лабораторной работе

- 1) Какова сущность методов анализа временных рядов
- 2) Раскройте метод скользящей средней
- 3) В чём заключается метод медианного сглаживания
- 4) В чём заключается метод экспоненциального сглаживания
- 5) Сформулируйте задачу оптимизации в методе медианного сглаживания
- 6) Сформулируйте задачу оптимизации в методе экспоненциального сглаживания

Список литературы

1. Гасников, А.В. Современные численные методы оптимизации. Метод универсального градиентного спуска: учебное пособие / А.В. Гасников. – М.: МФТИ, 2018. – 291с. – ISBN 978-5-7417-0667-1
2. Аббасов, М.Э. Методы оптимизации: учебное пособие / М.Э. Аббасов. – СПб.: Издательство «ВВМ», 2014. – 64с. ISBN 978-5-9651-0875-6
3. Пантелеев, А.В. Методы оптимизации. Практический курс: учебное пособие / А.В. Пантелеев. – Москва: Логос, 2020. – 424с. – ISBN 978-5-98704-540-4. – Текст: электронный – Режим доступа: <https://e.lanbook.com/book/163062>
4. Мицель, А.А. Методы оптимизации: учебное пособие / А.А. Мицель, А.А. Шелестов, В.В. Романенко. – Томск: ФДО, ТУСУР, 2017. – 198с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

Методы интеллектуального анализа естественного языка

Профиль подготовки

09.04.03 – Прикладная информатика

(Искусственный интеллект и бизнес-аналитика)

Квалификация выпускника

Магистр

Формы обучения

очная

г. Ульяновск, 2021

Содержание

1. Методические указания для лабораторных работ
2. Вопросы для промежуточной аттестации

1. Методические указания для лабораторных работ

Лабораторная работа 1

Парсинг текстов. Базовая обработка текстов на Python. Регулярные выражения.

Методические рекомендации:

1. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
2. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
3. Лабораторная работа состоит из 3-х заданий, каждое по 6 вариантов.
4. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции №1.

Строка представляет последовательность символов в кодировке Unicode. И мы можем обратиться к отдельным символам строки по индексу в квадратных скобках:

```
string = "hello world"
c0 = string[0] # h
print(c0)
c6 = string[6] # w
print(c6)

c11 = string[11] # ошибка IndexError: string index out of range
print(c11)
```

Индексация начинается с нуля, поэтому первый символ строки будет иметь индекс 0. А если мы попытаемся обратиться к индексу, которого нет в строке, то мы получим исключение `IndexError`. Например, в случае выше длина строки 11 символов, поэтому ее символы будут иметь индексы от 0 до 10.

Чтобы получить доступ к символам, начиная с конца строки, можно использовать отрицательные индексы. Так, индекс -1 будет представлять последний символ, а -2 - предпоследний символ и так далее:

```
string = "hello world"
c1 = string[-1] # d
print(c1)
c5 = string[-5] # w
print(c5)
```


При работе с символами следует учитывать, что строка - это неизменяемый (immutable) тип, поэтому если мы попробуем изменить какой-то отдельный символ строки, то мы получим ошибку.

При необходимости мы можем получить из строки не только отдельные символы, но и подстроку. Для этого используется следующий синтаксис:

- `string[:end]`: извлекается последовательность символов начиная с 0-го индекса по индекс `end`
- `string[start:end]`: извлекается последовательность символов начиная с индекса `start` по индекс `end`
- `string[start:end:step]`: извлекается последовательность символов начиная с индекса `start` по индекс `end` через шаг `step`

Используем все варианты получения подстроки:

```
string = "hello world"

# с 0 до 5 символа
sub_string1 = string[:5]
print(sub_string1) # hello

# со 2 до 5 символа
sub_string2 = string[2:5]
print(sub_string2) # llo

# со 2 по 9 символ через один символ
sub_string3 = string[2:9:2]
print(sub_string3) # lowr
```

Поскольку строка содержит символы Unicode, то с помощью функции `ord()` мы можем получить числовое значение для символа в кодировке Unicode:

```
print(ord("A")) # 65
```

Для получения длины строки можно использовать функцию `len()`:

```
string = "hello world"
length = len(string)
print(length) # 11
```

С помощью выражения `term in string` можно найти подстроку `term` в строке `string`. Если подстрока найдена, то выражение вернет значение `True`, иначе возвращается значение `False`:

```
string = "hello world"
```

```
exist = "hello" in string
print(exist) # True
```

```
exist = "sword" in string
print(exist) # False
```

С помощью цикла for можно перебрать все символы строки:

```
string = "hello world"
for char in string:
    print(char)
```

Рассмотрим основные методы строк, которые мы можем применить в приложениях:

- `isalpha()`: возвращает True, если строка состоит только из алфавитных символов
- `islower()`: возвращает True, если строка состоит только из символов в нижнем регистре
- `isupper()`: возвращает True, если все символы строки в верхнем регистре
- `isdigit()`: возвращает True, если все символы строки - цифры
- `isnumeric()`: возвращает True, если строка представляет собой число
- `startswith(str)`: возвращает True, если строка начинается с подстроки str
- `endswith(str)`: возвращает True, если строка заканчивается на подстроку str
- `lower()`: переводит строку в нижний регистр
- `upper()`: переводит строку в верхний регистр
- `title()`: начальные символы всех слов в строке переводятся в верхний регистр
- `capitalize()`: переводит в верхний регистр первую букву только самого первого слова строки
- `lstrip()`: удаляет начальные пробелы из строки
- `rstrip()`: удаляет конечные пробелы из строки
- `strip()`: удаляет начальные и конечные пробелы из строки
- `ljust(width)`: если длина строки меньше параметра width, то справа от строки добавляются пробелы, чтобы дополнить значение width, а сама строка выравнивается по левому краю
- `rjust(width)`: если длина строки меньше параметра width, то слева от строки добавляются пробелы, чтобы дополнить значение width, а сама строка выравнивается по правому краю
- `center(width)`: если длина строки меньше параметра width, то слева и справа от строки равномерно добавляются пробелы, чтобы дополнить значение width, а сама строка выравнивается по центру
- `find(str[, start [, end]])`: возвращает индекс подстроки в строке. Если подстрока не найдена, возвращается число -1
- `replace(old, new[, num])`: заменяет в строке одну подстроку на другую
- `split([delimiter[, num]])`: разбивает строку на подстроки в зависимости от разделителя
- `join(strs)`: объединяет строки в одну строку, вставляя между ними определенный разделитель

Для поиска подстроки в строке в Python применяется метод `find()`, который возвращает индекс первого вхождения подстроки в строку и имеет три формы:

- `find(str)`: поиск подстроки str ведется с начала строки до ее конца
- `find(str, start)`: параметр start задает начальный индекс, с которого будет производиться поиск
- `find(str, start, end)`: параметр end задает конечный индекс, до которого будет идти поиск. Если подстрока не найдена, метод возвращает -1.

Для замены в строке одной подстроки на другую применяется метод `replace()`:

- `replace(old, new)`: заменяет подстроку `old` на `new`
- `replace(old, new, num)`: параметр `num` указывает, сколько вхождений подстроки `old` надо заменить на `new`

Метод `split()` разбивает строку на список подстрок в зависимости от разделителя. В качестве разделителя может выступать любой символ или последовательность символов. Данный метод имеет следующие формы:

- `split()`: в качестве разделителя используется пробел
- `split(delimiter)`: в качестве разделителя используется `delimiter`
- `split(delimiter, num)`: параметр `num` указывает, сколько вхождений `delimiter` используется для разделения. Оставшаяся часть строки добавляется в список без разделения на подстроки

При рассмотрении простейших операций со строками было показано, как объединять строки с помощью операции сложения. Другую возможность для соединения строк представляет метод `join()`: он объединяет список строк. Причем текущая строка, у которой вызывается данный метод, используется в качестве разделителя.

Задание 1.

Используя базовые возможности языка Python, без использования сторонних библиотек выполните следующие задания.

Вариант 1.

Вашей программе на вход подаются три строки `s`, `a`, `b`, состоящие из строчных латинских букв.

За одну операцию вы можете заменить все вхождения строки `a` в строку `s` на строку `b`.

Например, `s = "abab"`, `a = "ab"`, `b = "ba"`, тогда после выполнения одной операции строка `s` перейдет в строку `"baba"`, после выполнения двух и операций – в строку `"bbaa"`, и дальнейшие операции не будут изменять строку `s`.

Необходимо узнать, после какого минимального количества операций в строке `s` не останется вхождений строки `a`. Если операций потребуется более 1000, выведите `Impossible`.

Выведите одно число – минимальное число операций, после применения которых в строке `s` не останется вхождений строки `a`, или `Impossible`, если операций потребуется более 1000.

Вариант 2.

Вашей программе на вход подаются две строки `s` и `t`, состоящие из строчных латинских букв.

Выведите одно число – количество вхождений строки `t` в строку `s`.

Пример:

`s = "abababa"`

`t = "aba"`

Результат:

Вхождения строки `t` в строку `s`:

`abababa`

`abababa`

`abababa`

Вариант 3.

Реализуйте алгоритм сжатия, который сжимает повторяющиеся символы в строке.

Кодирование осуществляется следующим образом:

`s = 'aaaabbcaa'` преобразуется в `'a4b2c1a2'`, то есть группы одинаковых символов исходной строки заменяются на этот символ и количество его повторений в этой позиции строки.

Напишите программу, которая считывает строку, кодирует её предложенным алгоритмом и выводит закодированную последовательность на стандартный вывод. Кодирование не должно учитывать регистр символов.

Вариант 4.

Напишите программу, которая считывает из файла строку, соответствующую тексту, сжатому с помощью кодирования повторов в задании варианта 3, и производит обратную операцию, получая исходный текст. Запишите полученный текст в файл.

Вариант 5.

Напишите программу, которая считывает текст из файла (в файле может быть больше одной строки) и выводит самое частое слово в этом тексте и через пробел то, сколько раз оно встретилось.

Если таких слов несколько, вывести лексикографически первое (можно использовать оператор `<` для строк).

Слова, написанные в разных регистрах, считаются одинаковыми.

Вариант 6.

Имеется файл с данными по успеваемости абитуриентов. Он представляет из себя набор строк, где в каждой строке записана следующая информация:

Фамилия;Оценка_по_математике;Оценка_по_физике;Оценка_по_русскому_языку

Поля внутри строки разделены точкой с запятой, оценки — целые числа.

Напишите программу, которая считывает исходный файл с подобной структурой и для каждого абитуриента записывает его среднюю оценку по трём предметам на отдельной строке, соответствующей этому абитуриенту, в файл с ответом.

Задание 2.

Используя регулярные выражения и библиотеку *re*, выполните следующие задания.

Вариант 1.

1. Вам дана последовательность строк. Выведите строки, содержащие "cat" в качестве подстроки хотя бы два раза.
2. Вам дана последовательность строк. В каждой строке замените все вхождения подстроки "human" на подстроку "computer" и выведите полученные строки.

Вариант 2.

1. Вам дана последовательность строк. Выведите строки, содержащие "cat" в качестве слова.
2. Вам дана последовательность строк. В каждой строке замените первое вхождение слова, состоящего только из латинских букв "a" (регистр не важен), на слово "argh".

Вариант 3.

1. Вам дана последовательность строк. Выведите строки, содержащие две буквы "z", между которыми ровно три символа.
2. Вам дана последовательность строк. В каждой строке поменяйте местами две первых буквы в каждом слове, состоящем хотя бы из двух букв.

Вариант 4.

1. Вам дана последовательность строк. Выведите строки, содержащие обратный слеш "\".
2. Вам дана последовательность строк. В каждой строке замените все вхождения нескольких одинаковых букв на одну букву.

Вариант 5.

1. Вам дана последовательность строк. Выведите строки, содержащие слово, состоящее из двух одинаковых частей (тандемный повтор).
2. Вам дана последовательность строк. Верните все даты в строках.

Вариант 6.

1. Вам дана последовательность строк. Верните первое слово из каждой строки. Верните первое слово из каждой строки.
2. Вам дана последовательность строк. В строке может быть указан(а может быть не указан) телефонный номер. Вам необходимо проверить его наличие и проверить его формат (корректный или же нет).

Задание 3.

Используя библиотеку запросов requests, а также библиотек BeautifulSoup осуществите парсинг следующих сайтов, извлеките и структурируйте требуемую информацию.

Вариант 1.

Со страницы сайта Кинопоиск <https://www.kinopoisk.ru/lists/top250/> извлеките информацию о 50-ти фильмах рейтинга, такую как: Название фильма, место, год, оценка(рейтинг).

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Вариант 2.

Со страницы сайта авто.ру https://auto.ru/cars/land_rover/all/ получите список автомобилей на странице, а также для каждого автомобиля год, пробег, цвет, тип КПП.

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Вариант 3.

Со страницы сайта Кинопоиск <https://www.kinopoisk.ru/media/news/> извлеките информацию обо всех новостях со страницы, такую как: Заголовок, ссылка на новость, ссылка на изображение, дата.

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Вариант 4.

Со страницы сайта ХОРОШИЕ НОВОСТИ ПРО ЖИВОТНЫХ <http://goodnewsanimal.ru/> извлеките информацию о обо всех новостях со страницы, такую как: Заголовок, ссылка на новость, ссылка на изображение, аннотация.

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Вариант 5.

Со страницы сайта cheb.ru <https://cheb.ru/news/> извлеките информацию о обо всех новостях со страницы, такую как: Заголовок, ссылка на новость, дата, тэги.

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Вариант 6.

Со страницы сайта Кинопоиск <https://www.kinopoisk.ru/comingsoon/> извлеките информацию о ТОП 20-ти ожидаемых фильмах, такую как: Название фильма, страна, год, рейтинг, ссылка на постер.

Данные структурируйте и отобразите в виде Pandas дата-фрейма, а затем сохраните его в .csv файл.

Лабораторная работа № 2

Базовые подходы к преобразованию и обработке текстов.

Методические рекомендации:

5. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
6. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
7. Лабораторная работа состоит из 1 задания, каждое по 6 вариантов.
8. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции №5.

Некоторые теоретические сведения.

Пример очистки текста от мусорных символов с помощью регулярных выражений.

```
import re

# sample review from the IMDB dataset.
review = "<b>A touching movie!!</b> It is full of emotions and wonderful acting.<br> I could have sat through it a second time."

cleaned_review = re.sub(re.compile('<.*?>'), '', review) #removing html tags
cleaned_review = re.sub('[^A-Za-z0-9]+', '', cleaned_review) #taking only words

print(cleaned_review)
```

A touching movie It is full of emotions and wonderful acting I could have sat through it a second time

Примеры удаления стоп-слов , лемматизации и стемминга:

```
# nltk.download('stopwords') # you will have to download the set of stop words the first time
from nltk.corpus import stopwords

stop_words = stopwords.words('english')

filtered_review = [word for word in tokens if word not in stop_words] #removing stop words.

print(filtered_review)
```

['touching', 'movie', 'full', 'emotions', 'wonderful', 'acting', 'could', 'sat', 'second', 'time']

```
from nltk.stem import PorterStemmer

stemmer = PorterStemmer()

stemmed_review = [stemmer.stem(word) for word in filtered_review]

print(stemmed_review)
```

['touch', 'movi', 'full', 'emot', 'wonder', 'act', 'could', 'sat', 'second', 'time']

```
# nltk.download('wordnet')
from nltk.stem import WordNetLemmatizer

lemmatizer = WordNetLemmatizer()

lemm_review = [lemmatizer.lemmatize(word) for word in filtered_review]

print(lemm_review)
```

['touching', 'movie', 'full', 'emotion', 'wonderful', 'acting', 'could', 'sat', 'second', 'time']

Пример токенизации текста и создания словаря частотности:


```
sample_text = ['один два три четыре пять два Три три четыре четыре четыре пять пять пять пять шесть семь восемь']
maxWordsCount = 5
tokenizer = Tokenizer(num_words=maxWordsCount, filters='!#$%&()*+,-./:;<=>@[\\]^_`{|}~«»\t\n\x0b\x0f\xff', lower=True, split=' ', oov_token='unknown', char_level=False)
tokenizer.fit_on_texts(sample_text)
```

```
tokenizer.word_index
```

```
{'unknown': 1,
 'восемь': 9,
 'два': 5,
 'один': 6,
 'пять': 2,
 'семь': 8,
 'три': 4,
 'четыре': 3,
 'шесть': 7}
```

Для создания мешка слов вы можете воспользоваться токенайзером библиотеки Keras, приведенной выше. Для это нужно воспользоваться соответствующим методом **texts_to_matrix**.

https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/text/Tokenizer

Кроме того, весьма полезная библиотека `py morphology` для работы с формами слов.

```
!pip install -q pymorphy2
import pymorphy2
import nltk
from nltk.stem import WordNetLemmatizer
nltk.download('wordnet')
```

Для работы с морфемами необходимо создать объект:

```
morph = pymorphy2.MorphAnalyzer()
```

Следующим шагом создается список из слов. В данном примере слова разделены пробелами. В зависимости от вашей задачи единицей текста могут быть как буквы, так и фразы:

```
words = text.split(' ')
```

Приводим каждое слово в начальную форму:

```
docs = [morph.parse(word)[0].normal_form for word in words]
```

Получаем список из слов предложения(фразы) в нормальной форме.

Развитие сетевых технологий и, не в последнюю очередь появление социальных сетей, дало возможность пользователям явно выражать свое мнение относительно каче-

ства услуг, товаров, фильмов, книг, результатах деятельности компаний или политических деятелей.

Наличие подобной информации привело к необходимости ее анализа с точки зрения установления обратной связи между поставщиками услуг и их потребителями. При этом, количество отзывов может достигать десятков тысяч, и обработка отзывов вручную оказывается невозможной. В связи с этим широкое распространение получили автоматизированные подходы к анализу тональности текстов (*sentiment analysis*).

Однако, непосредственно перед анализом необходимо проводить процедуру отсеивания имеющихся отзывов с точки зрения выявления их дубликатов.

Особенно этим «грешат» форумы на сайтах компаний, являющихся поставщиками таких услуг. Не является секретом то, что часть отзывов (особенно положительных) пишут менеджеры данных компаний. Но, еще чаще для их написания привлекаются специальные люди – копирайтеры. В интернет имеется множество публикаций на тему как правильно писать подобные отзывы.

Задачей настоящей работы является освоение технологии отсеивания «заказных» отзывов.

Для решения выше указанной задачи мало подходят стандартные методы классификации текстов (типа *bag of words*) прежде всего из-за небольших объемов сравниваемых текстов – максимум 1-2 стр. формата А4. Получающиеся при этом частотные словари являются статистически неустойчивыми и, как следствие, результаты классификации являются не надежными.

Для характеристики смысла текста вводится понятие семантического спектра. Данный спектр представляет собой своеобразную хэш-функцию заданной последовательности символов.

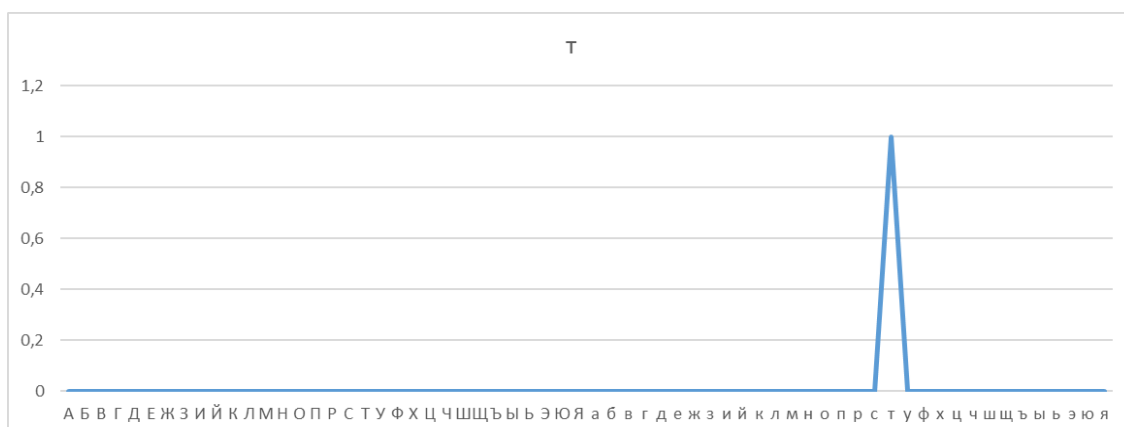
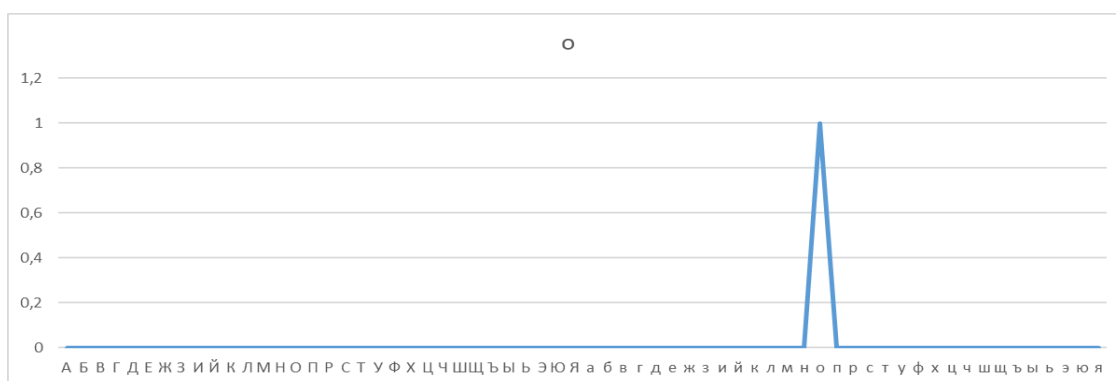
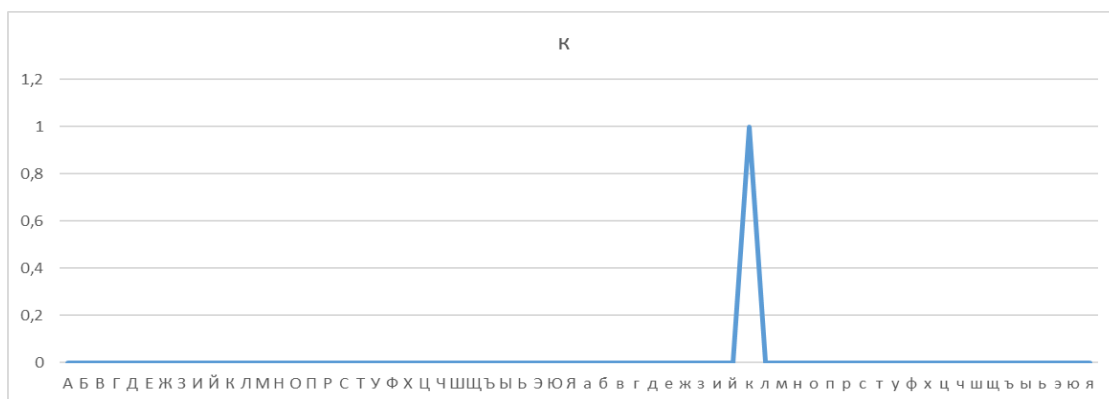
Указанная функция формируется следующим образом:

- задается набор символов, которые ожидаются в обрабатываемом тексте. Если это русскоязычные тексты, то в этот набор можно включить все строчные и прописные русские символы, цифровые символы и символы-разделители (точка, запятая, двоеточие и т.д.);

- выбранный набор символов упорядочивается. Упорядочивание наиболее естественно произвести в порядке возрастания ASCII-кодов символов;

- для каждого символа формируется стартовое значение семантического спектра. Формирование спектра производится по правилу: если номер элемента спектра равен номеру данного символа, то этот элемент спектра приравнивается 1. В противном случае элемент спектра равен 0.

Например. На рис. 5-7 показаны спектры букв «к», «о» и «т».

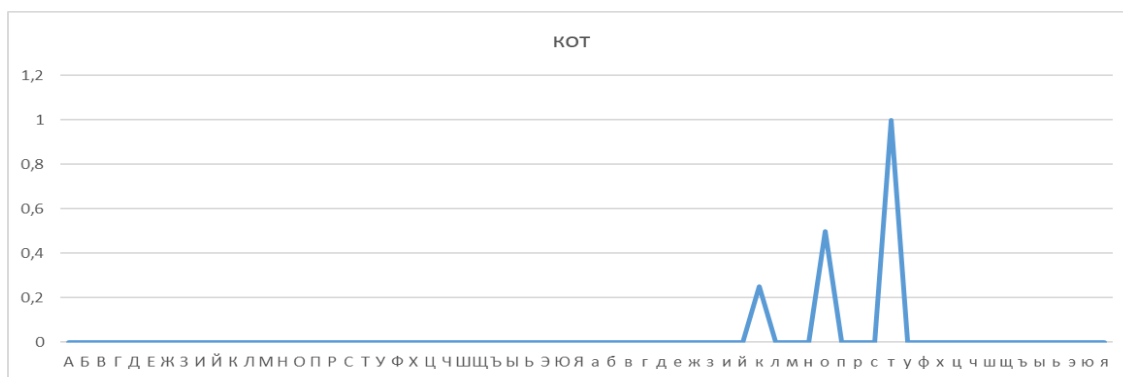


Для получения спектра последовательности символов (текста) выполняются следующие операции:

- задается начальный спектр текста, все элементы которого равны 0;
- поочередно берутся символы текста и их спектры поэлементно совмещаются с соответствующими элементами спектра текста по формуле:

$$S_{\text{текста}} = (S_{\text{текста}} + S_{\text{символа}})/2 \quad (1)$$

На рис. 8 показан полученный спектр для последовательности символов «кот».



Особенности формулы (1) приводят к следующим эффектам:

1. Получаемый спектр позволяет учитывать не только вхождение символа в последовательность, но и его порядок в этой последовательности.
2. Чем раньше символ появляется в последовательности, тем меньше его значение в спектре. Это приводит к тому, что для очень длинных последовательностей символов «сигнал» от начальных символов станет практически незаметным.

Наличие второго эффекта, в частности, означает, что нельзя механически применять уравнение (3) для очень длинных последовательностей символов.

Чтобы не потерять информацию о начальных символах длина обрабатываемой последовательности ограничивается длиной очередного слова в тексте. Типичная длина русских слов (5-6 символов) позволяет сохранять информацию о начальных символах слова в получающемся спектре. Получившийся спектр слова можно затем по той же формуле (3) объединять с начальным (нулевым) семантическим спектром текста.

Однако, нельзя и весь текст сводить к набору слов, поскольку для большого текста вклад спектров начальных слов в общую семантику текста будет постепенно нивелироваться. Чтобы этого не произошло, производится формирование промежуточных семантических спектров предложений. Для русского языка средняя длина простых предложений составляет 5-6 слов. Такой длины еще вполне достаточно, чтобы не потерять информацию о семантике начальных слов.

Если продолжить аналогичные рассуждения дальше, то просматривается следующая иерархическая последовательность получения семантики текста с помощью формулы (3):

стартовая семантика символов > семантика слов > семантика простых предложений > семантика составных предложений > семантика абзацев > семантика параграфов > семантика глав (разделов) > семантика текста.

Алгоритм обработки

1. Обрабатываемый текст просматривается и все входящие в него слова заменяются базовыми словоформами.

Например:

Исходная фраза: «часто проводятся акции на разные товары»

преобразуется к виду: «часто проводить акция на разный товар»

2. Базовые словоформы заменяются условным базовым синонимом.

Например:

вышеприведенная фраза: «часто проводить акция на разный товар»
 преобразуется к виду: «частый проводить поступок на отличный товар»

3. Формируется семантический спектр текста
4. С помощью шагов 1 – 3 формируются спектры других сравниваемых текстов.
5. Полученные спектры попарно сравниваются по некоторому критерию. В данной работе в качестве меры сходства семантики текстов используется косинусное сходство:

$$R = \text{ArcCos}(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

где n – размер спектра;
 x_i - i -ый элемент спектра в первом тексте;
 y_i - i -ый элемент спектра во втором тексте.

Значение R интерпретируется как некоторый угол между векторами X и Y в многомерном пространстве. Если вектора совпадают, угол между ними равен 0 и значение R будет равно 1 . Если вектора не совпадают, то максимальный угол между ними будет равен 90 градусов и значение R будет равно 0 .

Задание 1.

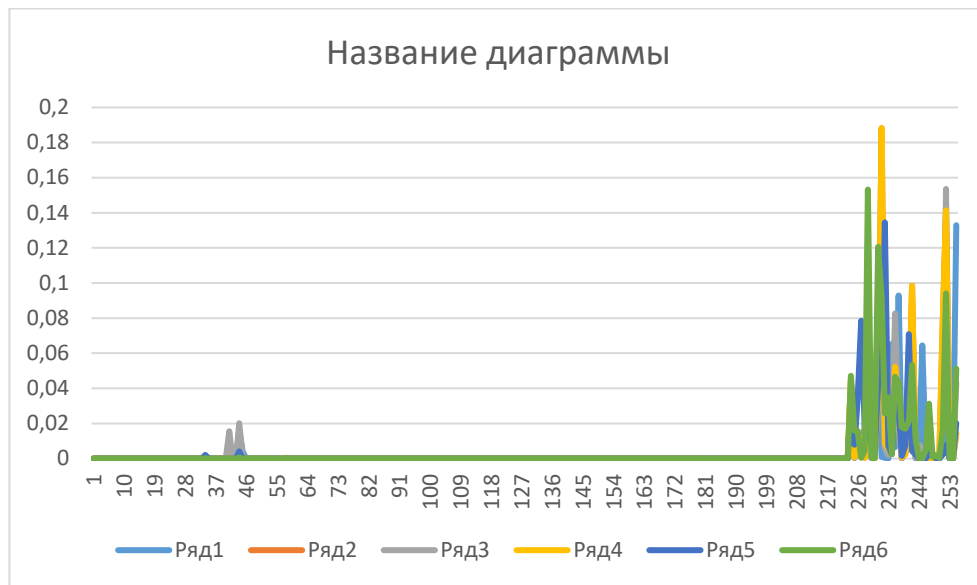
В папке, соответствующей вашему варианту находятся образцы текстов (отзывы различных сайтов).

- Считайте текстовые файлы, создайте список текстов
- Произведите предобработку текстовых данных, параллельно сопровождая ваши действия комментариями в ячейках Markdown в Jupyter ноутбуке. Объясните ваши действия, почему вы произвели именно эти указанные преобразования текстов.
- Рассчитайте спектр текстов, постройте матрицу взаимного косинусного сходства текстов, разместив в матрице коэффициенты R .
- Сформулируйте критерий определения порога близости двух текстов.
- Сформулируйте критерий определения «фейкового» отзыва.

Матрица косинусного сходства может иметь следующий вид:

	text1	text2	text3	text4	text5	text6
text1	1					
text2	0,444665	1				
text3	0,602006	0,7841	1			
text4	0,484063	0,6223	0,642163	1		
text5	0,288512	0,385	0,36231	0,51119	1	
text6	0,631552	0,7467	0,737299	0,67563	0,47	1

Примеры спектров отзывов на одной диаграмме:



- Отобразите наиболее близкие тексты согласно косинусному сходству спектров, и наиболее «дальние» тексты. Сформулируйте выводы по итогам работы.

Лабораторная работа 3

Введение в нейронные сети. Обучение нейронной сети. Переобучение. Метрики качества

Методические рекомендации:

9. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
10. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
11. Лабораторная работа состоит из 1 задания, 4 варианта.
12. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции №3 и 4 по введению в нейронные сети.

Градиент (от лат. *gradiens*, род. п. *gradientis* «шагающий, растущий») — вектор, своим направлением указывающий направление наибольшего возрастания некоторой скалярной величины φ , (значение которой меняется от одной точки пространства к другой, образуя скалярное поле), а по величине (модулю) равный скорости роста этой величины в этом направлении.

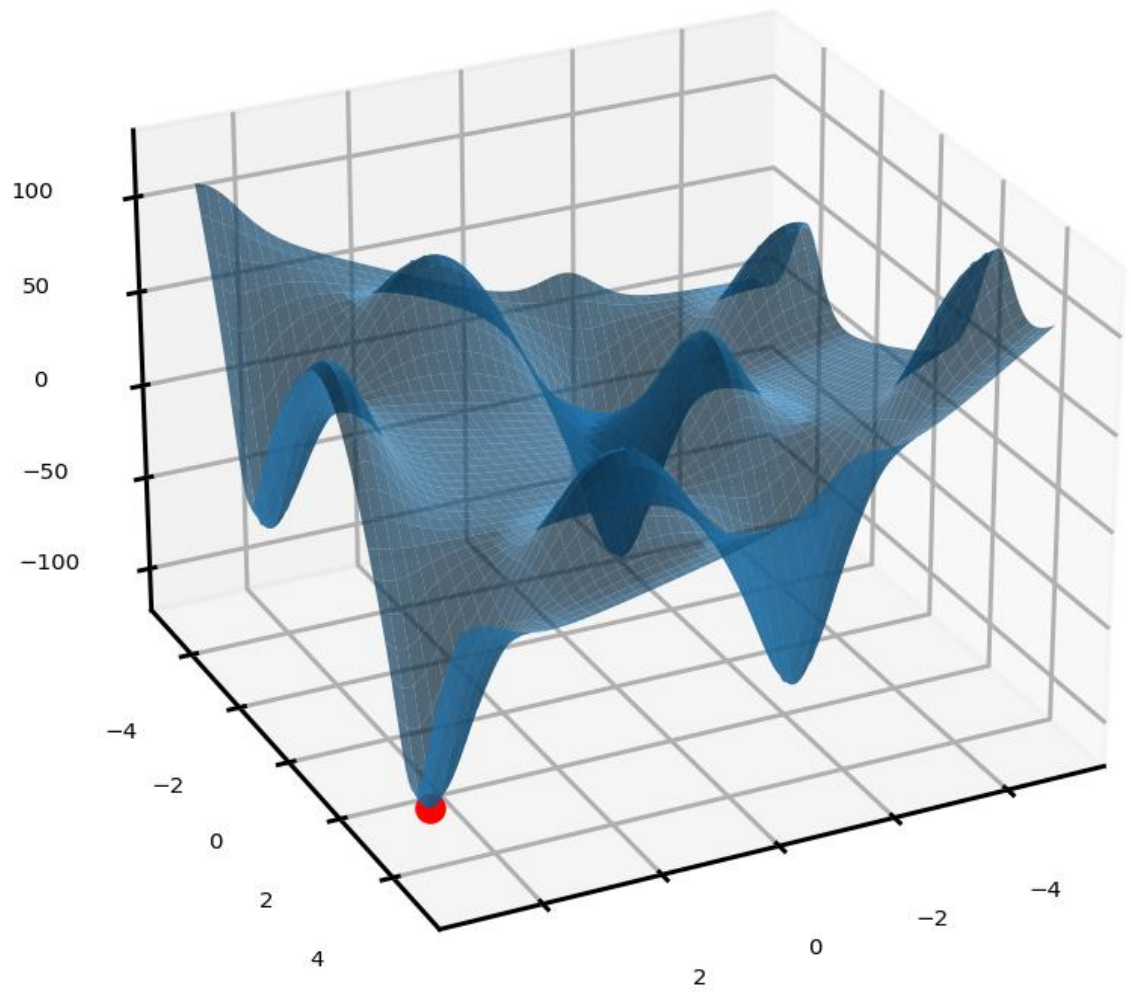
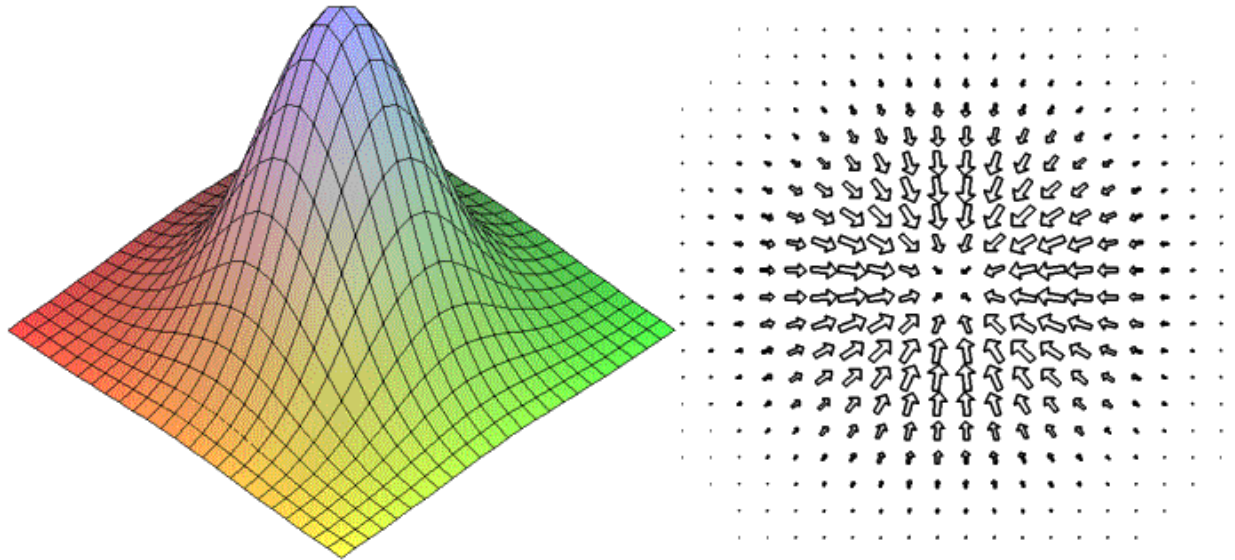
Например, если взять в качестве φ высоту поверхности земли над уровнем моря, то её градиент в каждой точке поверхности будет показывать «направление самого крутого подъёма», и своей величиной характеризовать крутизну склона.

Другими словами, градиент — это производная по пространству, но в отличие от производной по одномерному времени, градиент является не скаляром, а векторной величиной.

С математической точки зрения на градиент можно смотреть как на:

- Коэффициент линейности изменения значения функции многих переменных от изменения значения аргумента;
- Вектор в пространстве области определения скалярной функции многих переменных, составленный из частных производных;
- Строки матрицы Якоби содержат градиенты составных скалярных функций из которых состоит векторная функция многих переменных.

Пространство, на котором определена функция и её градиент, может быть, вообще говоря, как обычным трёхмерным пространством, так и пространством любой другой размерности любой физической природы или чисто абстрактным (безразмерным).



Задание 1.

Смоделируйте оптимизацию параметров некоей модели, согласно своему варианту, методом градиентного спуска, используя материалы лекции.

```
# Создаем класс своей модели
class Model(object):
    # Переопределяем методы инициализации
    def __init__(self):
        self.W = tf.Variable(5.0) # Создаем переменную класса W со стартовым значением 5
        self.b = tf.Variable(10.0) # Создаем переменную класса b со стартовым значением 10

    # Переопределяем метод __call__
    def __call__(self, input):
        return self.W * input + self.b # Метод возвращает результат wx+b

model = Model() # Создаем объект созданного класса Model()
```

Для создания класса модели рекомендуется использовать фреймворк Tensorflow, поскольку он лежит в основе фреймворка Keras. Данный фреймворк позволяет создавать тензоры, а также обладает богатым функционалом и широкими возможностями, например удобными средствами расчета градиентов и реализации методов градиентного спуска.

```
import tensorflow as tf
```

Вы можете ознакомиться с официальной документацией по ссылке <https://www.tensorflow.org/>.

Поскольку наша модель будет решать задачу регрессии, используем функцию ошибки MSE(среднеквадратичное отклонение).

```
def loss(target_y, predicted_y):
    return tf.reduce_mean(tf.square(target_y - predicted_y))
```

Функция ошибки принимает первым аргументом истинный вектор целевой переменной, вторым аргументом – предсказанный вектор, возвращая среднеквадратичную ошибку.

tf.square - возведение в вектора в степень 2.

tf.reduce_mean – вычисление среднего значения вектора.

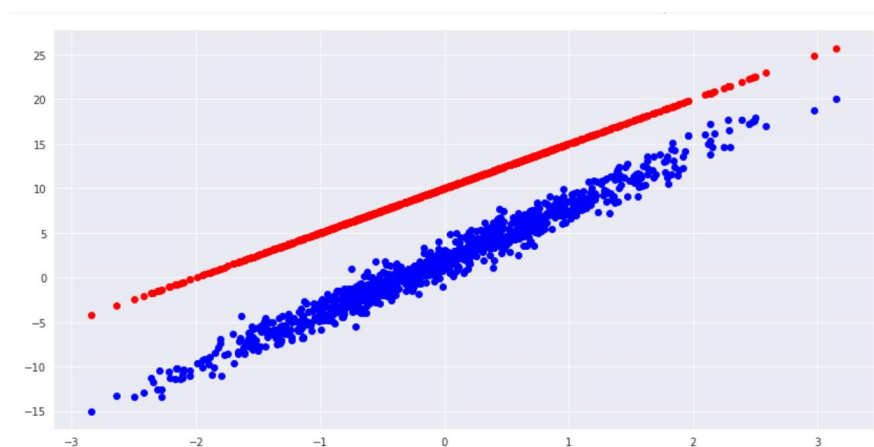
Следующим шагом задаются истинные параметры распределения:

```
TRUE_W = 6.0 # Задаем значение, к которому должен стремиться параметр W
TRUE_b = 2.0 # Задаем значение, к которому должен стремиться параметр b
NUM_EXAMPLES = 1000 # Количество элементов в списке x
```

Моделируем нормальное распределение точек (inputs), в количестве 1000 шт, рассчитываем истинные значения вектора y, добавив к нему шум(приближение к реальным условиям).

```
inputs = tf.random.normal(shape=[NUM_EXAMPLES]) # Задаем 1000 значение x
noise = tf.random.normal(shape=[NUM_EXAMPLES]) # Задаем 1000 значений для шума-смещения
outputs = inputs * TRUE_W + TRUE_b + noise # Получаем 1000 значений y_true (правильных ответов)
```

Отобразив реальное облако точек и аппроксимирующую функцию с начальными параметрами, получим график следующего вида:



Таким образом, необходимо подобрать параметры прямой, чтобы она максимально точно описывала облако точек с минимальной ошибкой MSE.

Функция, обучающая модель (подгоняющая веса модели) будет выглядеть следующим образом:

```
# Создаем функцию обучения модели
def train(model, inputs, outputs, learning_rate):
    # Создаем ленту GradientTape для автоматического расчета градиентов
    with tf.GradientTape() as t:
        current_loss = loss(outputs, model(inputs)) # Получаем значение ошибки
        dw, db = t.gradient(current_loss, [model.W, model.b]) # Считаем градиенты для параметров W и b
        model.W.assign_sub(learning_rate * dw) # Обновляем значение W
        model.b.assign_sub(learning_rate * db) # Обновляем значение b
```

`tf.GradientTape` осуществляет расчет градиентов. Метод `t.gradient` принимает первым аргументом функцию ошибки, вторым аргументом список параметров, по которым будут рассчитываться частные производные функции ошибки.

Метод `.assign_sub` позволяет уменьшить значения параметров (переменных `tensorflow`) с определенным шагом `learning_rate`, который является фактически скоростью обучения модели. Регулируя его значения мы можем ускорять или сокращать скорость движения к экстремумам оптимизационной функции.

1. Создайте функцию `model_fit()`, которая будет обучать модель по эпохам, с выводом промежуточной информации в консоль. Ваша функция должна выводить номер эпохи, значение параметра `w`, `b`, функцию ошибки `mse`. А также сохранять в словарь `history` под ключом `train_loss` список ошибок MSE на каждой эпохе.
2. Разбейте выборку точек на 2 подвыборки – тренировочную и валидационную. Причем доля валидационной выборки должна составлять не менее 30% от общего количества. Модифицируйте функцию обучения таким образом, чтобы она выводила на печать вдобавок ошибку MSE на валидационной выборке под названием `val-loss`.
3. Модифицируете функцию `model_fit()` таким образом, чтобы по окончании процесса обучения она автоматически на 1 графике изображала зависимость `train_loss` от номера эпохи и зависимость `val_loss` от номера эпохи. Необходимо оформить график по всем правилам: заголовок, подписи осей, координатная сетки, легенда графика с подписями. Посмотрите, наблюдаете ли вы переобучение модели?

4. Создайте функцию (либо класс) `auto_fitting()`, который будет принимать 2 аргумента: вектор X , вектор Y , которые являются координатами точек. Данная функция(либо класс) должна возвращать обученную модель, аппроксимирующую данное облако точек согласно вашему варианту. Подумайте, какие еще аргументы необходимо передать функции, чтобы максимально автоматизировать процесс обучения модели. Добавьте возможность автоматической генерации начальных значений весовых коэффициентов модели.
5. Добавьте к разработанной функции `auto_fitting()` аргумент, позволяющий задавать предел обучения модели в виде порогового значения ошибки на тренировочной (либо на тестовой) выборке, а также максимальное количество эпох. Действительно, мы заранее не можем знать достигнет ли функция заданной точности заранее, поэтому следует ее ограничивать каким-либо конечным числом эпох.
6. Смоделируйте и постройте зависимость величины ошибки на тестовой выборке от доли искусственного шума (noise)

Каждому варианту соответствует своя функция для аппроксимации:

Вариант 1.

$$Y = 6 * x^2$$

Вариант 2.

$$Y = -3 * x^3$$

Вариант 3.

$$Y = 2 * \sin(0.5 * x)$$

Вариант 4.

$$Y = -5 * \sin(0.5 * x)$$

В лабораторной работе использовались материалы из следующих источников:

1. Numpy градиентный. спусковой оптимизатор нейронных сетей. <https://pythobyte.com/numpy-gradient-75571/>
2. Numpy gradient. <https://numpy.org/doc/stable/reference/generated/numpy.gradient.html>
3. Градиент <https://ru.wikipedia.org/wiki/%D0%93%D1%80%D0%B0%D0%B4%D0%B8%D0%B5%D0%BD%D1%82>
4. Градиентный спуск в Python. <https://habr.com/ru/post/547424/>

Лабораторная работа 4

Полносвязные нейронные сети.

Методические рекомендации:

13. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
14. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
15. Лабораторная работа состоит из 2 заданий, 6 вариантов.
16. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции №3 и 4 по введению в нейронные сети.

В лабораторной работе вам будет необходимо спроектировать и обучить модели для задач классификации и регрессии.

Данные для подачи в нейронную сеть должны быть соответствующим образом преобразованы. Матрица предикторов, будем называть ее матрицей X должна быть в формате numpy числового массива, все категориальные признаки должны быть представлены сначала в числовом формате, а затем в формате One Hot Encoding. Процесс кодирования описан на следующей схеме:

Label Encoding			One Hot Encoding			
Food Name	Categorical #	Calories	Apple	Chicken	Broccoli	Calories
Apple	1	95	1	0	0	95
Chicken	2	231	0	1	0	231
Broccoli	3	50	0	0	1	50

Для осуществления Label Encoding рекомендуется использовать кодировщик библиотеки Scikit Learn.

<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>

Examples

`LabelEncoder` can be used to normalize labels.

```
>>> from sklearn import preprocessing
>>> le = preprocessing.LabelEncoder()
>>> le.fit([1, 2, 2, 6])
LabelEncoder()
>>> le.classes_
array([1, 2, 6])
>>> le.transform([1, 1, 2, 6])
array([0, 0, 1, 2]...)
>>> le.inverse_transform([0, 0, 1, 2])
array([1, 1, 2, 6])
```

It can also be used to transform non-numerical labels (as long as they are hashable and comparable) to numerical labels.

```
>>> le = preprocessing.LabelEncoder()
>>> le.fit(["paris", "paris", "tokyo", "amsterdam"])
LabelEncoder()
>>> list(le.classes_)
['amsterdam', 'paris', 'tokyo']
>>> le.transform(["tokyo", "tokyo", "paris"])
array([2, 2, 1]...)
>>> list(le.inverse_transform([2, 2, 1]))
['tokyo', 'tokyo', 'paris']
```

Methods

<code>fit(y)</code>	Fit label encoder.
<code>fit_transform(y)</code>	Fit label encoder and return encoded labels.
<code>get_params([deep])</code>	Get parameters for this estimator.
<code>inverse_transform(y)</code>	Transform labels back to original encoding.
<code>set_params(**params)</code>	Set the parameters of this estimator.
<code>transform(y)</code>	Transform labels to normalized encoding.

Для кодирования лейблов в формат ONE можно использовать процедуру библиотеки Keras из модуля `utils`.

https://keras.io/api/utils/python_utils/#to_categorical-function

```
1 import numpy as np
2 from keras.utils import to_categorical
3 ### Categorical data to be converted to numeric data
4 colors = ["red", "green", "yellow", "red", "blue"]
5
6 ### Universal list of colors
7 total_colors = ["red", "green", "blue", "black", "yellow"]
8
9 ### map each color to an integer
10 mapping = {}
11 for x in range(len(total_colors)):
12     mapping[total_colors[x]] = x
13
14 # integer representation
15 for x in range(len(colors)):
16     colors[x] = mapping[colors[x]]
17
18 one_hot_encode = to_categorical(colors)
19 print(one_hot_encode)
```

Run

Output 4.13s

```
[[1. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 1.]
 [1. 0. 0. 0. 0.]
 [0. 0. 1. 0. 0.]
```

Пропущенные значения рекомендуется либо удалить, используя соответствующую функцию библиотеки Pandas:

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.dropna.html>

Либо заполнить какими-либо значениями, например с использованием соответствующей функции также из библиотеки Pandas:

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.fillna.html>.

Для контроля переобучения и качества обучения моделей выборку необходимо разделить на обучающую и валидационную соответствующей функцией библиотеки Scikit Learn:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

Рекомендуемые пропорции тренировочной и валидационной выборки зависят от размера данных. Обычно доля валидационной выборки составляет от 20 до 35 %.

Процесс создания нейронной сети описан в материалах лекции № 4, рекомендуется изучить ее перед выполнением данной лабораторной работы.

Задача классификации подразумевает использование в качестве функции активации последнего слоя либо сигмоиду, либо softmax, в зависимости от количества нейронов. А количество нейронов должно равняться количеству классов (в случае 2-х классов достаточно 1 нейрона с функцией активации сигмоида и бинарной кроссэнтропией).

В задачах регрессии функция активации зависит от характера предсказываемых данных. Если это стоимость дома (меняется от 0 до бесконечности), то выходной слой будет состоять из 1 нейрона с функцией активации Relu. Если диапазон значений переменной будет ограничиваться от -1 до 1, до оптимальным вариантом будет использование tanh, если же значения целевой переменной изменяются от $-\infty$ до $+\infty$, то оправдано использование линейной функции активации.

<https://keras.io/api/layers/activations/>

Задание 1.

Используя данные для задач классификации для вашего варианта:

- 1) Изучите описание данных, загрузите данные используя библиотеку Pandas.
- 2) Кратко опишите структуру данных, и целевую переменную, которую необходимо будет предсказывать.
- 3) Осуществите предобработку данных, предварительно удалив неинформативные параметры. Создайте обучающую и тестовую выборки.
- 4) Спроектируйте базовый вариант нейронной сети, выберите либо напишите функцию ошибки и метрику точности.
- 5) Обучите нейронную сеть и выведите на экран графики основных ее метрик.
- 6) Проведите исследования зависимости качества предсказания нейронной сети от:
 - a) Размера валидационной выборки.
 - b) Стандартизации данных
 - c) Предложите не менее 2 различных архитектур нейронной сети. Сравните качество их работы
 - d) Поэкспериментируйте с различным количеством нейронов в слоях.
- 7) Добейтесь максимальной точности модели

Задание 2.

Используя данные для задач регрессии для вашего варианта:

- 8) Изучите описание данных, загрузите данные используя библиотеку Pandas.
- 9) Кратко опишите структуру данных, и целевую переменную, которую необходимо будет предсказывать.
- 10) Осуществите предобработку данных, предварительно удалив неинформативные параметры. Создайте обучающую и тестовую выборки.

- 11) Спроектируйте базовый вариант нейронной сети, выберите либо напишите функцию ошибки и метрику точности.
- 12) Обучите нейронную сеть и выведите на экран графики основных ее метрик.
- 13) Проведите исследования зависимости качества предсказания нейронной сети от:
 - a) Размера валидационной выборки.
 - b) Стандартизации данных
 - c) Предложите не менее 2 различных архитектур нейронной сети. Сравните качество их работы
 - d) Поэкспериментируйте с различным количеством нейронов в слоях.
- 14) Добейтесь максимальной точности модели

Материалы, задания, данные для выполнения работ приложены в сопроводительных файлах.

Лабораторная работа 5.

Классификация текстов с использованием полносвязных нейронных сетей

Методические рекомендации:

17. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
18. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
19. Лабораторная работа состоит из 1 задания, 6 вариантов.
20. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции №6 классификации текстов.

Задание.

1. Изучите базу текстов, которая соответствует вашему варианту. Материалы с текстами приложены к лабораторной работе в виде архива. Изучите описание датасета. Датасет содержит набор текстов, разделенный на классы. В зависимости от варианта он может быть в виде csv файла, в виде нескольких csv или txt файлов, либо группа файлов, рассортированная по каталогам.
2. Примените подход, аналогичный примеру из лекции. Создайте функцию, которая считывает данные из файла(-ов), обрабатывает тексты токенайзером, переводит все слова в токены, разбивает их на фрагменты (окна) определенной длины, каждому фрагменту должен соответствовать лейбл(метка) класса.

Размер фрагмента вынесите в аргумент функции, чтобы была возможность изменять его.

Функция должна возвращать обучающую и тестовую выборки: `pr.array()` массив фрагментов токенов в качестве матриц параметров, а также `pr.array()` массив лейблов(меток) в формате ONE, которые должны указывать на класс текста для матриц Y .

Для создания ONE массива необходимо сначала создать вектор числовых лейблов, затем при помощи утилиты `utils.to_categorical()` перевести их в массив ONE.

Кроме того в аргументах функции должен присутствовать размер словаря токенайзера.

```
def data_preproc(path,window, vocabulary):  
    .....  
    .....  
    .....  
    return X_train, Y_train, X_test, Y_test
```

3. Переведите X_{train} и X_{test} в формат BOW, спроектируйте нейронную сеть, обучите ее.
4. Протестируйте влияние гиперпараметров на качество нейронной сети:
 - a. Влияние размера окна на точность и величину ошибки
 - b. Влияние размера словаря на точность и величину ошибки
 - c. Протестируйте различное число слоев нейронной сети и количество нейронов.
 - d. Протестируйте влияние слоев Dropout и BatchnormalizationДля всех случаев постройте графики обучения. Визуализируйте результат для сравнения параметров.
5. Добейтесь максимальной точности полносвязной нейронной сети.
6. Выведите отчет о точности распознавания каждого класса, а также информацию о размерах классов. Сделайте выводы.

По итогам выполнения работы необходимо оформить отчет в печатном виде, либо в формате Jupyter Notebook с комментариями, графиками и объяснениями.

Лабораторная работа 6.

Классификация текстов с использованием эмбедингов, рекуррентных и сверточных нейронных сетей

Методические рекомендации:

21. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
22. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
23. Лабораторная работа состоит из 3 заданий, 6 вариантов.
24. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции № 7 классификации текстов.

Задание 1. Классификация текстов с использованием сверточных нейронных сетей.

7. Изучите базу текстов, которая соответствует вашему варианту. Материалы с текстами приложены к лабораторной работе в виде архива. Изучите описание датасета. Датасет содержит набор текстов, разделенный на классы. В зависимости от варианта он может быть в виде csv файла, в виде нескольких csv или txt файлов, либо группа файлов, рассортированная по каталогам.
8. Используя наработки из лабораторной работы № 5 преобразуйте тексты в токены, а также создайте обучающую и тестовую выборки токенизированных текстов. Обратите внимание - нас интересуют именно токенизированные выборки, формат BOW в лабораторной работе использоваться не будет.
9. Реализуйте архитектуру нейронной сети и материалов лекции.
10. Протестируйте влияние гиперпараметров на качество нейронной сети:
 - a. Влияние размера окна на точность и величину ошибки
 - b. Влияние размера словаря на точность и величину ошибки
 - c. Протестируйте различное число слоев нейронной сети и количество нейронов.
 - d. Проанализируйте влияние размера эмбединга на качество работы нейронной сети
 - e. Проанализируйте размера ядра свертки, функции активации, количества нейронов сверточной нейронной сети на ее точность.Для всех случаев постройте графики обучения. Визуализируйте результат для сравнения параметров.
11. Добейтесь максимальной точности сверточной нейронной сети.
12. Выведите отчет о точности распознавания каждого класса, а также информацию о размерах классов. Сделайте выводы.

Задание 2. Классификация текстов с использованием LSTM нейронных сетей.

1. Реализуйте архитектуру рекуррентной нейронной сети и материалов лекции.
2. Протестируйте влияние гиперпараметров на качество нейронной сети:
 - a. Проанализируйте влияние числа нейронов LSTM слоя на качество распознавания текстов
 - b. Проанализируйте влияние процедуры Bidirectional, параметра return_sequences на качество работы нейронной сети.Для всех случаев постройте графики обучения. Визуализируйте результат для сравнения параметров.
3. Добейтесь максимальной точности нейронной сети.

4. Выведите отчет о точности распознавания каждого класса, а также информацию о размерах классов. Сделайте выводы.

Задание 3.

Спроектируйте собственную архитектуру нейронной сети, комбинируя различные полносвязные слои, сверточные, рекуррентные. Добейтесь максимальной точности.

По итогам выполнения работы необходимо оформить отчет в печатном виде, либо в формате Jupyter Notebook с комментариями, графиками и объяснениями.

Лабораторная работа 7. Кластеризация текстов

Методические рекомендации:

25. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
26. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
27. Лабораторная работа состоит из 1 задания, 6 вариантов.
28. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции № 7 кластеризация текстов.

Задание 1. Кластеризация текстов.

- Возьмите текстовую базу из материалов урока.
- Сделайте классификатор с embedding слоем в начале (не Bag of words).
- Обучите классификатор.
- Возьмите предобученный embedding слой и перегоните все слова из полного словаря в embedding-пространство.
- Запустите кластеризацию.
- Выведите устойчивые классы слов.
- Проведите анализ для разного числа классов и разной размерности embedding-пространства.
- Напишите выводы.

Оформите отчет в виде Google Colab Notebook либо Jupyter Notebook.

** Примечание. При оформлении эмбединг слоя в виде внутренней модели вы точно так же можете ей пользоваться, подавая на вход токены, на выходе получать векторное представление этих токенов.

Лабораторная работа 8. Сегментация текстов

Методические рекомендации:

29. Рекомендуемый язык программирования – Python. Предполагается базовые знания языка программирования Python, умение работать со строками в Python, а также базовое понимание ООП на Python.
30. Рекомендуемые среды разработки:
 - a. PyCharm <https://www.jetbrains.com/pycharm/>
 - b. Пакет Anaconda (Jupyter Notebook) <https://www.anaconda.com/products/individual>
31. Лабораторная работа состоит из 2 заданий.
32. Перед выполнением лабораторной работы необходимо изучить теоретический материалы из Лекции № 8 сегментация текстов.

Задание 1. Реализуйте нейронную сеть из материалов лекции.

Разработайте модель, позволяющую выявлять сущности в текстах договоров. Для этого считайте договора, тексты переведите в начальную форму при помощи PyMorphy, преобразуйте слова в эмбединги с помощью предобученной модели Gensim, сформируйте обучающую матрицу Y, обучите модели и добейтесь точности не хуже, чем в материалах лекции.

Задание 2. Разработайте собственную архитектуру нейронной сети с собственным Embedding слоем. Сравните качество работы модели по сравнению с Gensim Word2Vec.

Оформите отчет в виде Google Colab Notebook либо Jupyter Notebook.

Оptionальное задание для самостоятельного выполнения.

<https://github.com/synalp/NER/tree/master/corpus/CoNLL-2003>

Используя данный датасет размеченных русских новостей сформируйте обучающую выборку, разработайте архитектуру нейронной сети и обучите модель для поиска сущностей.

Примерные вопросы при собеседовании.

1. Каким образом производилась предобработка текста.
2. Как формировалась обучающая и валидационная выборка.
3. Обоснуйте выбор архитектуры и параметров нейронной сети.
4. Аргументируйте выбор метрик качества предсказания модели и методов расчета ошибок предсказания.
5. Проанализируйте графики ошибки обучения модели на обучающей и валидационной выборке.
6. Какие способы повышения точности модели вы видите.
7. Проведите прогноз по построенным моделям.
8. Сравните качество вашей модели с предобученными эмбедингами.
9. В чем основные недостатки вашей модели и какие существуют пути их решения.

2. Вопросы для промежуточной аттестации

1. Автоматическая обработка естественного языка в кругу смежных дисциплин.
2. Особенности естественного языка и возможности его автоматической обработки.
3. Основные задачи автоматического анализа текстов и подходы к их решению.
4. Оценка систем автоматической обработки текстов.
5. Предобработка текста. Регулярные выражения.
6. Стеммеры, лемматизаторы, морфологические анализаторы.
7. Проблемы языковых моделей и способы их решения. Методы оценки языковых моделей.
8. Задачи разметки текста, применение разметки.
9. Классификация текстов: формулировка задачи и методы решения.
10. Наивный байесовский классификатор. Проблемы классификации текстов.
11. Информационный поиск и векторные модели текстов.
12. Задача парсинга, его применение. Синтаксис составляющих и синтаксис зависимостей.
13. Биологический и искусственный нейрон.
14. Основные функции активации нейронов. Преимущества нейронных сетей.
15. Классификации нейронных сетей, области применения и решаемые задачи.
16. Персептрон Розенблата.
17. Алгоритм обучения персептрона и правило Хебба.
18. Теорема о сходимости алгоритма обучения персептрона для линейноразделимых множеств. Проблема исключаящего «или».
19. Многослойный персептрон. Представление булевых функций.
20. Нейронные сети как универсальные аппроксиматоры.
21. Общая идея градиентных методов решения задач безусловной оптимизации. Метод наискорейшего спуска.
22. Алгоритм обратного распространения ошибки. Достоинства и недостатки алгоритма. Понятие паралича сети и причины его возникновения.
23. Эвристические приемы улучшения сходимости и качества градиентного обучения (нормализация, выбор функции активации, выбор начальных значений весов, порядок предъявления обучающих примеров, выбор величины шага, сокращение числа весов, выбивание из локальных минимумов, проблема переобучения и разделение выборки).
24. Методы упрощения структуры нейронной сети. Общие принципы обучения.
25. Задача кластеризации данных. Основные метрики для количественных и нечисловых переменных.
26. Подготовка и предобработка текстов перед подачей данных в нейросеть.
27. Предобработка данных в задачах кластеризации текстов.
28. Переобучение моделей машинного обучения. Причины, признаки. Способы решения данной проблемы.
29. Метрики качества. Методика выбора. Примеры.
30. Задача классификации текстов с использованием BOW. Реализация при помощи библиотеки Keras(либо TensorFlow, либо PyTorch).

31. Задача классификации текстов с использованием эмбедингов. Реализация при помощи библиотеки Keras(либо TensorFlow, либо PyTorch).
32. Современные тенденции, направления решения задачи классификации текстов.
33. Косинусное расстояние между текстами. Семантический спектр. Пример реализации.
34. Морфологические анализаторы естественного языка на примере библиотек Python.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

В. В. Воронина

Методы искусственного интеллекта в предиктивной аналитике и бизнес-аналитике

Методические указания для проведения лабораторных занятий
для студентов направлений

09.04.04 «Программная инженерия»

профиль Искусственный интеллект и предиктивная аналитика

09.04.03 «Прикладная информатика»

профиль Искусственный интеллект и бизнес-аналитика

Ульяновск
УлГТУ
2021

УДК 004 (076)

ББК 32.973-018.1я7

В12

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Воронина, Валерия Вадимовна

- В12 Методы искусственного интеллекта в предиктивной и бизнес-аналитике : практикум для проведения лабораторных занятий для студентов направлений 09.04.04 «Программная инженерия» профиль Искусственный интеллект и предиктивная аналитика, 09.04.03 «Прикладная информатика» профиль Искусственный интеллект и бизнес-аналитика / В. В. Воронина. – Ульяновск : УлГТУ, 2021. – 27 с.

Составлен в соответствии с учебными планами направлений 09.04.04 «Программная инженерия» профиль Искусственный интеллект и предиктивная аналитика, 09.04.03 «Прикладная информатика» профиль Искусственный интеллект и бизнес-аналитика. Цель данного практикума – ориентировать студентов на содержание и порядок выполнения лабораторных работ во время прохождения ими курсов «Методы искусственного интеллекта в предиктивной аналитике», «Методы искусственного интеллекта в бизнес-аналитике», «Методы глубокого обучения в бизнес-аналитике». Даются задания для лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК 004 (076)

ББК 32.973-018.1я7

© В. В. Воронина, 2021.

© Оформление. УлГТУ, 2021.

Оглавление

Общая информация.....	3
Задания на лабораторные работы по предмету «Методы искусственного интеллекта в бизнес-аналитике»	3
Задания на лабораторные работы по предмету «Методы искусственного интеллекта в предиктивной аналитике»	5
Задания на лабораторные работы по предмету «Методы глубокого обучения в бизнес-аналитике»	7
Предлагаемые массивы данных для анализа	10
Список используемых источников.....	27

Общая информация

В рамках лабораторного практикума вам предлагается проанализировать данные с платформы Kaggle. Вы можете взять один из предложенных массивов данных, или выбрать свой. Согласно [1]: «Kaggle – это платформа для исследователей разных уровней, где они могут опробовать свои модели анализа данных на серьезных и актуальных задачах. Суть такого ресурса – не только в возможности получить неплохой денежный приз в случае, если именно ваша модель окажется лучшей, но и в том (а, это, пожалуй, гораздо важнее), чтобы набраться опыта и стать специалистом в области анализа данных и машинного обучения. Ведь самый важный вопрос, зачастую стоящий перед такого рода специалистами – где найти реальные задачи? Здесь их достаточно».

Общее требование к лабораторным по предметам «Методы искусственного интеллекта в бизнес-аналитике» и «Методы искусственного интеллекта в предиктивной аналитике»: организовать следующую архитектуру системы: каждый метод анализа данных оформить в виде веб-сервиса и создать управляющую систему, вызывающую реализованные веб-сервисы.

Задания на лабораторные работы по предмету «Методы искусственного интеллекта в бизнес-аналитике»

Лабораторная работа 1. Работа с данными и проектирование общей системы

Исходя из общих требований, разработать архитектуру системы для выполнения всех лабораторных работ. Выбрать массив данных для анализа, определить, какие задачи классификации, кластеризации и прогнозирования вы сможете на нем решить (по две задачи каждого типа). Выбрать

необходимые признаки, осуществить загрузку, очистку и визуализацию данных.

Вопросы для собеседования по лабораторной работе:

1. В каких бизнес-процессах может быть использован выбранный вами набор данных?
2. Из каких источников данных может быть получен выбранный вами набор данных?
3. Решение каких задач с использованием искусственного интеллекта будет эффективно для выбранного вами набора данных?
4. Каким образом вы подбирали признаки для ваших задач?
5. Какова будет архитектура вашей системы и какие функции?

Лабораторная работа 2. Решение задач прогнозирования

Для выбранного вами массива данных решите две различных задачи прогнозирования различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи прогнозирования?
3. В чем особенность нейросетевых методов для решения задачи прогнозирования?
4. В чем особенность нечетких методов для решения задачи предсказания?
5. Расскажите об особенностях TimeSeries DataMining в контексте решаемой задачи.

Лабораторная работа 3. Решение задач классификации

Для выбранного вами массива данных решите две различных задачи классификации различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи классификации?
3. В чем особенность нейросетевых методов для решения задачи классификации?
4. В чем особенность нечетких методов для решения задачи классификации?
5. В чем особенность деревьев решений для решения задачи классификации?

Лабораторная работа 4. Решение задач кластеризации

Для выбранного вами массива данных решите две различных задачи кластеризации различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. В чем особенность задачи кластеризации в отличие от классификации?
2. Почему в задачах кластеризации так важен выбор критерия качества?
3. Для решения каких задач используется кластеризация, как метод обработки данных?
4. Какие вы знаете методы кластеризации данных?
5. Какие методы для решения задач вы выбрали и почему?

Лабораторная работа 5. Оценка качества и выбор лучшей модели

Для лабораторных работ 2-4 выполните решение каждой из задач не используемым в лабораторных методом. Оцените каждое из решений и выберите лучшую модель для каждой задачи.

Вопросы для собеседования по лабораторной работе:

1. Какими критериями можно оценить качество работы кластеризационной модели?
2. Какими критериями можно оценить качество работы предсказательной модели?
3. Как переобучение влияет на качество работы модели?
4. Какие есть методы борьбы с переобучением?
5. Расскажите подробнее про кросс-валидацию.

Задания на лабораторные работы по предмету «Методы искусственного интеллекта в предиктивной аналитике»

Лабораторная работа 1. Работа с данными и проектирование общей системы

Исходя из общих требований, разработать архитектуру системы для выполнения всех лабораторных работ. Выбрать массив данных для анализа. Выбрать перспективные признаки для решения задачи прогнозирования, осуществить загрузку, очистку и визуализацию данных.

Вопросы для собеседования по лабораторной работе:

1. В каких бизнес-процессах может быть использован выбранный вами набор данных?

2. Из каких источников данных может быть получен выбранный вами набор данных?
3. Решение каких задач предсказания с использованием искусственного интеллекта будет эффективно для выбранного вами набора данных?
4. Какими критериями вы можете оценить эффективность внедрения искусственного интеллекта в анализ выбранного вами набора данных?
5. Какова будет архитектура вашей системы и какие функции?

Лабораторная работа 2. Решение задач прогнозирования регрессионными методами

Для выбранного вами массива данных решите задачу прогнозирования различными регрессионными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задачи вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи прогнозирования?
3. Как подготовить текстовые данные для обработки линейной регрессией?
4. Что общего и чем отличаются модели Ridge и Lasso?
5. Для каких задач хорошо подходят линейные регрессионные модели?

Лабораторная работа 3. Решение задач прогнозирования нейросетевыми методами

Для выбранного вами массива данных решите задачу прогнозирования нейросетевыми методами (различной архитектуры). Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задачи вы выбрали и почему?
2. В чем особенность нейросетевых методов для решения задачи прогнозирования?
3. Как работа с нейронными сетями реализована в python?
4. Как необходимо готовить данные для обработки нейронными сетями?
5. Как вы выполнили визуализацию данных?

Лабораторная работа 4. Решение задач прогнозирования методами нечеткой логики или деревьев решений

Для выбранного вами массива данных решите задачу прогнозирования методами нечеткой логики или методами деревьев решений. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какой метод вы выбрали для решения и почему?
2. В чем особенность нечетких методов для решения задачи предсказания?
3. Расскажите о TimeSeries DataMining.
4. Расскажите подробнее о деревьях решений.
5. В чем преимущества и недостатки использования методов нечеткой логики?

Лабораторная работа 5. Оценка качества и выбор лучшей модели

Оцените каждое из решений в лабораторных работах 2-4 и выберите лучшую модель для решения задачи прогнозирования на вашем массиве данных.

Вопросы для собеседования по лабораторной работе:

1. Какими критериями можно оценить качество работы предсказательной модели?
2. Как переобучение влияет на качество работы модели?
3. Какие есть методы борьбы с переобучением?
4. Расскажите подробнее про кросс-валидацию.
5. Как оценить качество работы системы нечеткого логического вывода?

Задания на лабораторные работы по предмету «Методы глубокого обучения в бизнес-аналитике»

Лабораторная работа 1. Работа со сверточной нейронной сетью. Часть 1

Ознакомьтесь с пунктом «Сравнение эффективности моделей из библиотеки Keras» (ресурс [3], стр. 263). Проведите серию экспериментов на тестовых данных, интерпретируйте полученные результаты, ответьте на вопросы, сформулированные в конце пункта. Для одного из приведенных ниже наборов данных (или найденных вами на Kaggle) предложите задачу и соответствующее дополнение этого набора данных, чтобы применение сверточной сети для его анализа было целесообразным.

Вопросы для собеседования по лабораторной работе:

1. Как необходимо подготовить данные для работы со сверточной нейронной сетью?
2. Для решения каких задач лучше всего подходят сверточные нейронные сети?

3. В чем сложность работы с данным видом моделей глубокого обучения?
4. Что вы научились делать в ходе лабораторной работы?
5. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?

Лабораторная работа 2. Работа со сверточной нейронной сетью. Часть 2

Ознакомьтесь с пунктом «Работа с библиотекой OpenCV» (ресурс [3], стр. 265). Используя библиотеку OpenCV, сравните эффективность моделей, построенных в предыдущей лабораторной работе на массиве данных фотографий.

Вопросы для собеседования по лабораторной работе:

1. В чем преимущества и недостатки сверточной нейронной сети по сравнению с другими моделями?
2. Что вы научились делать в ходе лабораторной работы?
3. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?
4. Какой объем экспериментов вам пришлось проделать?
5. Что вы можете сказать о качестве исследуемых моделей?

Лабораторная работа 3. Работа с рекуррентной сетью. Часть 1

Ознакомьтесь с пунктом «Работа с рекуррентными сетями» (ресурс [3], стр. 274), выполните сформулированные задачи:

1. Модифицируйте код из листинга 94: поэкспериментируйте с разной структурой сети, разным количеством точек, по которым строится прогноз, и с разным количеством LSTM-блоков. Оцените качество работы сети в разных экспериментах. Интерпретируйте результаты. Данные для анализа можно взять тут: <https://www.kaggle.com/andreazzini/international-airline-passengers>
2. Для работы с временным рядом найдите массив данных на Kaggle, выполните прогноз нерекуррентной и рекуррентной сетями. Сравните и интерпретируйте результаты.
3. Объясните работу кода из листинга 98, а также проведите два эксперимента, модифицируя код из листинга 98 следующим образом:
 - Изменив значение `seq_length`, сделав его равным 3 и изменив форму входных данных следующим образом: `X = numpy.reshape(dataX, (len(dataX), 1, seq_length))`
 - Изменив значение `seq_length`, сделав его равным 3 и изменив форму входных данных следующим образом: `X = numpy.reshape(dataX, (len(dataX), 1, seq_length))`

Интерпретируйте полученные результаты.

Вопросы для собеседования по лабораторной работе:

1. Как необходимо подготовить данные для работы с рекуррентной нейронной сетью?
2. Для решения каких задач лучше всего подходят рекуррентные нейронные сети?
3. В чем сложность работы с данным видом моделей глубокого обучения?
4. Что вы научились делать в ходе лабораторной работы?
5. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?

Лабораторная работа 4. Работа с рекуррентной сетью. Часть 2

Проведите серию экспериментов использования рекуррентной сети, как генеративной модели на разных наборах данных (кириллица-латиница, поэзия-проза и т.д.). Сделайте выводы.

Вопросы для собеседования по лабораторной работе:

1. В чем преимущества и недостатки рекуррентной нейронной сети по сравнению с другими моделями в плане решения задачи обработки текста?
2. Что вы научились делать в ходе лабораторной работы?
3. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?
4. Какой объем экспериментов вам пришлось проделать?
5. Что вы можете сказать о качестве вашего решения?

Лабораторная работа 5. Разработка коммерческого предложения

Подберите комплектацию самого дешевого, среднего и самого дорогого компьютера (и «железо» и программное обеспечение) для использования методов глубокого обучения в бизнес-аналитике. Рассмотрите варианты аренды серверов. Сделайте выводы, при каких задачах какой вариант предпочтительнее. Оформите лабораторную работу в виде коммерческого предложения для какого-либо бизнеса с представленным анализом различных вариантов и обоснованием основного.

Вопросы для собеседования по лабораторной работе:

1. Что такое CUDA-вычисления? Когда они необходимы и на каком «железе» выполняются?
2. Что может служить критерием целесообразности использования методов глубокого обучения в бизнес-аналитике?
3. В чем сложность внедрения глубокого обучения в бизнес-аналитику?

4. Из чего складывается стоимость внедрения глубокого обучения в бизнес-аналитику?
5. Что вы научились делать в ходе лабораторной работы?

Предлагаемые массивы данных для анализа

Рестораны Бангалора

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

Приведенное описание набора данных:

«The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry has'nt been saturated yet and the demand is increasing day by day. In spite of increasing demand it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality. Do the entire locality loves vegetarian food. If yes then is that locality populated by a particular sect of people for eg. Jain, Marwaris, Gujaratis who are mostly vegetarian. These kind of analysis can be done using the data, by studying different factors.»

Перевод:

«Основная идея анализа набора данных Zomato состоит в том, чтобы получить четкое представление о факторах, влияющих на совокупный рейтинг каждого ресторана, создание различных типов ресторанов в разных местах, поскольку в одном из таких городов в Бангалоре насчитывается более 12 000 ресторанов, в которых подают блюда из по всему миру. С каждым днем новые рестораны открываются, индустрия еще не насыщена, и спрос растет день ото дня. Несмотря на растущий спрос, новым ресторанам стало трудно конкурировать с известными ресторанами. Большинство из них подают одинаковую еду. Бангалор, являющийся ИТ-столицей Индии. Большинство людей здесь зависят в основном от ресторанной еды, так как у них нет времени готовить для себя. Поэтому при таком огромном спросе на рестораны стало важно изучать демографию места. Какая еда более популярна в местности? У всего населенного пункта любит вегетарианскую еду. Если да, то является ли эта местность населенной определенной сектой людей, например. Джайны, марвари, гуджараты, которые в основном вегетарианцы. Этот вид анализа может быть сделан с использованием данных, изучая различные факторы».

- Пример строки данных (одной из 60 149):

url	https://www.zomato.com/bangalore/caf-eleven-banashankari-bangalore?context=eyJzZSI6eyJlIjpbIjE4MTQ4OTQxIiwMTg0NzA3NzEiLCIxODcxNjA2MSIsIjE4NjMyOTc3IiwxODIwMjk5NywiNTkwOTAiLCIxODU1OTIxMSIsIjE4MzIzNjM5IiwMTg1NzQyNzciLCIxODcyMzQ5NCIsIjE4ODY3MTIyIiwMTg2ODg0ODkiLCIxODUyOTA5MCIsljU2MTY2IiwjA5MTkiLCIxODM5Njc2MCIsljE4NTkzNTE5IiwMTg1MDAwMjAiLCIxODg0NDI2MCIsljU3OTk0IiwjNTUwOTAiLCI1NzQzNSJdLCJ0IjoiQ2FmXHUwMGU5cyBhbmQgRGVsaXMgaW4gQmFuYXNoYW5rYXJpIn19
address	111, Sapphire Toys Building, 100 Feet Ring Road, Banashankari, Bangalore
name	Caf-Eleven
online_order	No
book_table	No
rate	4.0/5
votes	424
phone	8 049 577 715
location	Banashankari
rest_type	Cafe
dish_liked	Sandwich, Omelette, Ice Tea, Virgin Mojito, Hot Chocolate, Pasta Arrabiata, Hazelnut Cappuccino
cuisines	Cafe, Continental
approx_cost(for two people)	450
reviews_list	[('Rated 2.0', "RATED\n This is a hookah cafe. It was not mentioned anywhere in the page that it was a hookah place. There's also no separate seating for non smokers. Horrible ambience. Ppl including waiters smoking hookah everywhere. I'm never gonna visit this place again. Also, I found an eyelash in the hot chocolate shake. The hot and cold chocolate are below average."), ('Rated 4.0', "RATED\n Loved the ambience and the interior, very colourful and lively.\nThe seating is spacuous!\nThe food tasted really good\nThe chicken was evenly cooked and was soft\nWe ordered lemon chicken dragon chicken and jalapeno poppers\nThe poppers were oily but otherwise it was good!\nGreat music\nNice sheesha!\nOne disadvantage , the sheesha is only for one hour on the weekends ... We'd like to obviously sit for more time!\nOtherwise everything is great.\nStaff can be a little more well learnt or greetable"), ('Rated 4.0', "RATED\n It's actually a hookah cafe with awesome ambience and good food and a bit slow service , I ordered chicken burger which wasn't very good but the barbecue chicken wings were awesome Its a great place for hanging out with your friends."), ('Rated 3.5', "RATED\n Visited this place on a weekday and only 2 or 3 tables were filled. This place is located in the same building of sapphire toys in the main road of kamakya road.\n\nOrdered KitKat shake, ide to milkshake, onion rings and chilly chicken. Everything was pretty decent.\nChilly chicken was spicy and tasty. Onion rings could have been done more well. Both the milkshake was above average. Sheesha is the main

attraction.\n\nAmbience is decent. They've used an old scooter and made it look attractive. Service is prompt and good."), ('Rated 4.0', "RATED\n A hookah lounge!\n\nBakasura makes it a point to mention such info before every review. As some are particular in such ambience.\n\nIt's bakasura Recommended!!\n\nThe food bakasura ordered were fries, nachos, pizza, blueberry Oreo and virgin mojito!\n\nWhen it comes to food. Every single thing was good. The nachos was crazily special. The blueberry Oreo milkshake is something that bakasura Recommends. Pizza was thick crust and very well stuffed.\n\nComing to the ambience. It's rooftop and it's crazily colourful. The yellow chairs and the Christmas tree that was put up. Everything is perfect.\n\nStaff needs a slight improvement.\n\nOverall it's great\n\nFollow Bakasura for more such reviews.\n\n#BakasuraForever"), ('Rated 1.0', "RATED\n For the price we pay here worst food I have tasted , steaks are bit ok but not that good. I love pasta to the core but red sauce pasta had lots and lots of basil in it and taste was also not gud .I wouldn't recommend this place."), ('Rated 3.0', "RATED\n Ambience: This cafe is located in banashankari 2nd stage near kamakya theatre. This is basically a cafe on the rooftop which is very busy all the time. The place has nostalgia feels because old scooters are present inside which attracts everyone. Has small LED lights, lamps, dream catchers and plants around which looks cute during evening.\n\nFood : 1) BBQ veg pizza was very big and was good. But wasn't yummy as expected.\n2) Arrabiata pasta, This is a must try. Was really tasty and yummy and creamy.\n3) Jalapeno cheese balls was very cheesy and hot hot served with mayo.\n4) Oreo milkshake was okay okay.\n\nService: was really bad."), ('Rated 5.0', 'RATED\n Great hangout spot to spend long time over conversations with friends with delicious food.\n\nMust try: cheese toasts, babycorn pepper dry\n\n#rajeshwrites'), ('Rated 3.0', "RATED\n Good place to hang out with friends. Food wise nothing great. Not advisable to go with family. Even if it's roof and sometimes it's suffocating"), ('Rated 4.0', "RATED\n First thing first, do book a table if you're going here on a weekend. We had to wait for about 20 minutes.\n\nTalking about the service, I think it has to get slightly better. 3/5\n\nThough people prefer this place for sheesha, I'd been there just to grab a few snacks. The cafe frappe and mushroom pepper were great! No complains. 4.5/5\n\nAmbience: 3/5 I wish to go there again, some time during the week."), ('Rated 2.0', 'RATED\n Before cafe eleven was good enough with the food stuffs and all but now staff has been changed and it is not a pocket friendly cafe too and service is too poor ? hookahs are good enough but food taste is up to the mark ? ambience is good enough ??'), ('Rated 3.0', "RATED\n It's a decent rooftop with pretty funky ambience like scooters, colourful seats and green pots on it. Food tastes good! I didn't liked brownie fraape and our order took a lot of time too, just look towards that once. It's pretty good place for hookah lovers. Reasonable price too! Overall one time visit is ok.")]

menu_item	[]
listed_in(type)	Cafes
listed_in(city)	Banashankari

Цены на топливо в Бразилии

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/matheusfreitag/gas-prices-in-brazil>

Приведенное описание набора данных:

«The Data

The National Agency of Petroleum, Natural Gas and Bio fuels (ANP in Portuguese) releases weekly reports of gas, diesel and other fuels prices used in transportation across the country. These datasets bring the mean value per liter, number of gas stations analyzed and other information grouped by regions and states across the country.

Source

As stated before, these datasets are provided by ANP, and are regularly updated with new dates and information - which can be retrieved here (in portuguese).

What can be done with this?

How different regions of Brazil saw their gas prices change?

Within a region, which states increased more their prices?

Which states are the cheapest (or most expensive) for different types of fuels?

Перевод:

«Данные

Национальное агентство нефти, природного газа и биотоплива (ANP на португальском языке) публикует еженедельные отчеты о ценах на газ, дизельное топливо и другие виды топлива, используемые при перевозках по всей стране. Эти наборы данных содержат среднее значение на литр, количество проанализированных заправочных станций и другую информацию, сгруппированную по регионам и штатам по всей стране.

Источник

Как указывалось ранее, эти наборы данных предоставляются ANP и регулярно пополняются новыми датами и информацией, которые можно найти здесь (на португальском языке).

Что можно сделать с этим? Можно ответить на вопросы:

Как в разных регионах Бразилии цены на газ изменились?

В пределах региона, какие государства увеличили свои цены?

Какие государства являются самыми дешевыми (или самыми дорогими) для различных видов топлива?»

- Пример строки данных (одной из 106 824):

DATA INICIAL	09.05.2004
DATA FINAL	15.05.2004
REGIÃO	CENTRO OESTE

ESTADO	DISTRITO FEDERAL
PRODUTO	ETANOL HIDRATADO
NÚMERO DE POSTOS PESQUISADOS	127
UNIDADE DE MEDIDA	R\$/l
PREÇO MÉDIO REVENDA	1,288
DESVIO PADRÃO REVENDA	0,016
PREÇO MÍNIMO REVENDA	1,19
PREÇO MÁXIMO REVENDA	1,35
MARGEM MÉDIA REVENDA	0,463
COEF DE VARIAÇÃO REVENDA	0,012
PREÇO MÉDIO DISTRIBUIÇÃO	0,825
DESVIO PADRÃO DISTRIBUIÇÃO	0,11
PREÇO MÍNIMO DISTRIBUIÇÃO	0,4201
PREÇO MÁXIMO DISTRIBUIÇÃO	0,9666
COEF DE VARIAÇÃO DISTRIBUIÇÃO	0,133
MÊS	5
ANO	2004

Преступления в Бостоне

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/AnalyzeBoston/crimes-in-boston#crime.csv>

Приведенное описание набора данных:

«Context

Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred.

Content

Records begin in June 14, 2015 and continue to September 3, 2018.

Acknowledgements

The data is provided by Analyze Boston. The most up-to-date version can be found here.

Inspiration

What types of crimes are most common? Where are different types of crimes most likely to occur? Does the frequency of crimes change over the day? Week? Year?»

Перевод:

«Бостонское полицейское управление (BPD) предоставляет отчеты об инцидентах с преступностью, чтобы задокументировать первоначальные

подробности инцидента, на которые реагируют сотрудники BPD. Это набор данных, содержащий записи из новой системы отчетов о происшествиях с преступностью, которая включает в себя сокращенный набор полей, предназначенных для регистрации типа происшествия, а также того, когда и где он произошел.

содержание

Записи начинаются 14 июня 2015 года и продолжаются до 3 сентября 2018 года.

Подтверждения

Данные предоставлены Analyze Boston. Самую актуальную версию можно найти здесь.

возможные вопросы

Какие виды преступлений наиболее распространены? Где различные виды преступлений наиболее вероятны? Меняется ли частота преступлений за день? Неделю? Год?»

- Пример строки данных (одной из 106 824):

Incident_number	I182070945
Offense_code	619
Offense_code_group	Larceny
Offense_description	LARCENY ALL OTHERS
District	D14
Reporting_area	808
Shooting	
Occurred_on_date	02.09.2018 13:00:00
Year	2018
Month	9
Day_of_week	Sunday
Hour	13
Ucr_part	Part One
Street	LINCOLN ST
Lat	42.35779134
Long	-71.13937053
Location	"(42.35779134-71.13937053)"

Продажи женской обуви

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/datafiniti/womens-shoes-prices/>

«About This Data

This is a list of 10,000 women's shoes and their product information provided by Datafiniti's Product Database.

The dataset includes shoe name, brand, price, and more. Each shoe will have an entry for each price found for it and some shoes may have multiple entries.

Note that this is a sample of a large dataset. The full dataset is available through Datafiniti.

What You Can Do with This Data

You can use this data to determine brand markups, pricing strategies, and trends for luxury shoes. E.g.:

What is the average price of each distinct brand listed?

Which brands have the highest prices?

Which ones have the widest distribution of prices?

Is there a typical price distribution (e.g., normal) across brands or within specific brands?

Further processing data would also let you:

Correlate specific product features with changes in price.

You can cross-reference this data with a sample of our Men's Shoe Prices to see if there are any differences between women's brands and men's brands».

Перевод:

«Об этих данных

Это список из 10000 женской обуви и информации об их продукции, предоставленной базой данных продуктов Datafiniti.

Набор данных включает имя обуви, марку, цену и многое другое. Каждая обувь будет иметь запись для каждой найденной цены, а некоторые ботинки могут иметь несколько записей.

Обратите внимание, что это образец большого набора данных. Полный набор данных доступен через Datafiniti.

Что вы можете сделать с этими данными

Вы можете использовать эти данные для определения наценок бренда, ценовых стратегий и тенденций для роскошной обуви. Например.:

Какова средняя цена каждого отдельного бренда в списке?

Какие бренды имеют самые высокие цены?

Какие из них имеют наибольшее распределение цен?

Существует ли типичное распределение цен (например, нормальное) по брендам или по конкретным брендам?

Дальнейшая обработка данных также позволит вам:

Соотнесите конкретные особенности продукта с изменениями в цене.

Вы можете сопоставить эти данные с образцом наших цен на мужскую обувь, чтобы увидеть, есть ли различия между женскими и мужскими брендами».

- Пример строки данных (одной из 33 802):

id	AVpe_eOilAPnD_xSt-H
asins	
brand	Novica
categories	Access.,Clothing,Shoes,Women's Clothing
colors	Purple
count	
dateAdded	2017-02-03T22:06:24Z
dateUpdated	2017-03-28T11:53:51Z

descriptions	[{"dateSeen":["2017-03-26T05:04:46.539Z","2017-02-03T22:06:24Z"],"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":"This handmade creation is offered in partnership with NOVICA, in association with National Geographic.Peru's Alfredo Falcon designs an attractive purple sweater with bell sleeves. Knitted of a fine alpaca wool blend, lace centers the scoop neck while the sleeves and sweater feature a rolled hem.Product Features:Worldstock Country: PeruColors: PurpleFit: ContemporarySleeve type: RolledNeckline: ScoopUnlinedClosure: PulloverMaterials: 45-percent acrylic/40-percent alpaca/15-percent woolCare instructions: Hand wash in cold water, lay flat to dryThis item has been made with natural fibers that are soft to the touchModel: 231516Story Behind the Art:Alfredo Falcon says, 'When I was but a small child I learned from my parents the best weaving techniques. Like me, they had learned from their own parents. This makes us a family entirely devoted to the world of textiles, and we have been for generations. Nowadays I combine knitting techniques, remaining true to Inca and colonial motifs.'What is Worldstock?The handcrafted touch of artisan skill creates variations in color, size and design. If buying two of the same item, slight differences should be expected. Note: Color discrepancies may occur between this product and your computer screen.ImportedPlease allow 10 business days for the product to leave our warehouse and to receive tracking information. You should expect to receive this item within 15 business days."}]
dimension	
ean	
features	[{"key":"Pattern","value":["Solid"]}, {"key":"Worldstock Country","value":["Peru"]}, {"key":"Neck Style","value":["Scoop Neck"]}, {"key":"Sleeve-length","value":["Long Sleeve"]}, {"key":"Front","value":["Flat Front"]}, {"key":"Care Instruction","value":["Hand Wash"]}, {"key":"Closure","value":["Pull Over"]}, {"key":"Dress Length","value":["Short"]}, {"key":"Size","value":["L, M, S, XL"]}, {"key":"Dimensions","value":["See Details"]}, {"key":"Model Number","value":["231516"]}, {"key":"material","value":["Acrylic, Alpaca Wool"]}]
flavors	
imageURLs	
isbn	
keys	handcraftedalpacablendpurplecharismasweaterperu/19238764,novica/231516,handcraftedalpacablendpurplecharismasweaterperubynovica/19238764
manufacturer	
manufacturerNumber	231516
merchants	
name	Handcrafted Alpaca Blend 'Purple Charisma' Sweater (Peru)
prices.amount Min	62.99
prices.amount Max	62.99
prices.availability	

prices.color	
prices.condition	
prices.count	
prices.currency	USD
prices.dateAdded	2017-03-28T11:53:51Z
prices.dateSeen	2017-03-26T05:04:46.539Z
prices.flavor	
prices.isSale	False
prices.merchant	Overstock.com
prices.offer	
prices.returnPolicy	
prices.shipping	
prices.size	L
prices.source	
prices.sourceURLs	https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html
prices.warranty	
quantities	[{"dateSeen":["2017-03-26T05:04:46.539Z","2017-02-08T07:38:48.141Z"],"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":3}]
reviews	[{"date":"2017-01-10T00:00:00.000Z","dateAdded":"2017-03-06T14:47:43Z","dateSeen":["2017-03-02T18:10:15.094Z"],"rating":5.0,"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/customer-reviews.html"],"text":"well made sweater fit true to size","title":"AS","username":"Antoinette M."}]
sizes	
skus	[{"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":"19238764"}]
sourceURLs	https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html,https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/customer-reviews.html
upc	
vin	
websiteIDs	
weight	

Здравоохранение в Ганне

- URL с данными и приведенным описанием [2]:
<https://www.kaggle.com/citizen-ds-ghana/health-facilities-gh>
«Context»

This dataset is provided as part of Citizen Data Science project, to gather & provide fairly clean data (which is a challenge in these regions) to support the data science practice in Ghana and other regions at the beginning of their data science learning curve. So your support is welcome

This dataset provides a listing of healthcare facilities in Ghana, by exploring it we gain new understanding of the country's health infrastructure.

Content

This dataset contains information about health facilities in Ghana organised by Region and District. It also includes the type of health facility and the ownership as well as its geo-location.

Dataset Usecases (are you up to the task? try any of below)

1. Learning/familiarisation with cleaning data and resolving in challenging context of data acquisition.

2. Understanding Ghana's health infrastructure

3. Complex Join of health facilities and tier data

The health facilities data and the tier data are from different sources but we like to join them because they refer to the same facility however this join may not be a simple join because the health facility name in both datasets are not exact string matches.

4. Understanding the level of access to facilities

Combined with population data we want to understand whether some regions or areas are deprived?

Any other creative stuff you can do with this data»

Перевод:

«Контекст

Этот набор данных предоставляется в рамках проекта Citizen Data Science для сбора и предоставления достаточно чистых данных (что является проблемой в этих регионах) для поддержки практики обработки данных в Гане и других регионах в начале их кривой обучения науке о данных. Так что вы поддерживаете приветствуется

Этот набор данных содержит список медицинских учреждений в Гане, изучая его, мы получаем новое понимание инфраструктуры здравоохранения страны. содержание

Этот набор данных содержит информацию о медицинских учреждениях в Гане, организованных по регионам и районам. Он также включает в себя тип медицинского учреждения и владельца, а также его географическое положение.

Варианты использования набора данных (у вас есть задача? Попробуйте любой из нижеприведенных)

1. Изучение / ознакомление с очисткой данных и разрешением в сложных условиях сбора данных.

2. Понимание инфраструктуры здравоохранения Ганы

3. Комплексное соединение медицинских учреждений и уровня данных

Данные о медицинских учреждениях и данные об уровне взяты из разностных источников, но мы хотели бы объединить их, поскольку они ссылаются на одно и то же учреждение, однако это объединение может быть не простым объединением, поскольку имя медицинского учреждения в обоих наборах данных не является точным совпадением строк.

4. Понимание уровня доступа к объектам

В сочетании с данными о населении мы хотим понять, лишены ли некоторые регионы или районы?

Любые другие творческие вещи, которые вы можете сделать с этими данными».

- Пример строки данных (одной из 3775):

Region	Ashanti
District	Offinso North
FacilityName	A.M.E Zion Clinic
Type	Clinic
Town	Afrancho
Ownership	CHAG
Latitude	7.4080100000000001
Longitude	-1.96317

Пиццерии

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/datafiniti/pizza-restaurants-and-the-pizza-they-sell>

«About this Data

This is a list of over 3,500 pizzas from multiple restaurants provided by Datafiniti's Business Database. The dataset includes the category, name, address, city, state, menu information, price range, and more for each pizza restaurant.

Note that this is a sample of a large dataset. The full dataset is available through Datafiniti.

What You Can Do with this Data

You can use this data to discover how much you can expect to pay for pizza across the country. E.g.:

What are the least and most expensive cities for pizza?

What is the number of restaurants serving pizza per capita (100,000 residents) across the U.S.?

What is the median price of a large plain pizza across the U.S.?

Which cities have the most restaurants serving pizza per capita (100,000 residents)?»

Перевод:

«Об этих данных

Это список из более чем 3500 пицц из нескольких ресторанов, предоставленных бизнес-базой данных Datafiniti. Набор данных включает в себя категорию, имя, адрес, город, штат, информацию о меню, ценовой диапазон и многое другое для каждой пиццерии.

Обратите внимание, что это образец большого набора данных. Полный набор данных доступен через Datafiniti.

Что вы можете сделать с этими данными

Вы можете использовать эти данные, чтобы узнать, сколько вы можете заплатить за пиццу по всей стране. Например.:

Какие города наименее и самые дорогие для пиццы?

Каково количество ресторанов, где подают пиццу на душу населения (100 000 жителей) по всей территории США?

Какова средняя цена большой обычной пиццы в США?

В каких городах больше всего пиццы на душу населения (100 000 жителей)?».

- Пример строки данных (одной из 10001):

id	AVz3Y-7h3D1zeR_xDAqm
dateAdded	2017-06-30T05:05:40Z
dateUpdated	2019-05-01T15:43:09Z
address	4203 E Kiehl Ave
categories	Pizza,Restaurant,American restaurants,Pizza Place,Restaurants
primaryCategories	Accommodation & Food Services
city	Sherwood
country	US
keys	us/ar/sherwood/4203ekiehlave/-1051391616
latitude	34.8323
longitude	-92.1838
menuPageURL	http://www.citysearch.com/profile/menu/1550074?singlePlatformId=shotgun-dans-pizza
menus.amountMax	18.21
menus.amountMin	18.21
menus.currency	USD
menus.dateSeen	2018-05-01T04:25:37.197Z,2018-04-16T04:36:02.356Z,2018-02-15T19:58:01.612Z,2018-04-02T23:29:46.353Z,2018-06-28T11:37:25.942Z
menus.description	
menus.name	Cheese Pizza
name	Shotgun Dans Pizza
postalCode	72120
priceRangeCurrency	USD
priceRangeMin	0
priceRangeMax	25
province	AR

Вина

- URL с данными и приведенным описанием [2]:
<https://www.kaggle.com/zynicide/wine-reviews>

«Inspiration

I think that this dataset offers some great opportunities for sentiment analysis and other text related predictive models. My overall goal is to create a model that can identify the variety, winery, and location of a wine based on a description. If anyone has any ideas, breakthroughs, or other interesting insights/models please post them.»

Перевод:

«Я думаю, что этот набор данных предлагает большие возможности для анализа настроений и других прогнозирующих моделей, связанных с текстом. Моя общая цель - создать модель, которая могла бы определять сорт, винодельню и местоположение вина на основе описания».

- Пример строки данных (одной из 129927):

country	US
description	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.
designation	Reserve Late Harvest
points	87
price	14.0
province	Oregon
region_1	Willamette Valley
region_2	Willamette Valley
taster_name	Paul Gregutt
taster_twitter_handle	@paulgwineB
title	Rainstorm 2013 Pinot Gris (Willamette Valley)
variety	Pinot Gris
winery	Rainstorm

Аренда жилья в Берлине

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/brittabetendorf/berlin-airbnb-data>

«Context

Airbnb has successfully disrupted the traditional hospitality industry as more and more travelers decide to use Airbnb as their primary accommodation provider. Since its beginning in 2008, Airbnb has seen an enormous growth, with the number of rentals listed on its website growing exponentially each year. In Germany, no city is more popular than Berlin. That implies that Berlin is one of the hottest markets for Airbnb in Europe, with over 22,552 listings as of November 2018.

Content

The datasets were scraped on November 07th, 2018 and contain detailed listings data, review data and calendar data of current Airbnb listings in Berlin.

Acknowledgements

This data was created by Murray Cox and his Inside Airbnb project which can be found here.

Inspiration

Can you predict the price for each Berlin neighborhood using listing descriptions? What are the busiest times of the year to visit Berlin? By how much do prices spike? Can you uncover trends in reviews of Airbnb visitors to Berlin?»

Перевод:

«Контекст

Airbnb успешно разрушил традиционную индустрию гостеприимства, так как все больше и больше путешественников решают использовать Airbnb в качестве основного поставщика жилья. С начала своего существования в 2008 году Airbnb пережил огромный рост, причем количество арендных плат, перечисленных на его веб-сайте, растет в геометрической прогрессии с каждым годом. В Германии нет более популярного города, чем Берлин. Это означает, что Берлин является одним из самых популярных рынков для Airbnb в Европе, с 22,552 списками на ноябрь 2018 года.

содержание

Наборы данных были очищены 7 ноября 2018 года и содержат подробные данные о списках, обзорные данные и календарные данные о текущих списках Airbnb в Берлине.

Подтверждения

Эти данные были созданы Мюрреем Коксом и его проектом Inside Airbnb, которую можно найти здесь.

возможные вопросы

Можете ли вы предсказать цену для каждого района Берлина, используя описания списков?

Какое время года для посещения Берлина? На сколько цены растут?

Можете ли вы раскрыть тенденции в обзорах посетителей Airbnb в Берлине?».

- Пример строки данных (одной из 401964):

listing_id	2015
id	69544350
date	11.04.2016
reviewer_id	7178145
reviewer_name	Rahel
comments	Mein Freund und ich hatten gute gemГјtliche vier NГјchte in Jans Studio. Es ist unfassbar ruhig und hat sogar ein gedГјmpftes Klavier fГјr Leute die gerne spielen. Die Lage ist hervorragend und es gibt in der Strasse einen tollen alimentari und eine gГјnstige weinerei. Das Bad ist sehr alt aber funktionstГјchtig und der insgesamte Standard ist ok, ein wenig chaotisch aber dafГјr hat man das GefГјhl man ist bei jemandem zuhause ! Jan ist ein aufgeschlossener lustiger Mann und auch um frische HandГјcher und sonstiges bemГјht! Wir danken.

Японский фаст-фуд

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/residentmario/ramen-ratings>

«Context

The Ramen Rater is a product review website for the hardcore ramen enthusiast (or "ramenphile"), with over 2500 reviews to date. This dataset is an export of "The Big List" (of reviews), converted to a CSV format.

Content

Each record in the dataset is a single ramen product review. Review numbers are contiguous: more recently reviewed ramen varieties have higher numbers. Brand, Variety (the product name), Country, and Style (Cup? Bowl? Tray?) are pretty self-explanatory. Stars indicate the ramen quality, as assessed by the reviewer, on a 5-point scale; this is the most important column in the dataset!

Note that this dataset does *not* include the text of the reviews themselves. For that, you should browse through <https://www.theramenrater.com/> instead!

Acknowledgements

This dataset is republished as-is from the original BIG LIST on <https://www.theramenrater.com/>.

Inspiration

What ingredients or flavors are most commonly advertised on ramen package labels? How do ramen ratings compare against ratings for other food products (like, say, wine)?

How is ramen manufacturing internationally distributed?»

Перевод:

«Контекст

Ramen Rater - это веб-сайт с обзором продуктов для хардкорных энтузиастов раменов (или «ramenphile»), на сегодняшний день более 2500 обзоров. Этот набор данных является экспортом «Большого списка» (обзоров), преобразованного в формат CSV.

Содержание:

Каждая запись в наборе данных представляет собой отдельный обзор продукта рамен. Числа обзора являются смежными: более поздние исследованные сорта рамена имеют более высокие числа. Бренд, Разнообразие (название продукта), Страна и Стиль (Кубок? Чаша? Поднос?) Говорят сами за себя. Звезды показывают качество рамена, оцениваемое рецензентом, по 5-балльной шкале; это самый важный столбец в наборе данных!

Обратите внимание, что этот набор данных не включает в себя текст самих обзоров. Для этого вы должны просмотреть <https://www.theramenrater.com/> вместо этого!

Подтверждения

Этот набор данных переиздан как есть из оригинального БОЛЬШОГО СПИСКА на <https://www.theramenrater.com/>.

Возможные вопросы:

Какие ингредиенты или ароматизаторы чаще всего рекламируются на этикетках упаковки рамен?

Как рейтинги рамэн сравниваются с оценками других продуктов питания (например, вина)?

Как производство рамэна распределяется на международном уровне?»

- Пример строки данных (одной из 2581):

Review #		1350
Brand	Mamee	
Variety	Chef Curry Laksa Flavour	
Style	Pack	
Country	Malaysia	
Stars		5
Top Ten	2014 #7	

Игроки FIFA-18

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>

«Possible Explorations

Make your dream team

Analyse which Club or National Team has the best-rated players

Assess the strength of a team at a particular position

Analyse the team with the best dribbling speed

Co-relate between Age and Overall rating

Co-relate between Age and Nationality

Co-relate between Age and Potential»

Перевод:

«Возможные исследования

Сделай команду своей мечты

Проанализируйте, в каком клубе или сборной лучшие игроки

Оценить силу команды на определенной позиции

Проанализируйте команду с лучшей скоростью дриблинга

Сопоставить возраст и общий рейтинг

Взаимосвязь между возрастом и национальностью

Взаимосвязь между возрастом и потенциалом»

- Пример строки данных (одной из 17 982):

Name	Cristiano Ronaldo	
Age		32
Photo	https://cdn.sofifa.org/48/18/players/20801.png	
Nationality	Portugal	
Flag	https://cdn.sofifa.org/flags/38.png	
Overall		94
Potential		94
Club	Real Madrid CF	
Club Logo	https://cdn.sofifa.org/24/18/teams/243.png	
Value	в, -95.5M	
Wage	в, -565K	
Special		2228
Acceleration		89

Aggression		63
Agility		89
Balance		63
Ball control		93
Composure		95
Crossing		85
Curve		81
Dribbling		91
Finishing		94
Free kick accuracy		76
GK diving		7
GK handling		11
GK kicking		15
GK positioning		14
GK reflexes		11
Heading accuracy		88
Interceptions		29
Jumping		95
Long passing		77
Long shots		92
Marking		22
Penalties		85
Positioning		95
Reactions		96
Short passing		83
Shot power		94
Sliding tackle		23
Sprint speed		91
Stamina		92
Standing tackle		31
Strength		80
Vision		85
Volleys		88
CAM	89.0	
CB	53.0	
CDM	62.0	
CF	91.0	
CM	82.0	
ID		20801
LAM	89.0	
LB	61.0	
LCB	53.0	
LCM	82.0	
LDM	62.0	
LF	91.0	
LM	89.0	
LS	92.0	
LW	91.0	

LWB	66.0
Preferred Positions	ST LW
RAM	89.0
RB	61.0
RCB	53.0
RCM	82.0
RDM	62.0
RF	91.0
RM	89.0
RS	92.0
RW	91.0
RWB	66.0
ST	92.0

Список используемых источников

1. Хабр. <https://habr.com/ru/post/248395/>
2. Kaggle. <https://www.kaggle.com>
3. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. – Ульяновск : УлГТУ, 2017. – 290 с.

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**В. В. Воронина, А. В. Михеев,
Н. Г. Ярушкина, К. В. Святков**

ТЕОРИЯ И ПРАКТИКА МАШИННОГО ОБУЧЕНИЯ

Учебное пособие

Ульяновск
УлГТУ
2017

УДК 004.8 (075)
ББК 32.973-018.1я7
В12

Рецензенты:

Главный научный сотрудник ФНЦП АО «НПО «Марс»,
д-р техн. наук Токмаков Г. П.

Начальник расчетно-теоретического отдела ОАО «Ульяновское конструкторское бюро приборостроения», канд. техн. наук Сорокин М. Ю.

*Утверждено редакционно-издательским советом университета
в качестве учебного пособия*

Воронина, Валерия Вадимовна

В12 Теория и практика машинного обучения : учебное пособие /
В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. –
Ульяновск : УлГТУ, 2017. – 290 с.

ISBN 978-5-9795-1712-4

Учебное пособие рассматривает вопросы, связанные с анализом данных: модели, алгоритмы, методы и их реализацию на языке Python. Особое внимание уделено анализу временных рядов. Книга предназначена для студентов группы направлений 09, а также для студентов других групп направлений, изучающих дисциплины, связанные с разработкой приложений в области анализа данных, в том числе TimeSeriesDataMinig и DataMining.

**УДК 004.8 (075)
ББК 32.973.2-018.2я7**

© Воронина В. В., Михеев А. В.,
Ярушкина Н. Г., Святков К. В., 2017

© Оформление. УлГТУ, 2017

ISBN 978-5-9795-1712-4

© Дизайн обложки. Пермовская А. А., 2017

СОДЕРЖАНИЕ

Введение	7
О задачах машинного обучения	9
Пространство признаков	13
Формальное определение понятия «обучение»	14
Общий алгоритм машинного обучения	16
Типы задач машинного обучения	17
Способы обучения и оценки его качества	21
Типовые задачи при подготовке данных и обучении моделей	28
Учет пропусков	29
Кодирование нечисловых признаков	31
Приведение данных к единому масштабу и стандартизация	33
Разметка данных	35
Переобучение	36
Модели и алгоритмы машинного обучения	44
Методы теории вероятностей	47
Деревья решений	52
Статистические модели и методы	56
Модели и методы нечеткой логики	67
Нечеткие множества	68
Лингвистические переменные	72
Операции нечеткой логики	74
Нечеткие системы	79
Нечеткая логика в анализе временных рядов	85
Метод моделирования нечетких временных рядов	88
Пример моделирования временного ряда в нечетком подходе	89
Извлечение знаний из временных рядов	94

Нечеткое сглаживание временного ряда	99
Нечеткая регрессия	104
ACL-шкала и нечеткая кластеризация объектов	106
Искусственные нейронные сети	111
Особенности нейронных сетей.....	111
Определение модели искусственной нейронной сети	113
Первая формальная модель и первая реализация нейронной сети.	115
Многослойный персептрон (MLP).....	119
Сверточные (ConvolutionalNeuralNet) и Глубокие (DeepNet) Сети	122
Карты (ART, SFAM).....	126
Рекуррентные сети (Recurrent Neural Network)	131
Самоорганизующиеся карты (Self-organization map, SOM).....	132
Автокодировщики (AutoEncoder)	134
Импульсные (Спайковые) сети	136
Причины бурного развития ИНС сегодня.....	138
Борьба с переобучением в ИНС	139
Обратное распространение ошибки.....	140
Нечеткие нейронные сети	148
Генетические алгоритмы.....	156
Нечеткие системы с генетической настройкой	160
Нечеткие нейронные сети с генетическим проектированием.....	162
Генетическая оптимизация F-преобразования временных рядов ..	163
Разработка приложений в сфере машинного обучения	165
Основы работы с Python.....	167
Элементарные операции с данными	172
Работа с DataFrame.....	182
Предобработка данных. Стандартизация и нормализация.....	185
Работа с деревьями решений	188
Работа с линейной регрессией.....	191

Сохранение и загрузка обученной модели	197
Работа с логистической регрессией	200
Решение задачи ранжирования признаков	205
Работа с полиномиальной регрессией	213
Работа с простейшими моделями нейронных сетей	217
Реализация алгоритма обучения нейронной сети	223
Регуляризация и сеть прямого распространения	228
Работа с библиотеками Keras и Theano. Настройка под Windows.....	236
Получение данных средствами Keras	240
Создание и обучение модели сверточной сети.....	241
Загрузка и сохранение сложных моделей	246
Рекуррентные сети для прогнозирования временных рядов.....	247
Контрольные вопросы и тестовые задания.....	251
Тест «Общие сведения о машинном обучении».....	251
Проблема переобучения.....	252
Регрессия.....	253
Модели и методы нечеткой логики.....	253
Нечеткие временные ряды	254
Нечеткая регрессия	255
Генетические алгоритмы.....	255
Нечеткая кластеризация	256
Искусственные нейронные сети и глубинное обучение.....	256
Тест «Искусственные нейронные сети»	256
Практические задания	259
Работа с файлом данных Титаника	259
Работа по отбору признаков	260
Многослойный персептрон.....	260

Реализация алгоритма обратного распространения ошибки.....	261
Регуляризация и сеть прямого распространения.....	261
Сравнение эффективности моделей из библиотеки Keras.....	263
Работа с библиотекой OpenCV.....	265
Нечеткая логика.....	267
Генетические алгоритмы.....	269
Нечеткая кластеризация объектов.....	270
Анализ временных рядов.....	272
Работа с рекуррентными сетями.....	274
Заключение.....	277
Глоссарий.....	278
Предметный указатель.....	282
Библиографический список.....	283

ВВЕДЕНИЕ

На протяжении всего периода своего развития человечество постоянно ищет способы автоматизации тех или иных видов деятельности, в итоге перекладывая на «плечи» механизмов все более обширные и сложные обязанности. Изначально мечты о роботах, выполняющих всю работу, находили воплощение лишь в литературных произведениях, однако с развитием технологий им открылся путь и в реальность. Сейчас уже не столь фантастичными звучат слова песни «До чего дошел прогресс» (к/ф «Приключения Электроника»): «До чего дошел прогресс – труд физический исчез, да и умственный заменит механический процесс». Ведь эти строки достаточно полно отражают суть такой дисциплины, как «машинное обучение», благодаря которой их воплощение в жизнь становится возможным. Однако стоит заметить, что до полной реализации описанного в песне еще очень далеко: пока машинное обучение не способно заменить умственный труд, и в ближайшее время вряд ли сможет. Но шаги в этом направлении активно совершаются.

С теоретической стороны машинное обучение – дисциплина, находящаяся на пересечении математической статистики, численных методов оптимизации, теории вероятностей, а также дискретного анализа. С помощью ее методов происходит решение задачи извлечения знаний из данных, которой занимается еще только формирующаяся область «Интеллектуальный анализ данных» (DataMining). Согласно Википедии [1]: «В теории искусственного интеллекта и экспертных систем, знание – это совокупность утверждений о мире, свойствах объектов, закономерностях процессов и явлений, а также правил логического вывода одних утверждений из других и правил использования их для принятия решений. Главное отличие знаний от данных состоит в их структурности и активности: появление в базе знаний новых фактов или установление новых связей между ними может стать источником изменений в принятии решений». Если,

опираясь на это определение знания, рассмотреть DataMining как процесс, то его сутью будет нахождение в «сырых» данных знаний, обладающих свойствами нетривиальности, интерпретируемости и практической полезности для принятия решений в различных сферах деятельности человека. Причем, эти знания ранее неизвестны.

С практической же стороны машинное обучение нацелено на создание систем, способных адаптироваться к решению различных задач без явного кодирования алгоритма, то есть систем, способных обучаться.

В данной книге рассмотрены основные модели и алгоритмы машинного обучения, в том числе для анализа временных рядов. Наряду с этим представлены и их практические реализации на языке Python. В последних разделах книги обучающемуся предлагаются контрольные вопросы по пройденным темам, а также задачи для выполнения, с помощью которых он сможет проверить себя и закрепить полученные навыки.

Книга подготовлена при реализации проекта №2.4760.2017/8.9 «Исследование и разработка моделей, методов и алгоритмов гибридизации нечетких предметных онтологий, логического вывода и интеллектуального анализа временных рядов» в рамках государственного задания Минобрнауки Российской Федерации.

О ЗАДАЧАХ МАШИННОГО ОБУЧЕНИЯ

Как было сказано выше, машинное обучение нацелено на создание систем, способных получать знания из данных, а также способных с помощью обучения улучшать показатели своей работы. Таким образом, можно утверждать, что машинное обучение является одной из областей науки о данных (DataScience). На рисунке 1 представлены ее составляющие и показано место машинного обучения среди них.



Рисунок 1. Области науки о данных

Сложность работы в сфере DataMining обуславливается неточностью, противоречивостью, разнородностью, неполнотой данных, которые при этом могут иметь гигантские объемы. Для их обработки требуются специальные программные средства: инструменты преобразования «сырых» данных в информацию, а информацию в знания. Кроме того, алгоритмы обработки данных должны иметь возможность обучаться по прецедентам. Машинное обучение (МО) как раз и занимается разрешением подобных сложностей. Причем в настоящее время усилия сконцентрированы на повышении

не столько качества решения, сколько скорости и оптимальности его получения.

Как видно из рисунка 1, машинное обучение пересекается с анализом данных, то есть имеет общие черты, но также и свою специфику. Анализ данных – это огромная и уже относительно не молодая область математики и информатики. Под нее попадает всяческая обработка данных. Например, обработка данных физического эксперимента статистическими методами для получения обобщенной информации (усредненных значений, точных значений, доверительных интервалов, оценок и тому подобное), обработка сигнала методом Фурье, даже подсчет минимального и максимального значений или отношения каких-то величин может быть анализом данных, если эти операции влекут за собой важные выводы, знания о данных и позволяют решить поставленную задачу. Так получение дисперсии некой величины является анализом данных, но не является в чистом виде машинным обучением. Хотя важно отметить, что методы статистики имеют большое значение и в машинном обучении, о чем мы еще будем говорить. Отличительной особенностью методов машинного обучения является то, что они способны решать широкий спектр задач без явного кодирования алгоритма решения. Например, метод Регрессии входит в машинное обучение. Регрессия давно применяется в экономике для прогнозирования и оценки разнообразных параметров, потому что данная математическая модель в некотором роде универсальна. Однако стоит заметить, что регрессия появилась в математике гораздо раньше, чем ее стали использовать в экономике, а вот машинным обучением это стало только тогда, когда все эти методы начали рассматриваться не как обособленные механизмы решения задач из специальных областей, а как методы преобразования информации вообще.

На рисунке 2 представлена диаграмма распределения задач, которые приходится решать в сфере машинного обучения. Необходимо отметить, что соотношение областей на данной диаграмме

отражает лишь субъективное видение отрасли авторами на момент написания книги.



Рисунок 2. Типы задач машинного обучения

Понятно, что разработка программ в области машинного обучения весьма дорога. Поэтому применять его следует лишь в случаях, когда это действительно необходимо. В таблице 1 представлены некоторые признаки, по которым можно принять решение о необходимости применения к задаче методов машинного обучения.

Таблица 1. Ключевые характеристики решаемых задач

Целесообразно применение классических методов	Целесообразно применение методов машинного обучения
Наличие четко формализованного алгоритма (например, поиск клиентов, которые тратят больше всего денежных средств в месяц или чаще всего покупают. Так называемая RFM сегментация).	Отсутствие четко формализованного алгоритма решения ввиду высокой вариативности решения (например, распознавание рукописного текста).
Не допускается неопределенность поведения модели (например, расчет местоположения самолета по строго заданным формулам).	Допускается неопределенность поведения модели (например, предварительный контроль качества детали, где допускается определенная доля брака).

Целесообразно применение классических методов	Целесообразно применение методов машинного обучения
Экономическая целесообразность (например, методы МО дают 98% качества, классическое решение дает 90% качества, что является приемлемым, но решение на основе МО в 2 раза дороже и реализуется дольше). Поэтому здесь выбирается приоритетный критерий.	

Кроме того, для разработки решений на основе МО необходимы оцифрованные данные. Они являются ключевым аспектом, от которого зависит качество и полнота решения. Таким образом, очень важно понимать, что для применения методов машинного обучения требуются собранные и специально подготовленные данные.

Как правило, процесс машинного обучения проходит несколько этапов. На рисунке 3 представлен обобщенный алгоритм, выполняющийся при решении задач методами рассматриваемой области.

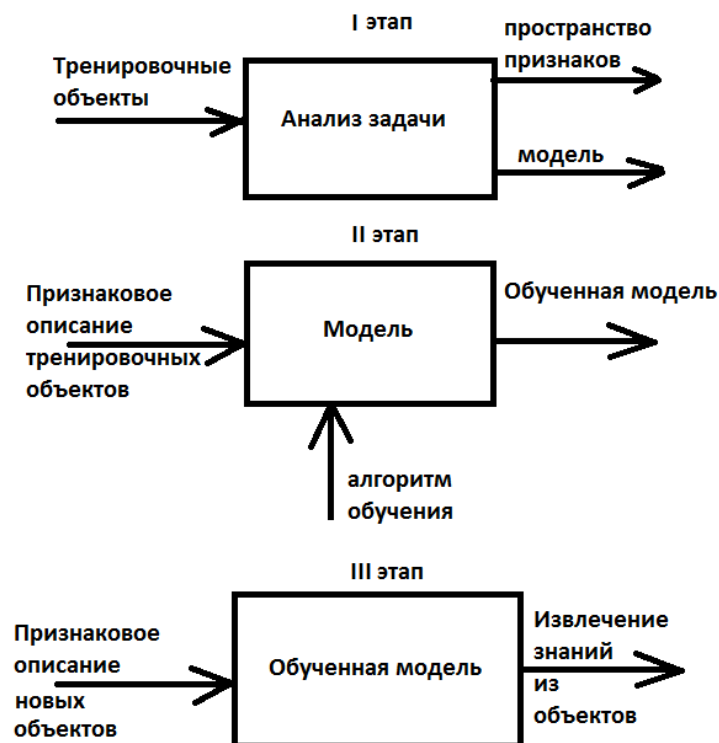


Рисунок 3. Этапы машинного обучения

Необходимо отметить, что данная иллюстрация отражает идеальный случай, когда после обучения мы сразу получили работающую модель. Однако в реальности, как правило, между вторым и третьим этапом есть промежуточное, но очень важное действие – оценка качества модели. И именно по его результатам принимается решение о переходе на следующий этап или о возврате к одному из предыдущих. Этот вопрос мы рассмотрим подробнее далее, здесь же введем определение пространства признаков, для лучшего понимания схемы с рисунка 3.

Пространство признаков

Пространство признаков – это N -мерное пространство, где N – число измеряемых характеристик объектов, выделенное для конкретной задачи. При этом объекты в пространстве признаков задаются N -мерными векторами, каждая компонента которых представляет собой значение определенной характеристики.

Рассмотрим пример. Пусть необходимо разработать функцию для сайта салона красоты, которая будет предлагать клиентам определенные косметические процедуры. Понятно, что они будут разными для разных возрастных и гендерных групп. То есть в этой задаче ключевыми будут характеристики пола и возраста людей. Графическое отображение пространства признаков, а также заданные в нем объекты A и B показаны на рисунке 4. В векторном виде объект A с характеристиками «пол» = «женский» и «возраст» = 40 будет задан как: $A = \{40, ж\}$. Объект B с возрастом 50 и мужским полом – $B = \{50, м\}$.

Как видно из рассмотренного примера, признаки объектов могут быть разными по своей природе. Существует следующая типизация:

1. Бинарные: $F \in \{0, 1\}$.
2. Номинальные: $|F| < \infty$.
3. Номинальные, упорядоченные: $|F| < \infty, F_i < F_{i+1} \dots < F_{i+n}$.
4. Количественные: $F \in \mathcal{R}$.

Каждый тип признаков обрабатывается по-разному, о чем скажем далее. Пока же продолжим разговор о других аспектах рассматриваемой дисциплины. Как неоднократно было сказано, работа в сфере машинного обучения нацелена на создание систем, способных обучаться. Рассмотрим формальное определение понятия «обучение».

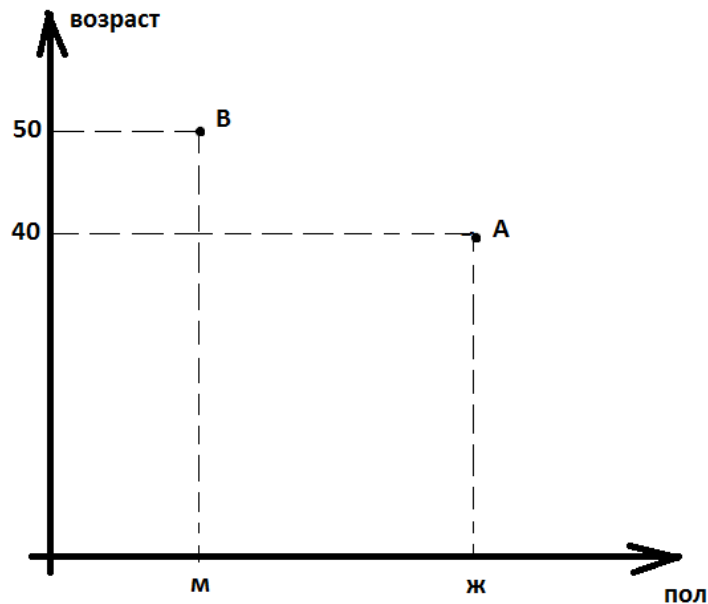


Рисунок 4. Объекты в пространстве признаков

Формальное определение понятия «обучение»

Формально понятие «обучение» специалисты определяют так: «Пусть есть некое множество задач класса C , для которого задано некоторое множество опыта EX и определена мера качества L . Тогда о наличии обучения на опыте EX относительно класса задач C в смысле меры качества L можно говорить в том случае, если при предъявлении нового опыта EX' возрастает качество решения задачи класса C , измеряемое мерой L ».

Существуют два основных типа обучения: с учителем и без учителя. Более подробно о них будет рассказано далее, здесь же рассмотрим формальную постановку задачи обучения для самого первого (и самого простого) типа – обучения с учителем.

Пусть X – некоторое множество объектов, по которым необходимо получить некоторое множество ответов Y . При этом предполагается, что существует некоторая зависимость $y: X \rightarrow Y$. Тогда задача обучения будет формулироваться так: дано $\{x_1 \dots x_n\} \in X$ – обучающая выборка, и $\{y_1 \dots y_n\}$ – известные ответы, причем $y_i = y(x_i)$. Найти: $a: X \rightarrow Y$ – решающую функцию, приближающую y на всем множестве X .

Данная задача в свою очередь порождает ряд подзадач:

1. Определение способа задания объектов;
2. Определение способа задания ответов;
3. Определение способа построения функции a ;
4. Определение способа приближения для a и y .

Все эти подзадачи решаются последовательно, и результаты решения предыдущей, как правило, влияют на решение текущей. Поэтому самой важной среди них будет определение способа описания данных. Можно выделить следующие типы входных данных при обучении:

1. Координаты объектов в пространстве признаков;
2. Временной ряд;
3. Сигнал;
4. Изображение;
5. Видеоряд;
6. Описание взаимоотношений между объектами.

На основе описанного выше введем формальное описание приведенных на рисунке 3 второго и третьего этапов машинного обучения.

Пусть $X = \{x_1 \dots x_k\}$ – исследуемые объекты. $F = \{f_1 \dots f_N\}$ – пространство признаков, в котором эти объекты будут рассматриваться. При этом $f_i(x_j)$, $i \in [1 \dots N]$, $j \in [1 \dots k]$ – значение i -го признака j -го объекта. Тогда признаковое описание тренировочных объектов задается в виде матрицы

$$D = \| \| f_i(x_j) \| \|_{k \times N} = \begin{pmatrix} f_1(x_1) & \dots & f_N(x_1) \\ f_1(x_k) & \dots & f_N(x_k) \end{pmatrix}.$$

Этап обучения может быть формализован следующим образом. Метод обучения $\mu: (X \times Y)_k \rightarrow A$ по выборке $XY_k = (x_i, y_i), i \in [1 \dots k]$ строит алгоритм $a = \mu(XY_k)$:

$$\begin{pmatrix} f_1(x_1) & \dots & f_N(x_1) \\ f_1(x_k) & \dots & f_N(x_k) \end{pmatrix} \xrightarrow{y} \begin{pmatrix} y_1 \\ y_k \end{pmatrix} \xrightarrow{\mu} a$$

Этап применения обученной модели формализуется следующим образом. Алгоритм a для новых объектов $X' = \{x'_1 \dots x'_k\}$ выдает ответы $y'_i = a(x'_i), i \in [1 \dots k]$:

$$\begin{pmatrix} f_1(x'_1) & \dots & f_N(x'_1) \\ f_1(x'_k) & \dots & f_N(x'_k) \end{pmatrix} \xrightarrow{a} \begin{pmatrix} y'_1 \\ y'_k \end{pmatrix}.$$

Общий алгоритм машинного обучения

С понятием обучения тесно связано понятие обобщающей способности. Обобщающая способность – это свойство модели отражать исходные данные в требуемые результаты ($X \rightarrow Y$) на всем множестве исходных данных (во всех сценариях, а не только на тренировочных примерах). Величину обобщения оценивают через обратную величину – отклонение или ошибку. Ошибка – это численно выраженная разница между ответом модели и требуемым (реальным) значением. В более общем смысле обобщающая способность – способность модели найти некий «закон природы», который будет описывать неизвестную нам и скрытую взаимосвязь входных и выходных данных. Таким образом, опираясь на понятие обобщающей способности, можно выделить основные проблемные вопросы машинного обучения:

1. Достаточно ли данных для нахождения в них полезных знаний?
2. Может ли в принципе данная модель обучиться на имеющихся данных?
3. Будет ли полученная модель приближать требуемый «закон природы» на всем возможном множестве X ?

4. Насколько хорошо будет работать обученная модель или насколько часто и насколько сильно будет ошибаться модель на реальных (контрольных) данных?

По сути в машинном обучении решается задача приближения алгоритма к некоторому «закону природы». С математической точки зрения задача приближения является задачей оптимизации или минимизации ошибки:

$$E=|y - y'| \rightarrow \min .$$

Таким образом, общий алгоритм решения задач в сфере машинного обучения будет состоять из следующих шагов:

1. Понимание задачи и исходных данных;
2. Формулировка решения на математическом языке (на этом шаге важно понять, что задача формализуема, а результаты работы модели могут быть проверены);
3. Предобработка данных и выделение ключевых признаков;
4. Построение модели (или моделей);
5. Обучение модели (моделей) и оценка качества;
6. Эксплуатация модели при достижении требуемого качества, либо возврат к одному из предыдущих шагов (перенастройка модели, добыча новых данных и т. п.).

Типы задач машинного обучения

Машинное обучение применяется для решения задач в следующих областях (основные сферы применения):

1. Медицинская диагностика.
2. Техника, в частности:
 - 2.1. Автоматизация и управление.
 - 2.2. Техническая диагностика.
 - 2.3. Робототехника.
 - 2.4. Компьютерное зрение.
 - 2.5. Распознавание речи.
3. Экономика, в частности:
 - 3.1. Кредитный скоринг (credit scoring).

- 3.2. Предсказание ухода клиентов (churn prediction).
- 3.3. Обнаружение мошенничества (fraud detection).
- 3.4. Биржевой технический анализ (technical analysis).
- 3.5. Биржевой надзор (market surveillance).
- 4. Офисная автоматизация, в частности:
 - 4.1. Распознавание текста.
 - 4.2. Обнаружение спама.
 - 4.3. Категоризация документов.
 - 4.4. Распознавание рукописного ввода.

Если же говорить об обобщенных типах задач машинного обучения, то можно выделить следующие:

1. Регрессия (или иногда встречается термин «аппроксимация»);
2. Классификация;
3. Кластеризация.

Помимо указанных видов существуют и другие, но остановимся на обозначенных, так как они наиболее распространены. Рассмотрим формальную постановку этих задач.

Задача регрессии – приближение неизвестной целевой зависимости на некотором множестве данных. Пусть X – множество данных – описаний некоторых объектов. Y – множество возможных ответов для X .

В задаче регрессии предполагается, что существует неизвестная целевая зависимость $y: X \rightarrow Y$, чьи значения известны только на объектах обучающей выборки $XY = \{(x_1, y_1) \dots (x_n, y_n)\}$, $x \in X$, $y \in Y$. Необходимо получить алгоритм $a: X \rightarrow Y$, приближающий целевую зависимость как на множестве XY , так и на X . То есть решить задачу регрессии – значит найти алгоритм, обладающий способностью к обобщению эмпирических фактов (способностью к выводу общих знаний из частных наблюдений, прецедентов).

Задача классификации – распределение некоторого множества объектов по заданному множеству групп (классов). При этом есть некоторое подмножество объектов, для которых распределение по классам известно, классовая принадлежность остальных – неизвестна.

Требуется построить алгоритм, который указывал бы классовую принадлежность для любого объекта из исходного множества.

Формально постановку задачи классификации можно описать следующим образом. Пусть X – множество данных – описаний некоторых объектов. Y – конечное множество классов, отмеченных метками. Существует неизвестная целевая зависимость – отображение $y: X \rightarrow Y$, чьи значения известны только на объектах обучающей выборки $XU = \{(x_1, y_1) \dots (x_n, y_n)\}$, $x \in X$, $y \in Y$. Необходимо получить алгоритм $a: X \rightarrow Y$, способный классифицировать произвольный объект $x \in X$.

Как можно заметить, данная задача схожа с предыдущей. Однако главная особенность задачи регрессии заключается в том, что функция $a: X \rightarrow Y$ является непрерывной вещественной функцией. Задача классификации отличается от этого тем, что Y – дискретное множество. Кроме того, в отличие от задачи аппроксимации у задачи классификации выделяют несколько типов. По количеству классов можно выделить:

1. Классификацию на два класса: множество Y содержит всего две метки.
2. Классификацию на множество классов: Y содержит от трех до нескольких тысяч меток.

По характеру разделения объектов на классы можно выделить:

1. Классификацию на непересекающиеся классы: один объект принадлежит только одному классу.
2. Классификацию на пересекающиеся классы: один объект может принадлежать нескольким классам.
3. Классификацию на нечеткие множества: объект принадлежит всем классам с определенной степенью принадлежности.

Задача кластеризации – разделение некоторого множества объектов на непересекающиеся группы (кластеры) таким образом, чтобы каждая группа состояла из схожих объектов, а объекты разных кластеров существенно отличались.

Формально постановку задачи кластеризации можно описать следующим образом. Пусть X – множество данных – описаний некоторых объектов. Y – множество кластеров, отмеченных метками. Определена функция расстояния между объектами из исходного множества X : $f(x, x')$, и есть некоторая обучающая выборка объектов $X_0 = \{x_1 \dots x_n\}$, $x \in X$. Необходимо разбить обучающую выборку на кластеры, приписав каждому x номер кластера y_i , так, чтобы близкие по метрике f объекты принадлежали одному кластеру, а объекты разных кластеров существенно отличались по метрике f . То есть необходимо построить алгоритм $a: X \rightarrow Y$, который любому $x \in X$ ставит в соответствие номер кластера $y \in Y$. Причем, иногда множество Y известно заранее, но чаще все-таки ставится задача получить оптимальное число кластеров, исходя из характера данных. Оптимальность оценивается по какому-либо критерию качества кластеризации.

Задача кластеризации сложнее аппроксимации и классификации. Это обусловлено следующими причинами:

1. Нет однозначного критерия качества кластеризации. Существует ряд эвристических критериев, а также ряд бескритериальных алгоритмов, выполняющих вполне осмысленную кластеризацию, но дающих на одних и тех же данных разные результаты.
2. Число кластеров, как правило, заранее неизвестно и задается субъективно.
3. На результат кластеризации существенное влияние оказывает выбранная метрика расстояния, которая, как правило, также выбирается субъективно.

Однако, несмотря на описанные выше сложности, кластеризация помогает достичь следующие цели:

1. Улучшить понимание данных за счет выявления их кластерной структуры: разбиение объемной выборки на группы схожих объектов может упростить дальнейшую обработку

данных за счет применения к каждому кластеру своих методов анализа.

2. Осуществить сжатие данных. В данном случае подразумевается сокращение объемной выборки за счет работы с наиболее яркими представителями кластеров (групп схожих объектов).
3. Выявить новизну в массиве данных: обнаружить нетипичные объекты, которые не удастся отнести ни к одному кластеру.
4. Решить задачу таксономии: построить древообразную иерархическую структуру, упорядочивающую исходные данные. Ее построение достигается за счет дробления крупных кластеров на более мелкие, которые в свою очередь также дробятся на еще более мелкие. Визуально таксономия отображается в виде графика – дендрограммы.

Способы обучения и оценки его качества

Как было сказано выше, основная характеристика систем, разрабатываемых с помощью методов машинного обучения, – способность к обучению. В зависимости от видов решаемых задач применяют различные алгоритмы реализации этой ключевой особенности. В рамках данного пособия рассмотрим три основных вида обучения, а также определим классы задач, подходящие для каждого из этих видов.

Первый тип – обучение с учителем. Формально он был описан ранее, поэтому здесь не будем на этом останавливаться подробно. Вкратце же обучение с учителем можно описать следующим образом: дано некоторое множество объектов и множество возможных реакций системы на эти объекты. При этом ответы и объекты связаны между собой некоторой неизвестной зависимостью. Есть конечная совокупность пар объект-ответ (прецедентов), называемая обучающей выборкой. На ее основе необходимо выявить алгоритм, который впоследствии для любого объекта из исходного множества даст достаточно точный ответ. Для измерения точности ответов используется один

из функционалов качества, как правило, завязанный на вычислении отклонения полученного ответа от ожидаемого, то есть вычислении ошибки. Рассмотрим некоторые их виды. В приведенных ниже формулах используются следующие обозначения: $XU = \{(x_1, y_1) \dots (x_n, y_n)\}$ – обучающая выборка, n – количество прецедентов, y_i – фактическое значение (ожидаемый ответ) в i -м прецеденте ($y_i \in XU$), \hat{y}_i – выданный системой ответ для x_i ($x_i \in XU$).

1. Средняя ошибка представляет собой усреднение ошибок для каждого образца и вычисляется по формуле

$$CO = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n}.$$

2. Средняя абсолютная ошибка представляет собой усреднение абсолютных ошибок на каждом шаге и вычисляется по формуле

$$CAO = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}.$$

3. Среднеквадратическая ошибка вычисляется как сумма средних квадратов ошибок. Формула:

$$CKO = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}.$$

4. Корень из среднеквадратической ошибки вычисляется по формуле

$$KCKO = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}.$$

5. Средняя относительная ошибка вычисляется как среднее относительных ошибок:

$$CO = \frac{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)}{y_i}}{n} \times 100\%.$$

6. Средняя абсолютная относительная ошибка вычисляется как среднее относительных ошибок по модулю:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|}{n} \times 100\% .$$

7. Симметричная средняя абсолютная относительная ошибка вычисляется как

$$SMAPE = \frac{\sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{(y_i + \hat{y}_i)/2} \right|}{n} \times 100\% .$$

У всех представленных мер качества есть свои достоинства и недостатки. Например, у первой и пятой недостаток заключается в том, что положительные и отрицательные ошибки аннулируют друг друга, поэтому в некоторых случаях они не являются достаточно хорошими индикаторами качества. В связи с этим чаще всего используется третья или четвертая меры.

Обучение с учителем используется при решении задач аппроксимации и классификации. В первом случае ответы являются действительными числами или векторами, во втором – выбираются из конечного множества меток-классов. Необходимо отметить, что приведенные выше формулы подходят только для случаев, когда ответ системы и требуемый ответ – действительные числа, получаемые при решении задачи аппроксимации. В задачах классификации же оценка качества чаще всего завязана на соотношении количеств правильно и неправильно отнесенных к классам объектов.

Второй тип обучения – обучение без учителя. Формально постановку задачи обучения без учителя можно описать следующим образом. Пусть X – множество данных – описаний некоторых

объектов. Необходимо найти множество Y , состоящее из взаимосвязей $f: (x, x')$ между объектами из X ($x, x' \in X, f \in Y$). Качество выявления взаимосвязей проверяется некоторой метрикой, выбранной исходя из решаемой задачи.

Обучение без учителя используется для решения следующих типов задач:

1. Задача кластеризации.
2. Поиск правил ассоциации.
3. Сокращение размерности данных.
4. Визуализация данных.

В определенной степени каждая из последних трех задач является производной от первой или ее частным случаем. Рассмотрим подробнее формулировки названных задач.

Под задачей поиска правил ассоциаций подразумевается выявление в признаковых описаниях объектов (исходных данных) таких наборов и значений признаков, которые особенно часто (неслучайно часто) встречаются в исходных данных. Если же проводить аналогию с первой задачей, то каждое правило в данном случае может быть представлено как кластер.

Задача сокращения размерности данных состоит в следующем. Существует большой (значительно большой) объем признаковых описаний объектов. Причем этот объем обуславливается внушительным количеством измерений признакового пространства. Необходимо представить те же данные в пространстве меньшей размерности, при этом минимизировав потери информации. Группировка по кластерам как раз и будет одним из вариантов решения проблемы.

Задача визуализации данных является по сути частным случаем предыдущей: ее цель – представить исходные данные в отображаемом пространстве, то есть пространстве размерности 2 или 3.

Как следует из описанного выше, обучение без учителя в какой-то мере так или иначе сводится к кластеризации. Поэтому для оценки качества обучения данным способом, как правило, используют метрики качества кластеризации. Причем при их выборе учитывается,

что эти метрики не должны зависеть от исходных данных, а только от результатов разбиения. Все оценки качества можно разделить на внешние и внутренние. Первые используют внешнюю информацию об истинном разбиении объектов на кластеры, вторые опираются только на набор исходных данных, то есть данные метрики могут работать с неразмеченной выборкой, когда заранее не известно истинное разбиение объектов на группы. И именно с их помощью определяют оптимальное число кластеров.

Приведем в качестве примера метрики, выделенные компанией ODS [2]:

1. Adjusted RandIndex (ARI). Данная метрика относится к группе внешних: предполагается, что истинные метки объектов известны (например, заданы экспертом), однако от самих значений меток она не зависит. Только от разбиения выборки на кластеры.

Рассчитывается мера следующим образом: пусть n – число объектов в выборке, a – число пар объектов, имеющих одинаковые метки и находящихся в одном кластере, b – число пар объектов, имеющих различные метки и находящихся в различных кластерах. Тогда можно посчитать долю объектов, для которых исходное и полученное в результате кластеризации разбиения согласованы:

$$RI = \frac{2(a+b)}{n(n-1)} .$$

Полученная величина называется RandIndex (RI) и выражает схожесть двух разных кластеризаций одной и той же выборки. Чтобы этот индекс давал значения, близкие к нулю, для случайных кластеризаций при любом n и числе кластеров, необходимо произвести его нормирование, то есть получить AdjustedRandIndex:

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]} ,$$

где E – математическое ожидание.

Мера ARI симметрична и не зависит от перестановок и значений меток. По сути этот индекс является мерой расстояний между различными разбиениями выборки. Его область значений – $[-1,1]$. Интерпретировать ее интервалы можно следующим образом: для «независимых» разбиений на кластеры – отрицательные значения, для случайных разбиений – близкие к нулю, для схожих разбиений – положительные значения, причем, $ARI=1$ говорит о совпадении разбиений.

2. Adjusted MutualInformation (AMI). Данная метрика схожа с предыдущей: она также симметрична и не зависит от значений и перестановок меток. Для ее определения используется функция энтропии. Разбиения выборки интерпретируются как дискретные вероятности: вероятность отнесения к кластеру равна доле объектов в нем. Как и в предыдущем случае, для данной метрики необходимо вычислить специальный индекс – MutualInformation (MI). Данная величина определяется как взаимная информация для двух распределений, соответствующих разбиениям выборки на кластеры. Интерпретировать это можно как долю информации, общей для обоих разбиений: насколько информация об одном из них уменьшает неопределенность относительно другого. Этот индекс рассчитывается по следующей формуле:

$$MI(XY) = \sum_{y \in Y} \sum_{x \in X} p(xy) \log \frac{p(xy)}{p(x)p(y)},$$

где $p(x, y)$ – совместная функция распределения вероятностей для X и Y ; $p(x)$ и $p(y)$ – функции распределения предельной вероятности для X и Y соответственно.

Областью значения индекса AMI является диапазон $[0,1]$. Интерпретируется он следующим образом: значения, близкие к нулю, говорят о независимости разбиений, а близкие к единице – об их схожести (или совпадении при $AMI=1$).

3. Гомогенность, полнота, V-мера. Данные метрики рассматривают разбиение выборки как дискретные распределения и определяются с использованием функции энтропии и условной энтропии.

Гомогенность определяет, насколько каждый кластер состоит из объектов одного класса, и рассчитывается по формуле

$$h = 1 - \frac{H(C|K)}{H(C)},$$

где K – результат кластеризации, C – истинное разбиение, H – функция энтропии. При этом

$$H(X) = -\sum_{i=1}^n P(x_i) \log_b P(x_i),$$

$$H(X|Y) = \sum_j p(y_j) \log \frac{p(y_j)}{p(x_i|y_j)}.$$

Полнота измеряет, насколько объекты одного класса относятся к одному кластеру и определяется по формуле

$$c = 1 - \frac{H(K|C)}{H(K)}.$$

Эти меры принимают значения в диапазоне $[0,1]$. Интерпретировать их можно следующим образом: чем больше значение, тем более точна кластеризация. Эти меры не являются симметричными и не являются нормализованными, в отличие от рассмотренных ранее, и поэтому они зависят от числа кластеров. Согласно ресурсу [2]: «Случайная кластеризация не будет давать нулевые показатели при большом числе классов и малом числе объектов. В этих случаях предпочтительнее использовать *ARI*. Однако при числе объектов более 1000 и числе кластеров менее 10 данная проблема не так явно выражена и может быть проигнорирована».

Чтобы учесть значения обеих величин, вводится симметричная V -мера, показывающая, насколько кластеризации схожи между собой. Ее расчет происходит по формуле

$$v = 2 \frac{hc}{h+c}.$$

Силуэт. Данная метрика относится к типу внутренних: она не предполагает знания об истинном разбиении и первоначально рассчитывается для каждого объекта следующим образом. Пусть a – среднее расстояние от данного объекта до объектов из того же кластера,

b – среднее расстояние от данного объекта до объектов из ближайшего кластера (но не того, в котором находится сам объект). Тогда силуэтом данного объекта будет величина, вычисляемая по формуле

$$s = \frac{b - a}{\max(ab)}$$

Силуэтом же всей выборки будет средняя величина силуэтов ее объектов. Интерпретировать метрику можно следующим образом: она показывает, насколько среднее расстояние до объектов своего кластера отличается от среднего расстояния до объектов других кластеров. Область значения данной величины – диапазон $[-1,1]$. При этом значения, близкие к левой границе, соответствуют плохим (разрозненным) кластеризациям; значения, близкие к середине, говорят о пересечении и наложении кластеров; значения, близкие к правой границе, соответствуют «плотным», четко выделенным кластерам. Согласно [2]: «чем больше силуэт, тем более четко выделены кластеры, и они представляют собой компактные, плотно сгруппированные облака точек».

Данная величина может быть использована для выбора оптимального числа кластеров, если оно заранее неизвестно: оно будет равно числу, максимизирующему значение силуэта. В отличие от рассмотренных ранее метрик, данная величина зависит от формы кластеров. Силуэт достигает больших значений при более выпуклых кластерах, которые получаются при помощи алгоритмов, основанных на восстановлении плотности распределения.

Типовые задачи при подготовке данных и обучении моделей

В разделе «Общий алгоритм машинного обучения» были представлены шаги разработки решения. Если задача типовая и не требует специального исследования или разработки специального метода, то существенная работа начинается с п. 3, который посвящен подготовке данных (обучающей и тестовой выборке). Исходные данные в подавляющем большинстве случаев требуют предварительной обработки перед тем как на них будет обучаться модель. Как правило, предварительная обработка подразумевает следующую работу: учет

пропусков, кодирование нечисловых данных, приведение данных к единому масштабу и стандартизация данных, разметка данных. Рассмотрим их подробнее.

Учет пропусков

В реальных задачах данные могут содержать пропуски. Это может быть вызвано тем, что клиенты не заполняли все поля в анкете или аккаунте, или тем, что не все параметры были оцифрованы за все время работы системы.

Таким образом, встает вопрос о том, как интерпретировать пропуски. Простейший вариант заключается в исключении объектов, имеющих неполные сведения (то есть удаление тех строк из матрицы признаков, в столбцах которых есть пропуски), или исключении признаков, содержащих неполные сведения (то есть удаления тех столбцов, в которых есть пропуски). Плюсы этого варианта в том, что он очень прост и его можно сразу опробовать на какой-либо модели. Минусы вполне очевидны. Если много объектов будут иметь пропуски, то мы рискуем удалить важные аспекты закономерности, которая сокрыта в данных, и, как следствие, получим низкое качество аппроксимации. Если мы удалим столбец, который содержит крайне важный признак, мы рискуем вовсе потерять ключевую информацию о разделении объектов.

Второй вариант борьбы с пропусками заключается в том, чтобы заменить их при помощи интерполяции. Это может быть среднее или медианное значение по столбцу. В случае если признак является функцией от времени, как и отдельные объекты, то можно интерполировать пропуски только по соседним временным значениям (например, если объект представляет собой вероятность покупки квартиры, а признак – среднюю заработную плату программиста в этот год, то за пропущенный год цифру можно приблизить по двум соседним).

Третий вариант можно также рассмотреть на примере с квартирой, который был описан выше. Если речь идет о каких-то открытых

экономических, социальных или других параметрах, то их можно найти в других источниках и тем самым дополнить данные.

Четвертый вариант заключается в том, чтобы закодировать пропуски специальным числовым значением (это может быть число, не встречающееся у других объектов) или категориальным значением (представление категориальных признаков будет рассмотрено ниже). Такой подход, как минимум, позволит внести информацию о том, что эти объекты по этому признаку отличаются от других (то есть мы не сообщим точной информации, но не удалим объекты полностью, как в первом подходе).

Пятый подход заключается в привлечении эксперта в соответствующей теме, который сможет сказать, как именно лучше всего интерполировать данные (на основе специфичных математических моделей, которые, вероятно, существуют или могут подойти). Например, для каких-то признаков можно сгенерировать псевдослучайные значения, подчиняющиеся некому распределению с учетом других известных признаков. Сюда же можно отнести генерацию синтетических данных (то есть искусственных) на основании собственного понимания предметной области. Долгий и тщательный анализ данных может прояснить поведение скрытого параметра, а также выявить способы, как сэмулировать его с определенной достоверностью. Однако этот подход опасен тем, что неверное понимание данных приведет к тому, что мы заложим в данные другие закономерности и потом их же и найдем при помощи методов машинного обучения. Иными словами, вместо решения исходной задачи мы решим искусственно созданную.

В сложных случаях на практике применяется некоторая комбинация всех вышеперечисленных подходов. Какие-то объекты или столбцы полностью исключаются, потому что их слишком сложно интерполировать или сэмулировать. Как правило, это объекты или признаки с большим числом пропусков (например, где пропусков больше, чем реальных значений). Какие-то признаки помечаются

особым кодом, какие-то данные находятся извне, а какие-то эмулируются синтетическими данными.

Многие нюансы зависят от специфики задачи. Но если никакие идеи не работают, то, вероятно, данных просто недостаточно и нужно получать еще непосредственно исходные данные (проводить физические эксперименты, опрашивать клиентов, измерять какие-то статистические метрики поведения пользователей на веб-сайтах и тому подобное).

Замечание: кроме пропусков данные могут содержать ошибки и выбросы (аномальные отклонения), вызванные неверным заполнением данных или ошибкой измерения. Решение этих проблем выходит за рамки данного учебного пособия и рекомендуется к самостоятельному изучению.

Кодирование нечисловых признаков

Очевидно, что далеко не все признаки объектов естественно описываются численным значением. Если говорить о размерах объекта или стоимости какого-то товара, то эти признаки, несомненно, будут числовыми. Если же речь идет о цвете, типе товара (категории) или вообще о текстовом описании некоторого объекта, то подобные признаки, как правило, поступают неоцифрованными.

Нечисловые признаки с неупорядоченными значениями (в которых между значениями не определена дистанция, то есть нельзя сказать, что больше или меньше) называют категориальными, или номинальными. Типичный подход к их обработке – кодирование категориального признака с m возможными значениями с помощью m бинарных признаков. Каждый бинарный признак соответствует одному из возможных значений категориального признака и является индикатором того, что на данном объекте он принимает данное значение. Такой подход иногда называют one-hot-кодированием. Например, у нас есть три варианта материала изделия: дерево, пластик, сталь. Причем мы не знаем наверняка, что лучше, а что хуже и как это материал влияет на итоговое качество товара. Тогда вместо

того, чтобы закодировать набор материалов в один столбец как значения 1, 2, 3, one-hot-кодирование даст три столбца и следующие коды: 001, 010, 100. Заметьте, что это не двоичное кодирование.

Если речь идет не просто о категориях, а о целых предложениях или больших текстах на естественном языке, то задача существенно усложняется. Кодирование текста побуквенно в большинстве случаев ничего не даст, так как разнообразие слов и смыслов настолько велико, что на таком уровне абстракции модель не построит нужную функцию. Обработка и понимание текста – это во многом направление для исследований.

Тем не менее на сегодняшний день существует ряд подходов, которые позволяют решать реальные задачи. Первый вариант – отнестись к тексту как к категориям. Допустим, в текстовых данных всего встречается 10 тыс. уникальных слов (не считая союзов, предлогов и тому подобное). Тогда каждое отдельное текстовое описание (присущее конкретному объекту) есть такая категория, где встречаются соответствующие слова, то есть каждый текст будет кодироваться в вектор из 10 тыс. элементов, где на месте соответствующих слов-элементов будут стоять единицы, а на месте слов, которых нет в данном тексте – нули. В итоге получим довольно разреженную матрицу (состоящую в основном из нулей) и при этом довольно большую. Из-за этого возникает проблема ее хранения и обработки. На сегодняшний день существует ряд техник оптимизации работы с разреженными матрицами как на фундаментальном уровне (например, сингулярное разложение), так и на уровне библиотек (например, в пакете `scipy`).

Однако учитывать все слова часто бывает избыточно и даже бессмысленно. Поэтому на практике применяется подсчет статистических характеристик текста, таких как TF-IDF. Подробнее о статистическом анализе текстов можно прочитать в [3].

Сегодня также актуально кодирование текста методом `word2vec` на основе машинного обучения. Ключевой особенностью метода является то, что большой массив слов отображается в вещественные

векторы небольшой размерности (100-200 элементов), причем, похожие слова имеют близкие друг к другу векторы (то есть вводится дистанция между словами).

Приведение данных к единому масштабу и стандартизация

«Сырые» данные имеют разный масштаб и разное распределение по каждому признаку. Например, какой-то химический показатель смеси может иметь значения в диапазоне от 0.0001 до 0.2, а другой показатель от -100 до 100. Или, скажем, возраст клиентов может быть от 16 до 40, причем гораздо больше клиентов имеет возраст от 18 до 25, иными словами, математическое ожидание смещено относительно центра распределения. Подобные различия в признаках могут вносить существенную ошибку для множества моделей (например, для регрессии, нейронных сетей), и потому требуется привести все признаки к единому виду.

Существует некоторая путаница в терминах «стандартизация» и «нормализация». Очень часто под стандартизацией и нормализацией понимаются разные вещи, а иногда стандартизацию рассматривают как часть нормализации. Поэтому важно понять общую суть и цель этих методов.

Стандартизация данных – это процесс приведения вектора каждого признака к такому виду, что его математическое ожидание станет нулевым, а дисперсия – единичной.

Нормализация данных – это процесс масштабирования вектора каждого признака, то есть приведение его к такому виду, что вектор будет иметь единичную норму (при этом есть разные способы оценки\подсчета нормы).

Так, стандартизация матрицы X:

[[1., -1., 2.],

[2., 0., 0.],

[0., 1., -1.]],

даст следующий результат:

$$\begin{bmatrix} 0. & -1.22 & 1.34 \\ 1.22 & 0. & -0.27 \\ -1.22 & 1.22 & -1.07 \end{bmatrix}$$

Видно, что значения вектора сместились (выровнялись относительно единого центра в нуле), а также произошло выравнивание разброса. Уже такого преобразования достаточно, чтобы повысить качество данных. Однако видно, что значения разных векторов не будут в одинаковом диапазоне (от -1 до 1 , например). Они будут лишь иметь стандартный разброс в рамках вектора\столбца\признака.

Для того чтобы достичь одинакового масштаба всех векторов, необходима нормализация. Есть несколько видов норм и, соответственно, нормализации.

Самый очевидный и простой метод – это max норма. Чтобы все значения лежали в одном диапазоне, нужно найти максимальное из возможных значений и все остальные поделить на него. Таким образом, максимальное значение будет единицей, а все остальные лягут в диапазон от 0 до 1 . Но это при условии, что нет отрицательных значений.

Менее очевидный способ – это L1 норма и нормализация. Формула следующая:

$$x'_i = \frac{x_i}{\|x\|_1} = \frac{x_i}{\sum_j |x_j|},$$

где $\|x\|_1$ и есть L1 норма, а вся формула целиком отображает процесс нормализации вектора x .

Еще один способ – это L2 норма. Формула нормализации вектора x :

$$x'_i = \frac{x_i}{\|x\|_2} = \frac{x_i}{\left(\sum_j x_j^2\right)^{1/2}},$$

где $\|x\|_2$ и есть L2 норма, а вся формула целиком отображает процесс нормализации вектора x .

Исследователи говорят, что лучшие результаты дают L2 и max нормализация, хотя любой вариант лучше, чем использование исходных данных.

По поводу плюсов-минусов этих способов можно сказать следующее: max нормализация не дает запас на неизвестные новые значения (то есть если заранее неизвестен весь диапазон данных, лучше не использовать эту норму), а L2 норма вычислительно дольше, но чаще всего дает оптимальный результат. L1 норма может дать слишком большой запас при большом разбросе данных, что может удалить нужную информацию из данных.

Разметка данных

Разметка данных – последняя из рассматриваемых нами (по порядку, но не по важности) операция с данными. Если речь идет о том, чтобы обучить модель прогнозировать будущие показатели по прошлому опыту (например, прогноз средней выручки по количеству посетителей магазинов), то такие данные, как правило, приходят с известной целевой переменной Y (то есть средней выручкой). Эти данные необходимо будет обработать в соответствии с пунктами выше, но их не потребуется дополнительно размечать. Однако не редкость, когда специалисту по анализу данных необходимо будет самостоятельно размечать выборку. Эта потребность может возникнуть из-за отсутствия размеченных исходных данных, так и может быть обусловлена экспериментами, возникающими в ходе решения общей задачи. Например, для распознавания визуальных образов машин или светофоров потребуется создать специальный файл, в котором будут проставлены ассоциации образов по имени файлов (то есть 001.png – «светофор»). В реальном бизнесе эту задачу может решать и не специалист по анализу данных или машинному обучению, а другие сотрудники (например, со стороны заказчика), но важно отметить, что этот этап может также возникнуть в ходе разработки решения, и его нельзя избежать в случае обучения с учителем.

Однако стоит отметить, что в случаях обучения и без учителя может потребоваться частичная разметка данных с целью тестирования и оценки качества модели, так как в противном случае мы просто получаем «черный ящик» без понимания того, как мы решили поставленную задачу.

Переобучение

Нами были рассмотрены типовые задачи по подготовке данных, теперь рассмотрим одну из основных проблем алгоритмов машинного обучения – переобучение.

Мы помним, что одной из важных характеристик алгоритмов машинного обучения является обобщающая способность. Однако с ней связаны еще два понятия: недообучения и переобучения. При подготовке данного пункта использовался материал (в том числе иллюстративный) из источника [4].

Недообучение возникает при обучении по прецедентам и характеризуется тем, что алгоритм не дает удовлетворительно малой средней ошибки на обучающем множестве. Как правило, это явление появляется вследствие использования недостаточно сложных моделей.

Противоположное этому явлению – переобучение. Его суть состоит в том, что вероятность ошибки натренированного алгоритма на объектах тренировочной выборки оказывается существенно меньше, чем на объектах тестовой. Чаще всего переобучение появляется из-за использования слишком сложных моделей.

Рассмотрим график, иллюстрирующий эффект переобучения (рисунок 5).

Точки на графике с рисунка 5 соответствуют разным способам обучения, и каждая из них получена усреднением по большому числу разбиений исходной выборки объемом в 72 образца на тестовую и обучающую части. Как видно из рисунка, точки имеют постоянное смещение вверх относительно диагонали графика. То есть наблюдается эффект переобучения: ошибки на тестовой выборке появляются чаще, чем на обучающей.

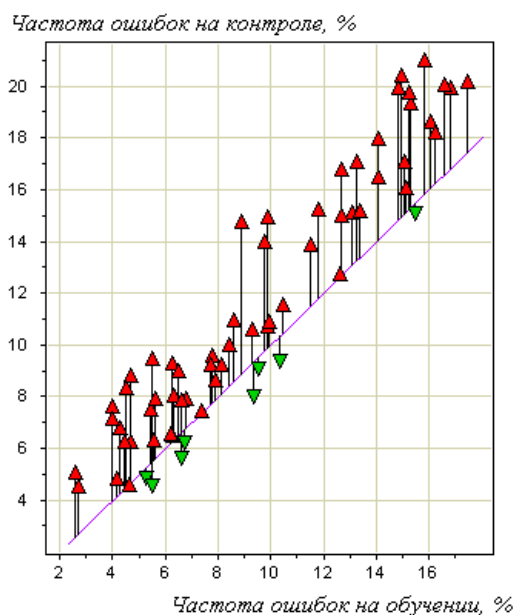


Рисунок 5. Иллюстрация переобучения

Чаще всего при построении алгоритмов обучения используется метод минимизации эмпирического риска (средней ошибки алгоритма на обучающей выборке). Его суть состоит в том, чтобы для текущей модели подобрать алгоритм, минимизирующий значение средней ошибки на данной обучающей выборке.

С переобучением метода минимизации эмпирического риска связаны три утверждения, объясняющие его причину:

1. Минимизация эмпирического риска не является гарантией малой вероятности ошибки на тестовых данных. Можно легко построить алгоритм, который минимизирует эмпирический риск до нуля, однако не будет способен к обучению. Суть состоит в том, что этот алгоритм запоминает обучающую выборку, потом сравнивает запомненный образец с предъявляемым. При совпадении предъявленного объекта и образца обучающей выборки алгоритм выдаст правильный ответ, иначе – выведется произвольный. То есть эмпирический риск равен нулю, однако обобщающей способности у алгоритма нет.

2. Согласно [4]: «Переобучение появляется именно вследствие минимизации эмпирического риска. Пусть задано конечное множество из D алгоритмов, которые допускают ошибки независимо и с одинаковой вероятностью. Число ошибок любого из этих

алгоритмов на заданной обучающей выборке подчиняется одному и тому же биномиальному распределению. Минимум эмпирического риска – это случайная величина, равная минимуму из D независимых одинаково распределенных биномиальных случайных величин, ожидаемое значение которой уменьшается с ростом D . Соответственно, с ростом D увеличивается переобученность – разность вероятности ошибки и частоты ошибок на обучении.

В данном модельном примере легко построить доверительный интервал переобученности, так как функция распределения минимума известна. Однако в реальной ситуации алгоритмы имеют различные вероятности ошибок, не являются независимыми, а множество алгоритмов, из которого выбирается лучший, может быть бесконечным. По этим причинам вывод количественных оценок переобученности является сложной задачей, которой занимается теория вычислительного обучения. До сих пор остается открытой проблема сильной завышенности верхних оценок вероятности переобучения».

3. Переобучение появляется в связи с избыточной сложностью модели. Всегда можно найти оптимальное значение сложности модели, при котором переобучение будет минимальным. Для примера приведем несколько графиков, на которых будет видна зависимость переобучения от сложности модели. При степени 2 полинома (рисунок 6) модель является недообученной. При степени 40 (рисунок 7) – переобученной и неустойчивой, а вот степень 20 (рисунок 8) – оптимальна.

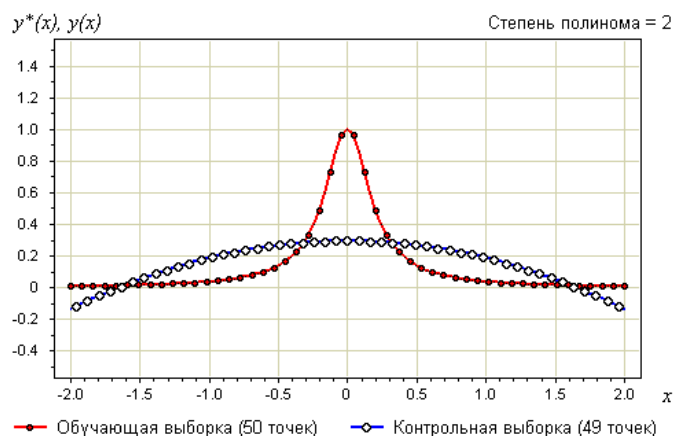


Рисунок 6. Недообученная модель

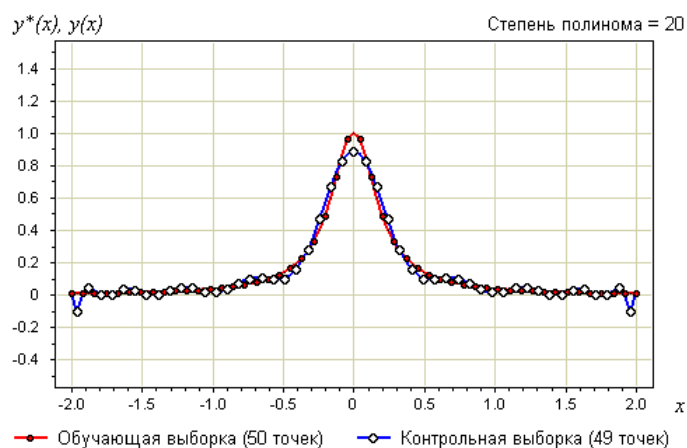


Рисунок 7. Оптимальная модель

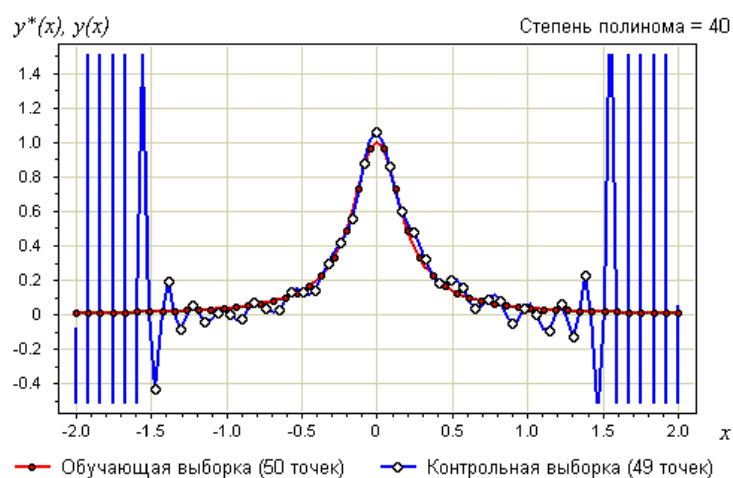


Рисунок 8. Переобученная модель

Одним из способов измерения вероятности переобучения является эмпирический метод Монте-Карло, или метод скользящего контроля. В литературе также встречаются названия кросс-проверка, или кросс-валидация. Суть его в следующем: производится некоторое количество разбиений исходной выборки на обучающую и контрольную. Для каждого разбиения происходит обучение алгоритма на обучающей подвыборке и оценка средней ошибки на контрольной. Затем вычисляется оценка скользящего среднего, как средняя по всем разбиениям величина вычисленной ошибки. Для независимой выборки этот показатель дает несмещенную оценку вероятности ошибки. Метод скользящего контроля является стандартным способом сравнения и оценивания алгоритмов классификации, регрессии, прогнозирования.

При использовании кросс-валидации можно сделать следующие выводы:

1. Если ошибка большая на большинстве участков, то скорее всего проблема в модели.

2. Если данные обучающей выборки характеризуются сильно смещенными математическим ожиданием и дисперсией, то уровень обобщения будет низким, что может быть связано с переобучением.

3. Если найдены сильные отклонения на определенных подвыборках, то, вероятно, проблема в этих участках данных или модель недообучается.

4. При малом объеме обучающей выборки кросс-валидация может стать способом борьбы с переобучением.

Если же говорить о прямых способах борьбы с переобучением, то можно выделить следующие:

1. Упрощение модели.

2. Подготовка большего числа обучающих данных (возможно, с помощью генерации).

3. Регуляризация.

Остановимся подробнее на последнем. Регуляризация представляет собой добавление некоторой дополнительной информации к условию минимизации ошибки. Выполняется это, чтобы решить некорректно поставленную задачу или предотвратить переобучение. Чаще всего добавляемая информация принимает вид штрафа за сложность модели. Например, введенные ограничения по норме векторного пространства или гладкости результирующей функции. Или же, с байесовской точки зрения, добавленные априорные распределения на параметры модели.

Согласно [5]: «Существуют следующие основные виды регуляризации:

L1-регуляризация (англ. Lasso regression):

$$RSS(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \lambda \sum_{j=0}^M |w_j| ,$$

где w – вектор весов полинома;
 λ – коэффициент регуляризации.

L2, или Регуляризация Тихонова (в английской литературе – ridge regression или Tikhonov regularization), для интегральных уравнений позволяет балансировать между соответствием данным и маленькой нормой решения:

$$\text{RSS}(\mathbf{w}) = \frac{1}{2} \sum_{i=1}^N (f(x_i, \mathbf{w}) - y_i)^2 + \frac{\lambda}{2} \|\mathbf{w}\|^2,$$

где $\|\mathbf{w}\|^2$ – квадратичная норма вектора весов (сумма квадратов каждого веса).

Иными словами, переобучение в большинстве случаев проявляется в том, что в получающихся многочленах слишком большие коэффициенты. Соответственно и бороться с этим можно довольно естественным способом: нужно просто добавить в целевую функцию штраф, который бы наказывал модель за слишком большие коэффициенты».

По поводу применения той или иной регуляризации приведем материал источника [6]: «Регуляризацию можно применять с любым методом МО-классификации, который основан на математическом уравнении. Примеры включают линейную, логистическую регрессию и нейронные сети. Поскольку это уменьшает величину весовых значений в модели, регуляризацию иногда называют сокращением весов. Основное преимущество применения регуляризации в том, что оно часто приводит к созданию более точной модели. Главный недостаток заключается во введении дополнительного параметра, значение которого нужно определить, – весового значения регуляризации. В случае логистической регрессии это не слишком серьезно, так как в этом алгоритм обычно используется лишь параметр скорости обучения, но при использовании более сложного метода классификации, в частности нейронных сетей, добавление еще одного так называемого гиперпараметра может потребовать массы дополнительной работы для подбора комбинации значений двух параметров.

L1- и L2-регуляризация – процессы схожие. Какой же из них лучше? В принципе, исследователями сформулированы кое-какие теоретические правила насчет того, какая регуляризация лучше для определенных задач, но на практике придется поэкспериментировать, чтобы найти, какой тип регуляризации лучше в вашем случае и стоит ли вообще использовать какую-либо регуляризацию.

Применение L1-регуляризации иногда может давать полезный побочный эффект, вызывающий стремление одного или более весовых значений к 0.0, а это означает, что соответствующий признак не оказывает значимого влияния на результирующий, то есть включение его в модель требуется. Это одна из форм того, что называют «селекцией признаков». В отличие от L1, L2-регуляризация ограничивает весовые значения модели, но обычно не приводит к полному обнулению этих значений. Поэтому может показаться, что L1-регуляризация лучше L2-регуляризации. Однако недостаток применения L1-регуляризации в том, что этот метод не так-то просто использовать с некоторыми алгоритмами машинного обучения. Например, с теми, в которых используются численные методы для вычисления так называемого градиента. L2-регуляризацию можно использовать с любым типом алгоритма обучения.

Таким образом, можно сделать вывод, что L1-регуляризация иногда дает полезный побочный эффект удаления ненужных признаков, присваивая связанным с ними весам значение 0.0, но L1-регуляризация стабильно работает не со всеми формами обучения. L2-регуляризация работает со всеми формами обучения, но не обеспечивает неявной селекции функций. На практике же следует использовать метод проб и ошибок, чтобы определить, какая форма регуляризации (если она вообще нужна) лучше для конкретной задачи».

Рассмотренные методы борьбы с переобучением подойдут для регрессии или нейронных сетей. Для борьбы же с переобучением в моделях Деревьев Решений используют `pruning` (так называемое усечение) – удаление наименее информативных узлов для упрощения модели. Дело в том, что сразу нельзя построить оптимальное

(маленькое) дерево, так как в дереве малого размера будет мало информативных параметров, чтобы корректно обработать все входные образы (модель будет неполной, неинформативной). Поэтому сперва делают большое, но информативное дерево, а затем удаляют наименее информативные узлы и подтягивают дерево, уменьшая его размер\глубину.

МОДЕЛИ И АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

Как было сказано во введении, машинное обучение находится на стыке математической статистики, численных методов оптимизации, теории вероятностей и дискретного анализа. Эта дисциплина комбинирует и использует различные методы в построении математических моделей объектов и явлений для последующего их изучения и добычи знаний. В данном разделе предлагается рассмотреть, основные элементы математики, использующиеся в машинном обучении, а также изучить методы анализа данных, которые заложили основу рассматриваемой дисциплины.

Первое, что необходимо отметить, – любая дисциплина, использующая математический аппарат, так или иначе занимается математическим моделированием – заменой реального объекта его абстрактным, идеализированным представлением и использованием полученного представления для изучения объекта и добычи знаний. То есть основное, с чем работает машинное обучение, – это математические модели. Их можно разделить на различные типы по некоторым признакам. С точки зрения математического вида функции, составляющей модель, выделяют линейные и нелинейные модели. По количеству переменных в модели они делятся на сосредоточенные и распределенные. С точки зрения учета случайности модели делятся на детерминированные и стохастические, а с точки зрения изменчивости во времени – на статические и динамические. По природе используемых в модели параметров и переменных их можно разделить на дискретные и непрерывные.

В общем виде математическая модель может быть представлена следующим образом:

$$Y = F(X, W_c, W_d, W_s),$$

где Y – выходные данные; X – входные данные; F – некоторая функция; W_c – константные параметры модели; W_d – динамические параметры модели; W_s – статические параметры модели.

При построении моделей элементы множеств W , как правило, заранее неизвестны и требуют нахождения. Есть два способа решения этой задачи: аналитический и численный. Первый предполагает нахождение корней, второй – приближение результата к некоторому удовлетворительному значению. Удовлетворительность в данном случае оценивается некоторым заранее определенным критерием. При этом для решения задачи численным методом нередко происходит ее переформулировка для определения и введения условия, определяющего качество решения. Необходимо отметить, что численные методы чаще всего опираются на отклонение получаемого результата от желаемого и стремятся минимизировать это отклонение. Причем данный подход будет работать, даже если неизвестна точная формула для получения выходного значения, однако обязательно существует способ измерения отклонения, а также можно найти значения, которые он должен принимать.

Появление численных методов оптимизации было обусловлено, во-первых, тем, что не все задачи можно решить аналитически. Классический пример подобной задачи – гравитационная задача N тел. Согласно Википедии [1], ее формулировка звучит так: «В пустоте находится N материальных точек, массы которых известны $\{m_i\}$. Пусть попарное взаимодействие точек подчинено закону тяготения Ньютона, и пусть силы гравитации аддитивны. Пусть известны начальные на момент времени $t=0$ положения и скорости каждой точки $\mathbf{r}_i|_{t=0} = \mathbf{r}_{i0}$, $\mathbf{v}_i|_{t=0} = \mathbf{v}_{i0}$. Требуется найти положения точек для всех последующих моментов времени». И согласно тому же источнику [1]: «На данный момент в общем виде задача N тел для $N>3$ может быть решена только численно, причем для $N=3$ ряды Зундмана даже при современном уровне компьютеров использовать практически невозможно».

Во-вторых, даже возможные аналитические решения систем могут обладать существенной сложностью. Матричные решения имеют вычислительную сложность $O(N^3)$, тогда как численные – линейную.

В-третьих, при противоречивости или мультиколлинеарности данных аналитические решения могут давать неопределенный результат. В отличие от них численные методы всегда сходятся пусть и к локальному, но все же определенному решению.

В основном, при решении задач минимизации ошибки, в машинном обучении используются градиентные методы, составляющие особый класс алгоритмов оптимизации. Градиент в данном случае – это (согласно Википедии [1]): «вектор, своим направлением указывающий направление наибольшего возрастания некоторой величины φ , значение которой меняется от одной точки пространства к другой (скалярного поля), а по величине (модулю) равный скорости роста этой величины в этом направлении».

Например, если взять в качестве φ высоту поверхности земли над уровнем моря, то ее градиент в каждой точке поверхности будет показывать «направление самого крутого подъема», и своей величиной характеризовать крутизну склона.

С математической точки зрения на градиент можно смотреть как на:

1. Коэффициент линейности изменения значения функции многих переменных от изменения значения аргумента.
2. Вектор в пространстве области определения скалярной функции многих переменных, составленный из частных производных.
3. Содержимое Матрицы Якоби. Ее строки содержат градиенты составных скалярных функций, из которых состоит векторная функция многих переменных.

Пространство, на котором определена функция и ее градиент, может быть как обычным трехмерным пространством, так и пространством любой другой размерности, любой физической природы или чисто абстрактным (безразмерным)».

Метод градиентного спуска – нахождение локального экстремума (минимума или максимума) функции путем движения вдоль градиента в направлении наискорейшего спуска, задаваемого антиградиентом. Существуют следующие типы градиентного спуска:

1. С постоянным шагом.
2. С дроблением шага.
3. Наискорейшего спуска.
4. Стохастический.

Более подробно о градиентном спуске мы поговорим далее, здесь же рассмотрим аналитические методы, используемые в машинном обучении, так как именно они составляют первое поколение алгоритмов, применяемых для анализа данных.

Методы теории вероятностей

Первый срез методов, который необходимо рассмотреть, – методы теории вероятностей. Теория вероятностей – раздел математики, в котором изучаются случайные величины и события, их свойства и возможные операции над ними. Таким образом, ключевое понятие, лежащее в основе этой дисциплины, – вероятность события P_i . Опираясь на него, можно дать определение полной группы событий. Она определяется как система случайных событий, которая обладает следующими свойствами:

1. В результате случайного эксперимента непременно произойдет одно и только одно из составляющих ее событий.
2. Сумма вероятностей всех событий полной группы равна 1.

Достаточно часто при исследовании данных (выборок) и подсчете определенных характеристик выборки считаются полными группами событий.

С помощью методов теории вероятностей можно проводить как простые, так и сложные операции по анализу данных. Среди простых можно выделить подсчеты вероятностных характеристик выборки: медианы, математического ожидания, дисперсии, а также величины среднеквадратического отклонения. Рассмотрим их подробнее.

Медиана – число, характеризующее выборку по среднему из ее значений. То есть, если все данные выборки различны, и она упорядочена по возрастанию, то ровно половина из элементов выборки

будет меньше медианы, и ровно половина – больше. Формально величину можно выразить следующим образом:

Если $X = \{x_1, \dots, x_n\}$ – характеризуемая выборка, то x_k – медиана, если $x_j < x_k$, при $j \in [1, k)$ и $x_k < x_i$, при $i \in (k, n]$ и $k = n/2$.

Рассмотрим пример. Пусть выборка $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, тогда ее медианой будет число 5.

Математическое ожидание – среднее значение вероятностных элементов выборки. Формально она рассчитывается так:

$$M|X| = \sum x_i p_i .$$

Рассмотрим пример. Пусть выборка $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$, тогда ее математическим ожиданием будет величина, равная 5. Вычисляется она следующим образом: выборка рассматривается как полная группа событий с равными вероятностями. То есть p_i для каждого элемента равно 0,11. Тогда $1 \times 0,11 + 2 \times 0,11 + 3 \times 0,11 \dots = 5$.

Дисперсия – мера разброса элементов выборки относительно ее математического ожидания. Рассчитывается данная величина по формуле

$$D|X| = \sum (x_i - M|X|)^2 p_i .$$

При этом p_i рассчитывается как $1/(n - 1)$, где n – количество элементов выборки. Это выполняется для того, чтобы не учитывать отклонение самого математического ожидания от себя. То есть, для выборки $X = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ дисперсия будет рассчитываться следующим образом:

$$(1-5) \times (1-5) \times 0,125 + (2-5) \times (2-5) \times 0,125 + (3-5) \times (3-5) \times 0,125 \dots = 7,5.$$

Среднеквадратическое отклонение – величина, характеризующая рассеивание значений выборки относительно ее математического ожидания. Формула расчета

$$\sigma = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2},$$

где n – количество элементов в выборке; \bar{x} – математическое ожидание.

Интерпретировать данную величину можно следующим образом: чем больше значение среднеквадратического отклонения, тем

большой разброс значений в представленном множестве (относительно его средней величины). Малое значение среднеквадратического отклонения говорит о том, что элементы выборки сгруппированы вокруг ее среднего значения. Если говорить о более практическом применении величины, то в экономике, например, она характеризует доходность портфеля и его риск.

Однако рассмотренные показатели характеризуют случайную величину лишь с какой-то одной стороны. Наиболее же полно и исчерпывающе ее описывает закон распределения – функция, определяющая для выборки X вероятность попадания в некоторый интервал или вероятность получения определенного значения x_i . Если какой-либо закон распределения описывает случайную величину, то говорят, что она ему подчиняется или по нему распределена. Иными словами, закон распределения описывает область значений случайной величины и вероятности их получения.

Среди законов распределения наиболее часто используется закон нормального распределения или распределения Гаусса, который характеризует большинство процессов, встречающихся в мире. Отсюда и произошло его название (нормальное). Закон (плотность распределения вероятности) описывается следующей формулой:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

где σ – среднеквадратическое отклонение; μ – математическое ожидание.

С математической точки зрения можно заметить, что данная функция зависит от двух параметров: математического ожидания и среднеквадратического отклонения, поэтому по сути данный закон представляет собой семейство распределений. В качестве примера приведем изображение из источника [7], показывающее графики нормальных распределений с разными параметрами (рисунок 9).

Обратите внимание!

- График с параметрами $\sigma^2=1$ и $\mu=0$ соответствует стандартному нормальному распределению.

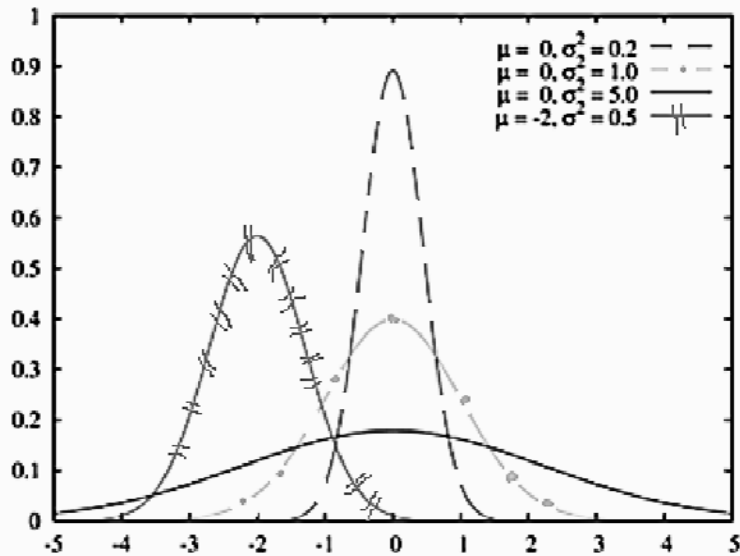


Рисунок 9. Плотность нормального распределения

Согласно Википедии [1]: «Важное значение нормального распределения во многих областях науки (например, в математической статистике и статистической физике) вытекает из центральной предельной теоремы теории вероятностей. Если результат наблюдения является суммой многих случайных слабо взаимосвязанных величин, каждая из которых вносит малый вклад относительно общей суммы, то при увеличении числа слагаемых распределение централизованного и нормированного результата стремится к нормальному. Этот закон теории вероятностей имеет следствием широкое распространение нормального распределения, что и стало одной из причин его наименования».

Кроме описанных выше величин и характеристик для анализа данных очень часто используется одна из основных теорем теории вероятностей – теорема Байеса.

Она позволяет определить вероятность наступления какого-либо события при условии, что произошло другое событие, статистически

взаимосвязанное с исследуемым. Ключевыми понятиями в данной теореме являются понятия априорной и апостериорной вероятностей. Рассмотрим их подробнее.

Априорная вероятность – назначенная событию вероятность, при условии отсутствия знаний, поддерживающих его наступление. Апостериорная вероятность – назначенная событию вероятность при условии наличия знаний, поддерживающих его наступление и полученных опытным путем.

Теперь перейдем непосредственно к теореме Байеса. Формулируется она следующим образом:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$

где $P(A|B)$ – апостериорная вероятность справедливости гипотезы A при наступлении B ; $P(B|A)$ – вероятность наступления события B при истинности гипотезы A ; $P(A)$ – априорная вероятность гипотезы A ; $P(B)$ – вероятность наступления события B .

Метод Байеса имеет свои достоинства и недостатки. К первому можно отнести возможность использования экспертных знаний, возможность точного описания явления и хорошую интерпретируемость результатов. К недостаткам же относится следующее:

1. Метод не ставит целью минимизацию ошибки классификации.
2. Метод требует работы эксперта.
3. Сильная зависимость результатов от выбора модели.
4. Плохая работа при малом количестве и высокой размерности данных.
5. Метод дает плохое обобщение, особенно на высокоуровневых признаках.
6. Метод дает плохие результаты при взаимозависимости признаков.

Таким образом, можно сделать вывод, что данный метод хорошо работает в тех случаях, где признаки и результат сильно завязаны на их частотных характеристиках.

Деревья решений

Следующий метод анализа данных, который нам необходимо рассмотреть, – применение деревьев решений. Дерево решений – средство поддержки принятия решений для прогнозных моделей. Суть его работы заключается в последовательном разбиении множества данных на непересекающиеся классы, которые в свою очередь также подвергаются разбиению по каким-либо критериям с оценкой эффективности разбиения. Как правило, дерево решений состоит из «узлов», «листьев» и «веток». «Ветки» содержат записи атрибутов, от которых зависит целевая функция, «листья» – значения целевой функции, а «узлы» – остальные атрибуты, по которым происходит классификация. Чаще всего выделяют два типа деревьев: для классификации (в этом случае предсказываемый результат – класс, которому принадлежат данные) и для регрессии (результат – прогнозируемое значение целевой функции).

Обобщенный алгоритм построения дерева решений по обучающей выборке состоит из следующих шагов:

1. Берем следующий атрибут и помещаем его в корень.
2. Для всех значений этого атрибута – оставляем в «листьях» данной «ветки» только те значения, которые соответствуют определенному условию.
3. Продолжаем строить дерево среди оставленных на предыдущем шаге «листьев».

Для выбора следующего атрибута может быть использован один из следующих основных алгоритмов:

1. ID3. Атрибут выбирается на основе прироста информации и минимизации энтропии. Напомним, что под энтропией в данном случае подразумевается мера неупорядоченности системы. Чем меньше ее величина, тем более упорядочены составляющие системы.
2. C4.5 – улучшенная версия предыдущего алгоритма, в которой используется нормализованный прирост информации.

3. CART – алгоритм, используемый для построения бинарных деревьев, производящий разбиение на основе модальных значений признаков.

Рассмотрим работу по построению дерева решений с использованием первого алгоритма. Более подробно об этом можно прочитать в [8, 9].

Кратко алгоритм ID3 может быть описан тремя шагами:

1. Посчитать энтропию разбиваемого множества.
2. Выбрать признак с минимальной энтропией и, соответственно, максимальной информационной выгодой.
3. На основе полученного признака создать узел дерева и повторить процедуру.

Рассмотрим построения дерева решений для классификации следующего множества из семи объектов: {красный круг, зеленый квадрат, красный квадрат, зеленый треугольник, зеленый круг, красный треугольник, красный прямоугольник}. Как мы видим, каждый из объектов характеризуется двумя признаками: цвет и форма, то есть именно по ним мы будем проводить разбиение. Для меры энтропии выберем формулу энтропии Шеннона:

$$H = - \sum_{i=1}^N p_i \times \log p_i$$

В таблице 2 представлены расчеты энтропии всей системы. Подсчеты энтропии для каждого из признаков приведены в таблицах 3 и 4, соответственно.

Таблица 2. Подсчет энтропии системы

Объект	pi	log(pi)	pi*log(pi)
красный квадрат	0,142857	-0,8451	-0,12073
красный прямоугольник	0,142857	-0,8451	-0,12073
красный круг	0,142857	-0,8451	-0,12073
зеленый квадрат	0,142857	-0,8451	-0,12073
зеленый треугольник	0,142857	-0,8451	-0,12073
зеленый круг	0,142857	-0,8451	-0,12073
красный треугольник	0,142857	-0,8451	-0,12073
Энтропия			0,845098

Таблица 3. Подсчет энтропии для признака «Цвет»

Объект	p_i	$\log(p_i)$	$p_i \cdot \log(p_i)$
красный	0,571429	-0,24304	-0,13888
зеленый	0,428571	-0,36798	-0,1577
Энтропия			0,296583

Таблица 4. Подсчет энтропии для признака «Форма»

Объект	p_i	$\log(p_i)$	$p_i \cdot \log(p_i)$
круг	0,285714	-0,54407	-0,15545
квадрат	0,285714	-0,54407	-0,15545
треугольник	0,285714	-0,54407	-0,15545
прямоугольник	0,142857	-0,8451	-0,12073
Энтропия			0,587072

Как мы видим, первым признаком, на основе которого будет происходить разбиение, станет цвет, а вторым – форма. Классификация исходного множества с помощью построенного дерева показана на рисунке 10.

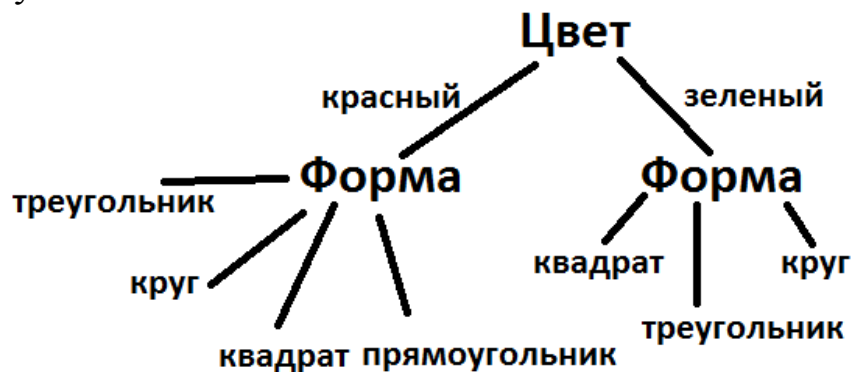


Рисунок 10. Классификация фигур с помощью дерева

Стоит отметить, что чаще всего деревья решений – бинарные. В узлах они содержат условия, а ветви соответствуют истинности или ложности этого условия. Поэтому, если дерево с рисунка 10 преобразовать в классическое дерево решений, то получится изображение, представленное на рисунке 11.

Теперь рассмотрим использование этого метода для решения задачи из источника [10]: спрогнозируем исход матча для футбольной команды исходя из следующих параметров:

- выше ли находится соперник по турнирной таблице;
- дома ли играется матч;
- пропускает ли матч кто-либо из лидеров команды;
- идет ли дождь.

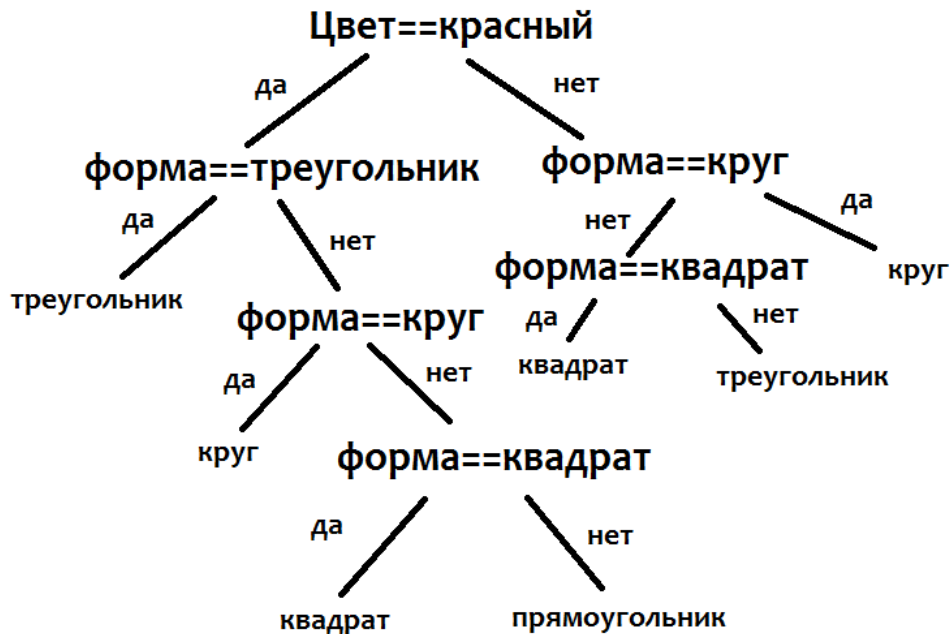


Рисунок 11. Классическое дерево решений

Прогноз предлагается выполнить на основе статистики, представленной в таблице 5.

Таблица 5. Статистические данные для задачи

Соперник	Играем	Лидеры	Дождь	Победа
Выше	Дома	На месте	Да	Нет
Выше	Дома	На месте	Нет	Да
Выше	Дома	Пропускают	Нет	Нет
Ниже	Дома	Пропускают	Нет	Да
Ниже	В гостях	Пропускают	Нет	Нет
Ниже	Дома	Пропускают	Да	Да
Выше	В гостях	На месте	Да	Нет
Ниже	В гостях	На месте	Нет	???

Первое, что мы рассчитываем, – энтропию для признаков. Результаты можно увидеть в таблице 6, причем данные приведены в округленном виде.

Таблица 6. Энтропия для признаков

Соперник	Играем	Лидеры	Дождь	Победа
0,29	0,26	0,29	0,29	0,29

Согласно таблице 6, первый признак, по которому будет происходить деление, – место игры. По нашим данным получатся два множества: со значением «дома» и со значением «в гостях». Если мы рассмотрим второе множество, то увидим, что следующий целевой признак с нулевой энтропией – исход матча: все элементы множества «в гостях» имеют значение поля победа «нет». Следовательно, по нашим статистическим данным с помощью деревьев решений мы прогнозируем поражение.

Если говорить о достоинствах деревьев решений, то можно выделить следующие. Во-первых, простота понимания и интерпретации. Во-вторых, минимальные требования к подготовке данных, а также способность работы с большими объемами данных. В-третьих, метод одинаково хорошо работает с разными видами признаков. В-четвертых, является надежным методом и позволяет оценить модель статистическими тестами. К недостаткам же можно отнести следующее:

1. Достаточно сложно построить оптимальное дерево решений.
2. Подверженность переобучению.
3. Не для всех задач может быть получено решение удовлетворительного качества.

Статистические модели и методы

Мы рассмотрели методы машинного обучения, предоставляемые теорией вероятностей, а также модель дерева решений. Теперь обратимся к статистике и кратко опишем ее модели и методы, которые могут быть использованы для решения задач машинного обучения, а именно методы регрессионного анализа, подробно рассмотренные в следующих источниках: [11,12,13].

В контексте машинного обучения под регрессионным анализом понимается процесс построения математической модели, описывающей зависимость некоторой целевой характеристики объекта или процесса от других его характеристик. Например, зависимость числа новых клиентов от величины зарплаты работающего на улице промоутера.

В задаче регрессионного анализа всегда есть обучающая выборка, состоящая из входных параметров и откликов, а также начальная параметрическая модель, в самом простом случае – линейная, однако не обязательно таковая. Для задачи из примера эта модель может иметь вид $y = \beta_0 + \beta_1 \times x$, где x – размер зарплаты промоутера; y – количество новых клиентов; β_0 и β_1 – параметры модели. Задача регрессии – оценить их, то есть найти такие значения β_0 и β_1 , чтобы полученная модель отражала зависимость между входом и выходом с требуемой точностью.

После получения адекватной модели мы можем решать задачу прогнозирования, подставляя в полученную формулу величину x и вычисляя величину y (с удовлетворяющей нас погрешностью). Приведенный выше пример модели – модель парной линейной регрессии, но помимо нее существует и множественная, и нелинейная регрессии. Начнем рассмотрение с самого простого.

Допустим, мы хотим описать зависимость между двумя факторами моделью вида $y = \beta_0 + \beta_1 \times x$. Первое, что необходимо учесть, – построенная линия никогда не будет точно проходить по опытным точкам, поэтому истинный вид регрессионной модели будет $y = \beta_0 + \beta_1 \times x + e$, где e – ошибки наблюдений. Второе – прежде чем переходить к оцениванию параметров модели, целесообразно построить диаграмму рассеяния, чтобы убедиться, что выбранная модель действительно может описать зависимость между факторами. На диаграмме рассеяния каждой паре «зависимый-влияющий параметр» соответствует точка на плоскости. Как правило, зависимый фактор откладывается по оси ординат, а второй – по оси абсцисс. Модель парной линейной регрессии графически представляет собой линию, следова-

тельно, использовать ее для описания зависимости целесообразно, если на диаграмме рассеяния точки располагаются вокруг (и достаточно близко) какой-либо прямой. Отклонение реальных точек от модельных – остатки или ошибки наблюдений, которые обозначаются e . Пример диаграммы рассеяния показан на рисунке 12.

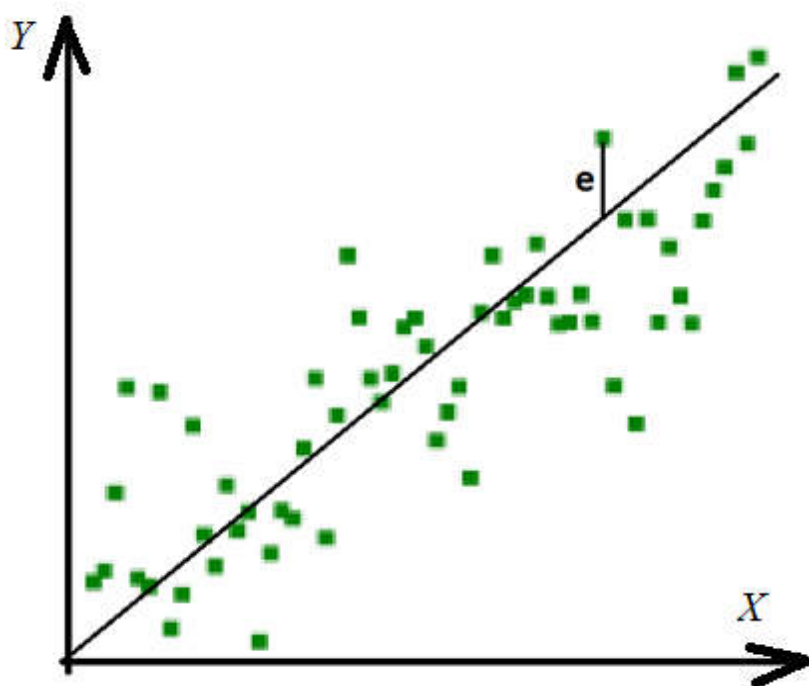


Рисунок 12. Диаграмма рассеяния

Если после анализа диаграммы рассеяния принимается решение использовать линейную парную регрессию для моделирования зависимости, то следующим шагом будет нахождение параметров модели β_0 и β_1 . Для решения этой задачи в девяносто девяти процентах случаев используется метод наименьших квадратов, предложенный Гауссом более двухсот лет назад. Суть данного метода заключается в минимизации суммы квадратов отклонений опытных данных от модельных. Формально эта задача описывается так:

$$Q = \sum e_i^2 \rightarrow \min$$

В данном случае e_i вычисляется следующим образом:

$$e_i = \beta_0 + \beta_1 * x_i - y_i,$$

где y_i – реальное выходное значение для входного значения x_i .

Чтобы решить указанную задачу оптимизации, решают систему уравнений:

$$\left. \begin{aligned} \frac{\partial Q}{\partial \beta_0} &= 0 \\ \frac{\partial Q}{\partial \beta_1} &= 0 \end{aligned} \right\}$$

Найдя необходимые частные производные и выполнив преобразования по упрощению выражений, получают нормальную систему уравнений для парной линейной регрессии:

$$\left. \begin{aligned} n * \beta_0 + \beta_1 * \sum x_i - \sum y_i &= 0 \\ \beta_0 * \sum x_i + \beta_1 * \sum x_i^2 - \sum x_i y_i &= 0 \end{aligned} \right\}$$

Рассмотрим для примера построение парной линейной регрессии для задачи зависимости ежемесячного количества новых клиентов от почасовой зарплаты промоутера по данным, представленным в таблице 7.

Таблица 7. Данные для анализа

Зарплата (x)	100	120	130	150
Количество клиентов (y)	70	100	120	140

Подставим данные в систему уравнений и получим следующий ее вид:

$$\left. \begin{aligned} 4 * \beta_0 + \beta_1 * 500 - 430 &= 0 \\ \beta_0 * 500 + \beta_1 * 63800 - 55600 &= 0 \end{aligned} \right\}$$

Выразим β_0 из первого уравнения:

$$\beta_0 = 107,5 - \beta_1 * 125$$

Подставим во второе и найдем β_1 :

$$\begin{aligned} 53750 - 62500 * \beta_1 + \beta_1 * 63800 - 55600 &= 0 \\ \beta_1 &= \frac{1850}{1300} \end{aligned}$$

В итоге получим уравнение регрессии:

$$y = 1,4231x - 70385$$

Полученные коэффициенты можно интерпретировать следующим образом. β_0 – значение при $x=0$. То есть, если у нас не будет работать промоутер (то есть мы не будем платить ему зарплату), ежемесячно мы прогнозируем отток 70 клиентов. Содержательная интерпретация второго коэффициента следующая: каждый рубль в зарплате промоутера дает 1,4 нового клиента в месяц.

Необходимо отметить, что коэффициенты для парной линейной регрессии можно найти средствами табличного процессора Excel. Для этого строится обычная точечная диаграмма, а затем отображается линия тренда и уравнение, как показано на рисунке 13.

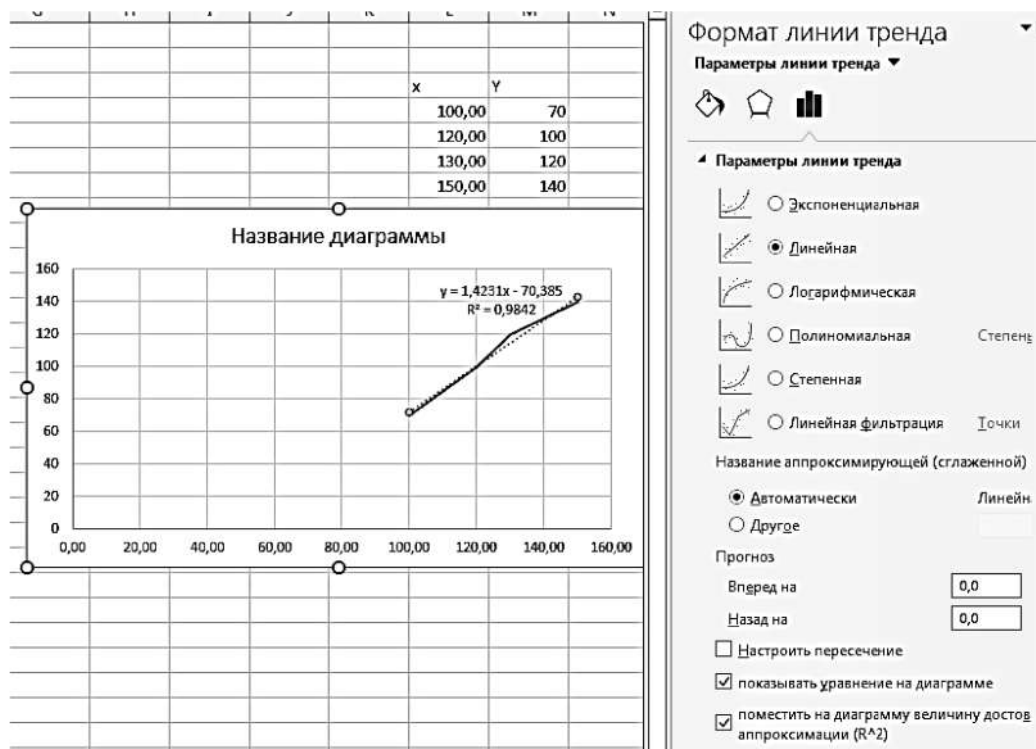


Рисунок 13. Работа с парной линейной регрессией в Excel

Там же можно отобразить величину достоверности аппроксимации (или коэффициент детерминации R^2), которая показывает, насколько модельные значения близки к реальным. Чем ближе эта величина к 1, тем лучше модель описывает реальность. То есть, по сути, это критерий качества модели. Рассчитывается он по формуле

$$R^2 = \frac{\sum \hat{y}_i^2 - n\bar{y}^2}{\sum y_i^2 - n\bar{y}^2},$$

где $\bar{y} = \frac{\sum y_i}{n}$ – среднее значение реальных результатов, \tilde{y} – прогнозное значение.

Областью значения данной величины будет отрезок от 0 до 1.

Кроме расчета величины достоверности аппроксимации есть еще несколько способов оценить качество регрессионной модели. Однако в данном пособии не будем подробно останавливаться на этих методах, а лишь перечислим их:

1. Анализ диаграммы рассеяния;
2. Проверка статистических гипотез о значимости модели;
3. Анализ остатков.

По третьему пункту поясним: метод наименьших квадратов справедлив, если остатки обладают следующими свойствами:

1. Нормальным распределением.
2. Взаимонезависимостью.
3. Нулевым математическим ожиданием.
4. Постоянной дисперсией.

То есть, если остатки разбросаны хаотично, то метод наименьших квадратов можно использовать для решения поставленной задачи.

Если не хаотично, то нужно смотреть: если остатки растут при росте x , то есть их график напоминает диаграмму рассеяния, показанную на рисунке 12, это говорит о гетероскедастичности остатков. В данном случае необходимо использовать взвешенный метод наименьших квадратов. Если график остатков похож на отрезок параболы, то скорее всего была выбрана неправильная модель и необходимо выбрать другую. Если график остатков напоминает, например, синусоиду, то есть они циклически колеблются вверх-вниз, то это свидетельствует об автокорреляции остатков, которая может быть оценена количественно с помощью критерия Дарбина-Уотсона. В этом случае метод наименьших квадратов неприменим, и необходимо использовать другие подходы.

Как было сказано выше, парная линейная регрессия не всегда хорошо описывает данные. Тогда зависимость можно попробовать смоделировать либо кусочно-линейной моделью, то есть построить несколько разных линейных моделей для разных диапазонов значений величины x , либо использовать нелинейные модели: гиперболическую, параболическую, степенную и обратную.

В общем случае при втором подходе оптимизационная задача формулируется так:

$$Q = \sum (f(x_i) - y_i)^2 \rightarrow \min,$$

где $f(x_i)$ – некоторая нелинейная функция.

Для ее решения можно воспользоваться, например, методом Ньютона-Рафсона. Но в некоторых случаях при работе с парной нелинейной регрессией используют прием преобразования модели к линейному виду.

Рассмотрим способ преобразования более подробно. Например, мы выбрали гиперболическую модель, задаваемую уравнением вида

$$y = \beta_0 + \beta_1/x.$$

Тогда, если мы введем замену $z=1/x$, то модель преобразуется к привычному нам линейному виду.

Для параболической модели вида

$$y = \beta_0 + \beta_1 \times x + \beta_2 * x^2$$

замена не выполняется. При ее решении используется подход, аналогичный случаю линейной модели, но в итоговой системе будет три уравнения.

Обратная модель вида

$$y = 1/(\beta_0 + \beta_1 * x)$$

приводится к линейной через замену $z=1/y$.

Степенная модель вида

$$y = \beta_0 \times x^{\beta_1}$$

приводится к линейной логарифмированием:

$$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x).$$

Вводя замены, получаем уравнение:

$$z = \beta_0' + \beta_1 \times t.$$

Однако, выполняя подобные преобразования, мы рискуем исказить взаимосвязи и понизить адекватность модели, поэтому линеаризацию нужно применять крайне аккуратно. К тому же данный подход работает не для всех моделей. В таком случае, как было сказано выше, требуется использовать численный метод Ньютона-Рафсона. Опишем кратко его идею, не вдаваясь в подробности. Поиск решения в нем происходит посредством построения последовательных приближений, то есть на основе простой итерации. Первым шагом задается стартовое приближение около предполагаемого корня. Затем в точке приближения строится касательная к графику функции, для которой находится пересечение с осью абсцисс. Эта точка считается следующим приближением, для которого повторяется процесс. Алгоритм продолжает свою работу до тех пор, пока не будет достигнута необходимая точность.

Мы рассмотрели различные модели парной регрессии, то есть случаи, когда отклик зависит от одного фактора. Однако в некоторых задачах на выходную переменную (отклик) влияет несколько факторов. В таком случае эта зависимость описывается множественной регрессией. Чаще всего в этом случае используются линейные модели. Общий вид зависимости описывается следующим образом:

$$y_i = \beta_0 + \beta_1 \times x_{1i} + \dots + \beta_k \times x_{ki} + \varepsilon_i,$$

где $i=1\dots n$, n – количество наблюдений; k – количество факторов.

При работе с множественной регрессией, как правило, переходят к матричной записи системы:

$$Y=X \times \beta + \varepsilon,$$

где $Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}$ – вектор откликов; $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_n \end{pmatrix}$ – вектор ошибок;

$\beta = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_n \end{pmatrix}$ – вектор параметров; $X = \begin{pmatrix} 1 & x_{11} \dots & x_{k1} \\ 1 & x_{12} \dots & x_{k2} \\ \dots & \dots \dots & \dots \\ 1 & x_{1n} \dots & x_{kn} \end{pmatrix}$ – регрессионная

матрица; n – количество наблюдений; k – количество факторов.

В данном случае задача формулируется следующим образом: для известных X и Y необходимо найти такие β , чтобы $Q = \sum \varepsilon_i^2 \rightarrow \min$.

В матричном виде последнее выражение можно переписать так:

$$Q = \varepsilon^T \varepsilon.$$

При этом

$$\varepsilon = Y - X\beta.$$

Тогда аналогом нормальной системы уравнений парной линейной регрессии в нашем случае будет следующее выражение:

$$\tilde{\beta} = (X^T X)^{-1} X^T Y.$$

Однако при работе с этой формулой могут возникнуть проблемы, связанные с тем, что обратную матрицу $(X^T X)^{-1}$ сложно посчитать. Как правило, это происходит, если ее столбцы сильно коррелированы. Например, если столбцы практически линейно зависят один от другого (имеет место мультиколлинеарность). Если коэффициент корреляции столбцов будет больше 0,8, то обратную матрицу уже не посчитать. Для решения этой проблемы либо исключают один из взаимозависимых факторов, либо прибегают к использованию гребневой регрессии (Ridge regression). Ее суть состоит в том, что для решения проблемы плохой обусловленности матрицы $(X^T X)$ к ней добавляется некое число λ . То есть обращается матрица $(X^T X + \lambda I)$. Полученные с помощью гребневой регрессии оценки параметров являются смещенными (в отличие от МНК-оценок), но доказано, что существует такое λ при котором оценки гребневой регрессии более эффективны, чем МНК-оценки. Однако четких рекомендаций относительно выбора числа λ нет.

Продолжая тему проблем множественной регрессии, необходимо отметить еще одну. В случае непарной регрессии диаграмма рассеяния не применима для отслеживания выбросов, так как пространство будет многомерно. Для решения этой проблемы используют робастные методы. Основную информацию о них можно найти в источнике [14].

Если же говорить об общем алгоритме работы с множественной регрессией, то можно выделить следующие шаги:

1. Подготовка исходных данных, как правило, в виде таблицы (см. таблицу 8):

Таблица 8. Шаблон подготовки данных

№ опыта	y	x_1	x_2	...	x_k
1					
...					
n					

2. Оценивание параметров модели по формуле

$$\tilde{\beta} = (X^T X)^{-1} X^T Y,$$

или при использовании гребневой регрессии – по формуле

$$\tilde{\beta} = (X^T X + \lambda I)^{-1} X^T Y.$$

3. Проверка значимости модели с использованием критерия Фишера. Если по критерию Фишера линейная модель множественной регрессии оказалась незначима, то можно перейти к неполной квадратичной модели и проверить ее. Для двух факторов данная модель записывается следующим образом:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

Далее производят замену $\beta_3 = \beta_{12}$ и $x_3 = x_1 x_2$ и проверяют значимость получившейся модели. Если и она не значима, то переходят к полной квадратичной модели, затем (при ее незначимости) – к неполной кубической и полной кубической. Если же и последняя оказывается незначимой, то переходят к степенной, но здесь возрастает риск возникновения мультиколлинеарности. В принципе, даже кубическая модель уже дает высокую мультиколлинеарность. В случае, когда не удастся найти ни одной значимой модели, признается, что регрессионными методами поставленную задачу решить нельзя, и ищется иной способ решения.

4. Проверка значимости каждого фактора по критерию Стьюдента. Если какой-то фактор оказывается незначимым, то его

убирают из расчета и повторяют его заново. Если незначимых факторов несколько, то убирают по одному.

5. Оценка качества модели с помощью коэффициента детерминации. В зависимости от области задачи приемлемы различные значения коэффициента детерминации. Например, для техники значение 0,9 будет пороговым. Все, что выше его, – хорошее, ниже – не очень. Для экономики порогом будет значение 0,5.

Теперь рассмотрим еще один особый вид регрессии – логистическую регрессию. Данный вид используется, если значение отклика (y) должно быть ограничено каким-либо диапазоном. Например, если y – вероятность некоторого события, то она должна лежать строго в диапазоне от 0 до 1. Рассмотренные ранее модели никак этого не учитывают, поэтому в подобных задачах используют логистическую регрессию. Чаще всего она применяется в задачах бинарной классификации. При $y < 0,5$ – один класс, иначе – другой.

В основе данного вида регрессии лежит следующая функция:

$$y = \frac{1}{1+e^{-x}}.$$

Ее график представлен на рисунке 14.

Соответствующая этой функции регрессионная модель будет иметь вид

$$y = \frac{1}{1+e^{-(\beta_0+\beta_1x_1+\dots+\beta_kx_k)}}.$$

Теоретически, выполнив преобразование линеаризации, ее можно привести к линейной модели множественной регрессии, произведя следующую замену:

$$z = -\ln\left(\frac{1}{y-1}\right).$$

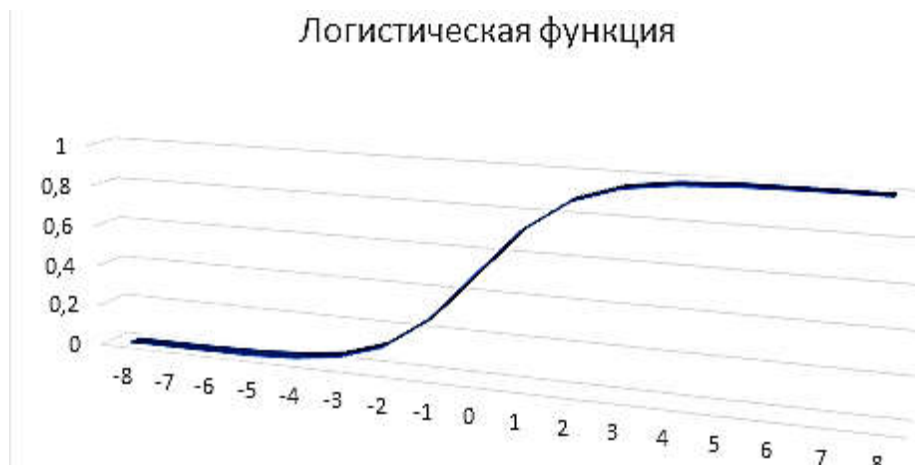


Рисунок 14. График логистической функции

Но так как в процессе преобразований нарушаются основные предпосылки МНК, то в практике этот подход не используется. Вместо этого для оценки параметров модели применяют метод максимального правдоподобия, с которым вам предлагается ознакомиться самостоятельно при необходимости.

Мы закончили рассмотрение статистических методов анализа данных. Теперь перейдем к более абстрактным моделям и алгоритмам, а именно к нечеткой логике.

Модели и методы нечеткой логики

Природа нечетких объектов обусловлена использованием экспертных оценок, которым свойственна неопределенность класса нечеткости. В отличие от стохастической неопределенности, нечеткость затрудняет или даже исключает применение статистических методов и моделей, но может быть использована для принятия предметно-ориентированных решений на основе приближенных рассуждений человека. Основные проблемы, решаемые в нечеткой логике, связаны с моделированием интеллектуальных операций приближенных рассуждений человека (эксперта), а также объектов, над которыми эти операции выполняются:

1. Объектами интеллектуальных операций, используемых в приближенных рассуждениях человека, являются переменные нового

класса – лингвистические переменные, значениями которых являются нечеткие множества. Важным является тот факт, что наименования лингвистической переменной и ее значений должны соответствовать словам, которые использует человек при решении прикладных задач. Таким образом, операндами и результатом интеллектуальных операций являются значения особого вида – *нечеткие множества*.

2. Основные интеллектуальные операции строятся с помощью *операций нечеткой логики*.

3. Алгоритмы вычисления нечетких значений предназначены для манипулирования со значениями, представленными нечеткими множествами на основе операций нечеткой логики, поэтому они классифицируются как нечеткие системы логического вывода. Часто используют сокращенную форму обозначенного класса моделей – *нечеткие модели или нечеткие системы*.

Нечеткие множества

Теория нечетких множеств, введенная Л. Заде [15], – это раздел прикладной математики, посвященный методам анализа неопределенных данных, в которых описание неопределенностей реальных явлений и процессов проводится с помощью понятия о множествах, не имеющих четких границ.

В дальнейшем для указания неопределенности экспертных оценок будем использовать эквивалентное выражение – лингвистическая неопределенность. Л. Заде предложил по аналогии с теорией вероятностей использовать в качестве математической модели лингвистической неопределенности объекта $x \in X$ функцию вида

$$Y = \mu(x, B),$$

где Y – результат вычисления функции, выражающий меру неопределенности (нечеткости) для конкретного объекта $x \in X$;

μ – непрерывная функция, такая, что $\mu: X \rightarrow [0, 1]$. Содержательно функция μ определяет распределение неопределенности на X ;

X – область определения функции μ . Область определения задается упорядоченным множеством значений произвольной природы, называемым *универсальным множеством (или универсумом)*. Носителем функции $\mu(x, B)$ является подмножество $w \subset X$, на котором функция $\mu(x, B)$ принимает значение, отличное от нуля. В качестве универсального множества обычно задается множество действительных чисел;

B – вектор параметров функции, обычно числовых.

Функциональная модель лингвистической неопределенности получила название *нечеткого множества*, так как указанная функция μ рассматривается как характеристическая функция, определенная на множестве объектов X . Таким образом, с математической точки зрения, нечеткое множество моделируется параметрической функцией особого класса, называемого классом *функций принадлежности*. В том случае, если значения функции принадлежности нечеткого множества представлены точными числовыми значениями, такие нечеткие множества относят к *нечетким множествам типа 1*. Если значения функции принадлежности нечеткого множества моделируются другими нечеткими множествами, то такое нечеткое множество относят к *нечетким множествам типа 2*.

На практике используют несколько способов задания функции принадлежности. Среди них выделим следующие:

Структурный способ. Данная форма определения нечетких множеств основана на табличном представлении функций. В случае, если известен вектор параметров B , табличное представление функции принадлежности может быть задано явно путем табулирования функции $Y = \mu(x, B)$ на множестве значений w , являющемся ее носителем. При неизвестном векторе параметров B – путем прямого перечисления множества пар в виде

$$Y = \{\mu_1/x_1, \mu_2/x_2, \dots, \mu_n/x_n\}.$$

Данная форма удобна для графического отображения нечеткого множества и используется часто в тех случаях, когда затруднительно

задать математический вид функции $Y = \mu(x, B)$, например, если X не является множеством чисел.

Рассмотрим пример записи нечеткого множества в явной форме.

Пусть $w = \{x_1, x_2, x_3, x_4, x_5\}$, A – нечеткое множество, для которого $\mu_A(x_1) = 0,3; \mu_A(x_2) = 0; \mu_A(x_3) = 1; \mu_A(x_4) = 0,6; \mu_A(x_5) = 0,9$.

Тогда A можно представить в виде

$$A = \{0,3/x_1; 0/x_2; 1/x_3; 0,6/x_4; 0,9/x_5\}.$$

Функциональный способ. При этом предполагается, что форма функции принадлежности, моделирующей нечеткое множество, известна и определена на множестве действительных чисел X . Для представления $Y = \mu(x, B)$ используют различные функции (показанные, например, на рисунке 15: а) – треугольная, б) – трапецеидальная, в) – Гауссова).

Гауссова функция принадлежности описывается вектором параметров $B = \{\sigma, c\}$ и формулой

$$\mu(x, B) = \exp\left(-\left(\frac{x-c}{\sigma}\right)^2\right),$$

где c – среднее значение;

σ – среднее квадратичное отклонение.

Треугольная функция принадлежности характеризуется тройкой чисел, $B = \{a, b, c\}$, и вычисляется по формуле

$$\mu(x, B) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{c-x}{c-b}, & b < x \leq c \\ 0, & x < a, x > c, \end{cases}$$

где b – задает координату вершины треугольника;

a, c – определяют основание треугольника.

По аналогии задается и трапецеидальная функция принадлежности, которая характеризуется четверкой чисел $B = \{a, b, c, d\}$:

$$\mu(x, B) = \begin{cases} \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x < c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & x < a, x > d. \end{cases}$$

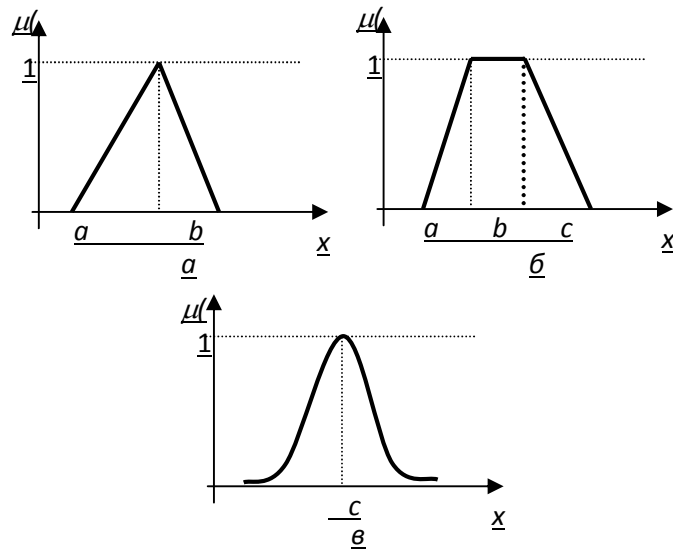


Рисунок 15. Типовые формы функций принадлежности

На практике часто параметр B явно не указывается для обеспечения более компактной записи функции принадлежности, то есть используется функциональная запись вида $Y = \mu(X)$ вместо $Y = \mu(x, B)$. При дальнейшем изложении в учебном пособии будем использовать компактную запись функции принадлежности.

С каждой функцией принадлежности $\mu(X)$ сопоставляется лингвистическое обозначение нечеткого множества (лингвистический терм). Тогда функция принадлежности нечеткого множества Z (функциональный способ) будет иметь следующую запись $Z = \mu_Z(X)$. Расширив традиционное понятие множества, Л. Заде построил «математический мостик» при описании свойств понятий в виде нечетких множеств между числом x , свойством Z , выраженным лингвистически, и степенью соответствия числа x свойству Z в виде функции

$\mu_Z(x)$. Фактически обозначение нечеткого множества через Z позволяет именовать функцию принадлежности $\mu_Z(X)$, в общем случае параметрическую, лингвистическими терминами, то есть оперировать с ней как со значениями лингвистической переменной. Часто эти два понятия – Z и $\mu_Z(X)$ – рассматриваются как эквивалентные.

Пусть $X = \{x\}$ – совокупность объектов (универсальное множество), обозначаемых через x . Пусть на X определены три нечетких множества: A = «низкое», B = «удовлетворительное», C = «хорошее», обозначающие имена свойств понятия «качество». Тогда, следуя структурному способу для всех $x \in X$, нечеткое множество A может быть задано совокупностью упорядоченных пар $A = \{x, \mu_A(x)\}$, нечеткое множество B – совокупностью упорядоченных пар $B = \{x, \mu_B(x)\}$, нечеткое множество C – совокупностью упорядоченных пар $C = \{x, \mu_C(x)\}$. Используя функциональный способ записи, приведенный выше, нечеткие множества будут представлены следующим образом:

$$A = \mu_A(x), B = \mu_B(x), C = \mu_C(x).$$

Лингвистические переменные

Лингвистическая переменная – это переменная, значениями которой являются слова или высказывания естественного или искусственного языка.

Согласно [16]: «Поскольку слова в общем смысле менее точны, чем числа, понятие лингвистической переменной дает возможность приближенно описывать явления, которые настолько сложны, что не поддаются описанию в общепринятых количественных терминах... высокая точность несовместима с высокой сложностью. Таким образом, быть может, именно по этой причине обычные методы анализа систем и моделирования на ЭВМ, основанные на точной обработке численных данных, по существу не способны охватить огромную сложность процессов человеческого мышления и принятия решений. Отсюда напрашивается вывод о том, что для получения

существенных выводов о поведении гуманистических систем придется, по-видимому, отказаться от высоких стандартов точности и строгости, которые мы, как правило, ожидаем при математическом анализе четко определенных механистических систем, и относиться более терпимо к иным подходам, которые являются приближенными по своей природе».

Любая переменная описывается множеством допустимых значений, а лингвистические понятия описываются набором присущих им свойств. Л. Заде расширил понятие обычной лингвистической переменной, допустив, что в качестве ее значений (термов) выступают нечеткие переменные [16]. Пример лингвистической переменной, заимствованный из [17], представлен на рисунке 16.

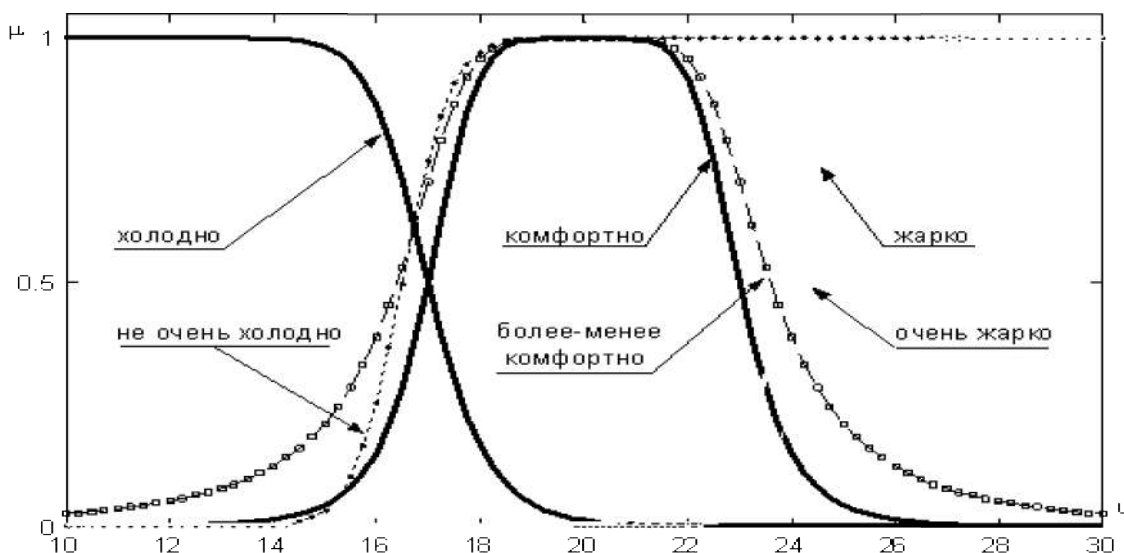


Рисунок 16. Пример лингвистической переменной «Температура»

Формально лингвистическая переменная описывается набором

$$\langle Name, \tilde{X}, X, G, P \rangle,$$

где *Name*– наименование лингвистической переменной;

X– универсальное множество объектов *x*;

\tilde{X} – базовое терм-множество, образующее совокупность термов лингвистической переменной, например, $\tilde{X} = \{ \langle \text{Отличный} \rangle, \langle \text{Хороший} \rangle, \langle \text{Плохой} \rangle, \langle \text{Удовлетворительный} \rangle \text{ и др.} \}$;

G – синтаксические правила вывода (порождения) новых термов \tilde{X}^* , не входящих в базовое терм-множество, задаваемые обычно на основе контекстно-свободной грамматики;

P – семантические правила, контекстно-зависимый способ вычисления смысла на основе функций принадлежности каждого терма из $\tilde{X} \cup \tilde{X}^*$.

Операции нечеткой логики

Нечеткая логика – это логика, оперирующая нечеткими высказываниями и рассуждениями на базе частичной истинности.

В основе операций нечеткой логики лежит понятие нечеткого множества, выраженного функцией принадлежности. Поэтому операндами и результатами операций нечеткой логики являются также функции, определяющие новые нечеткие множества.

В нечеткой логике для моделирования основных логических связей И (\wedge), ИЛИ (\vee) над нечеткими множествами используют триангулярные нормы [18].

Триангулярной нормой (t-нормой) называют отображение $T : [0,1] \times [0,1] \rightarrow [0,1]$, удовлетворяющее следующим условиям:

$$T(0, 0) = 0; T(x, 1) = x; T(1, x) = x \text{ – ограниченность;}$$

$$T(x, y) \leq T(a, b), \text{ если } x \leq a, y \leq b \text{ – монотонность;}$$

$$T(x, y) = T(y, x) \text{ – коммутативность;}$$

$$T(x, T(y, z)) \leq T(T(x, y), z) \text{ – ассоциативность.}$$

Триангулярной конормой (s-конормой) называют отображение $S : [0,1] \times [0,1] \rightarrow [0,1]$, удовлетворяющее следующим условиям:

$$S(1, 1) = 1; S(x, 0) = x; S(0, x) = x \text{ – ограниченность;}$$

$$S(x, y) \geq S(a, b), \text{ если } x \geq a, y \geq b \text{ – монотонность;}$$

$$S(x, y) = S(y, x) \text{ – коммутативность;}$$

$$S(x, S(y, z)) \leq S(S(x, y), z) \text{ – ассоциативность.}$$

t-норма и s-конорма в определенном смысле являются двойственными понятиями. Эти функции могут быть получены друг из

друга, например, с помощью инволютивного отрицания и законов Де Моргана следующим образом:

$$S(x, y) = n(T(n(x), n(y))), \quad T(x, y) = n(S(n(x), n(y))) .$$

Простейшими примерами t-норм и s-конорм, взаимно связанных этими соотношениями для $n(x)=1-x$, являются следующие (таблица 9).

Таблица 9. Примеры t-норм и s-конорм

Формула	Интерпретация
$T(x, y) = \min\{x, y\}$	(минимум)
$S(x, y) = \max\{x, y\}$	(максимум)
$T(x, y) = xy$	(произведение)
$S(x, y) = x+y-xy$	(вероятностная сумма)
$T(x, y) = \max\{x+y-1, 0\}$	(t-норма Лукасевича)
$S(x, y) = \min\{x+y, 1\}$	(t-конорма Лукасевича ограниченная сумма)

Основные операции с нечеткими множествами

1. Операция эквивалентности

$$A \equiv B \Leftrightarrow \forall x \in X \quad \mu_A(x) = \mu_B(x)$$

2. Операция включения

$$A \subseteq B \Leftrightarrow \mu_A(x) \leq \mu_B(x), \forall x \in X$$

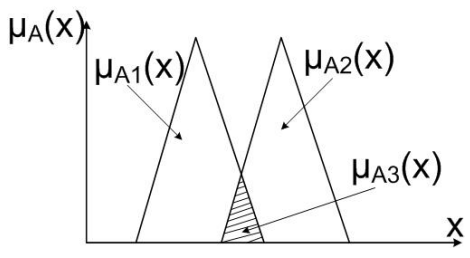
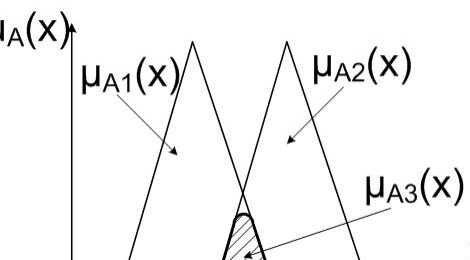
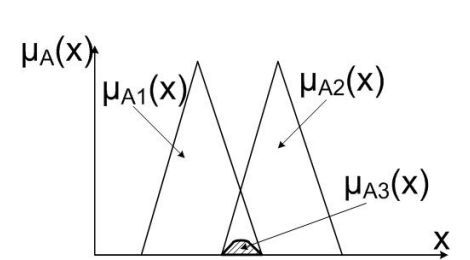
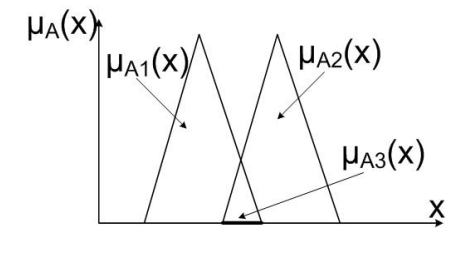
4. Нечеткая операция «НЕ» (дополнение) показана в таблице 10.

Таблица 10. Нечеткое «НЕ»

Иллюстрация	Формула
	$\overline{\mu}_A(x) = 1 - \mu_A(x) \text{ — нечеткая операция НЕ по Заде}$ $\overline{\mu}_A(x) = \frac{1 - \mu_A(x)}{1 + \lambda \cdot \mu_A(x)} \text{ — НЕ по Сугено}$

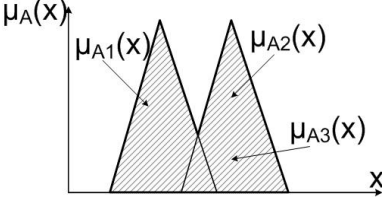
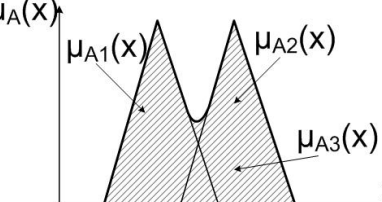
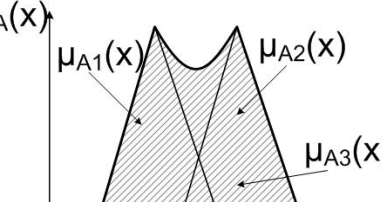
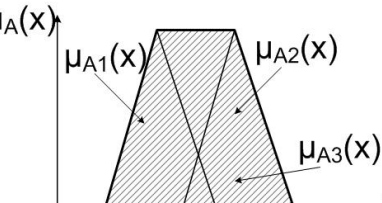
5. Нечеткая операция «И» показана в таблице 11.

Таблица 11. Нечеткое «И»

Иллюстрация	Формула
логическое произведение (Заде)	
	$\mu_{A_3}(x) = \mu_{A_1 \wedge A_2}$ $= \min(\mu_{A_1}(x), \mu_{A_2}(x))$ <p style="text-align: center;">min – конъюнкция</p>
алгебраическое произведение (Бандлер, Коходт)	
	$\mu_{A_3}(x) = \mu_{A_1}(x) \cdot \mu_{A_2}(x)$
граничное произведение (Лукасевич, Гринс)	
	$\mu_{A_3}(x) = \mu_{A_1 \otimes A_2}$ $= \max(\mu_{A_1}(x) + \mu_{A_2}(x) - 1, 0)$
драстическое произведение (Вебер)	
	$\mu_{A_3}(x) = \mu_{A_1}(x) \triangle \mu_{A_2}(x) = \begin{cases} \mu_{A_1}(x), & \text{если } \mu_{A_2}(x) = 1 \\ \mu_{A_2}(x), & \text{если } \mu_{A_1}(x) = 1 \\ 0, & \text{иначе} \end{cases}$ $0 \leq \mu_{A_1 \wedge A_2} \leq \mu_{A_1 \otimes A_2} \leq \mu_{A_1 \cdot A_2} \leq \mu_{A_1 \wedge A_2}$

6. Нечеткая операция «ИЛИ» показана в таблице 12.

Таблица 12. Нечеткое «ИЛИ»

Иллюстрация	Формула
логическая сумма (Заде)	
	$\mu_{A_3}(x) = \mu_{A_1 \vee A_2} = \max(\mu_{A_1}(x), \mu_{A_2}(x))$ <p style="text-align: center;">max-дизъюнкция</p>
алгебраическая сумма	
	$\mu_{A_3}(x) = \mu_{A_1}(x) + \mu_{A_2}(x) - \mu_{A_1}(x) \cdot \mu_{A_2}(x),$ <p style="text-align: center;">где $\forall x \in X$</p>
граничная сумма	
	$\mu_{A_3}(x) = \mu_{A_1 \oplus A_2} = \min(\mu_{A_1}(x) + \mu_{A_2}(x), 1)$
драстическая сумма	
	$\mu_{A_3}(x) = \begin{cases} \mu_{A_1}(x), & \text{если } \mu_{A_2}(x) = 0 \\ \mu_{A_2}(x), & \text{если } \mu_{A_1}(x) = 0 \\ 1, & \text{иначе} \end{cases}$

По существу, все человеческие понятия являются нечеткими, так как они получаются в результате группировки (clumping) точек или объектов, объединяемых по сходству. Тогда нечеткость подобных

групп (clumps) есть прямое следствие нечеткости понятия сходства. Простыми примерами таких групп являются понятия «средний возраст», «деловая часть города», «немного облачно», «бестолковый» и др. Данную группу в нечеткой логике называют «гранулой» (*granule*). В естественном языке (ЕЯ) слова играют роль меток гранул и служат для сжатия данных. Сжатие данных с помощью слов является ключевым аспектом человеческих рассуждений и формирования понятий.

В нечеткой логике гранулирование информации лежит в основе понятий лингвистической переменной и нечетких правил типа «ЕСЛИ-ТО», задаваемой операцией импликации.

Операция импликации

В качестве основного математического инструмента при определении импликации $A \rightarrow B$ для нечетких множеств A и B используют композиционное правило Л. Заде [16], являющееся обобщением правила *modusponens* (таблица 13).

Таблица 13. Импликация

Описание	Формула
Предпосылка	$A \rightarrow B$
Событие	A^*
Вывод	$A^* \circ (A \rightarrow B)$

Пусть U и V – два универсальных множества с базовыми переменными u и v соответственно. Пусть A и F – нечеткие подмножества множеств U и $U \times V$. Тогда композиционное правило вывода утверждает, что из нечетких множеств A и F следует нечеткое множество $B = A \circ F$. Функция принадлежности результата вычисляется с помощью триангулянтных норм следующим образом:

$$\mu_B(v) = S(T(\mu_A(u), \mu_F(u, v))),$$

при моделировании s-конормы и t-нормы операциями (\vee) *max* и (\wedge) *min* соответственно:

$$\mu_B(v) = \bigvee_{u \in U} ((\mu_A(u) \wedge \mu_F(u, v))).$$

Другие формулы, реализующие операцию импликации, и математические объекты теории нечетких множеств и нечеткой логики приведены в учебном пособии [19].

Формализация нечеткой импликации позволила задать правила «ЕСЛИ-ТО» в виде нечетких продукционных правил и заложило основу нечеткого моделирования опыта и знаний экспертов, выраженных в виде приближенных зависимостей.

Нечеткие системы

Целью построения нечетких систем сложных явлений является приближенное описание зависимости (аппроксимация некоторой функции) $Y = f(X)$,

где Y – выходная лингвистическая переменная;

X – вектор входных лингвистических переменных. Его размерность – n ;

f – зависимость между X и Y , описываемая совокупностью нечетких правил.

В основе нечетких систем лежат совокупность нечетких правил «ЕСЛИ-ТО», описывающих зависимости между нечеткими переменными предметной области, композиционное правило вывода и способ вычисления значений нечетких переменных (способ нечеткого вывода).

Системой нечеткого логического вывода в теории нечетких множеств и нечеткой логики называется модель, которая описывает поведение систем на естественном (или близком к естественному) языке в виде приближенных рассуждений на основе композиционного правила вывода.

В систему нечеткого логического вывода входят следующие объекты (см. рисунок 17):

- совокупность нечетких правил (база правил);
- набор функций принадлежности базы нечетких переменных (база переменных);

- блок фаззификации;
- блок дефаззификации;
- блок вывода.

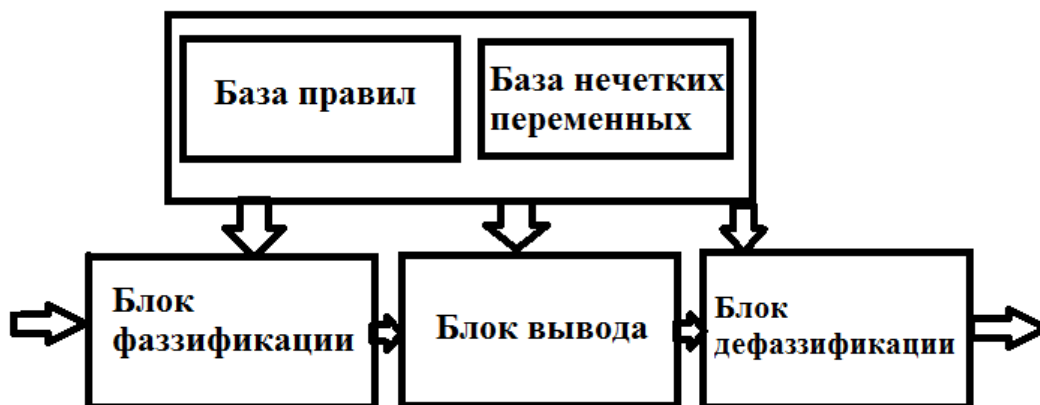


Рисунок 17. Система нечеткого вывода

База правил хранит множество логических правил вывода, а также их порядок (иерархическую структуру) применения. База нечетких переменных содержит названия лингвистических термов и параметры их функций принадлежности. База правил вместе с базой нечетких переменных образуют *базу знаний* (БЗ) системы нечеткого вывода.

Простейшие системы нечеткого логического вывода основаны на правилах вида:

R_i : Если X есть A_i и Y есть B_i , то Z есть C_i ,

R_i : Если X есть A_i и Y есть B_i , то $z=f_i(x,y)$,

где X, Y – входные нечеткие переменные;

Z – выходная нечеткая переменная;

A_i, B_i – входные значения (функции принадлежности);

C_i – выходные нечеткие значения (функции принадлежности);

f_i – некоторые вещественные функции.

При этом должны соблюдаться следующие условия:

- Существует хотя бы одно правило для каждого лингвистического терма выходной переменной.

- Для любого термина входной переменной имеется хотя бы одно правило, в котором этот терм используется в качестве предпосылки (левая часть правила).

В противном случае имеет место неполная база нечетких правил.

Распространены пять способов реализации нечеткого логического вывода [18].

Схема 1: Алгоритм Мамдани (Mamdani). Импликация моделируется минимумом, а агрегация – максимумом.

Схема 2: Алгоритм Цукамото (Tsukamoto). Исходные посылки – как у предыдущего алгоритма, но предполагается, что функции принадлежности являются монотонными.

Схема 3. Алгоритм Суджено (Sugeno). Алгоритм предполагает, что правые части правил вывода представлены в виде линейных функций.

Схема 4. Алгоритм Ларсена (Larsen). В алгоритме Ларсена нечеткая импликация моделируется с использованием операции умножения.

Схема 5. Упрощенный алгоритм нечеткого вывода. Исходные правила в данном случае задаются в виде

Если X есть A_i и Y есть B_i , то $z = Z_i$,

где Z_i – четкое значение.

Рассмотрим алгоритм нечеткого вывода по схеме Мамдани для базы правил вида R_i : *Если X есть A_i и Y есть B_i , то Z есть C_i , $i=[1,r]$.*

1. Фаззификация.

Определяются степени истинности по функциям принадлежности для левых частей каждого правила:

$$a_i = \mu_{A_i}(X)$$

$$b_i = \mu_{B_i}(Y),$$

где a_i – степень принадлежности X к A_i ;

b_i – степень принадлежности Y к B_i ;

$i = [1,r]$;

r – количество правил.

2. Импликация.

Определяется сила каждого правила, t -нормой является логический минимум:

$$\alpha_i = \min(a_i, b_i).$$

Модифицируются функции принадлежности переменной z в каждом правиле:

$$\mu'_i(z) = \min(\alpha_i, \mu_i(z)),$$

где $\mu_i(z)$ – функция принадлежности переменной Z_i .

3. Агрегация.

Объединение выходов каждого правила логическим максимумом (s -конорма):

$$\mu'(z) = \max_i(\mu_i(z))$$

4. Деффазификация.

Простейшим способом выполнения процедуры дефаззификации является выбор четкого числа, соответствующего максимуму функции принадлежности. Однако пригодность этого способа ограничивается одноэкстремальными функциями принадлежности. Для многоэкстремальных функций на практике часто используется метод центра тяжести.

Дефаззификация нечеткого множества по методу центра тяжести осуществляется по формуле

$$x^0 = \frac{\int x \cdot \mu(x) dx}{\int \mu(x) dx}.$$

Физическим аналогом этой формулы является нахождение центра тяжести плоской фигуры, ограниченной осями координат и графиком функции принадлежности нечеткого множества. В случае дискретного универсального множества дефаззификация нечеткого множества по методу центра тяжести осуществляется по формуле

$$x^0 = \frac{\sum_{i=1}^k x_i \cdot \mu(x_i)}{\sum_{i=1}^k \mu(x_i)}$$

На рисунке 18 графически представлен процесс нечеткого вывода по алгоритму Мамдани.

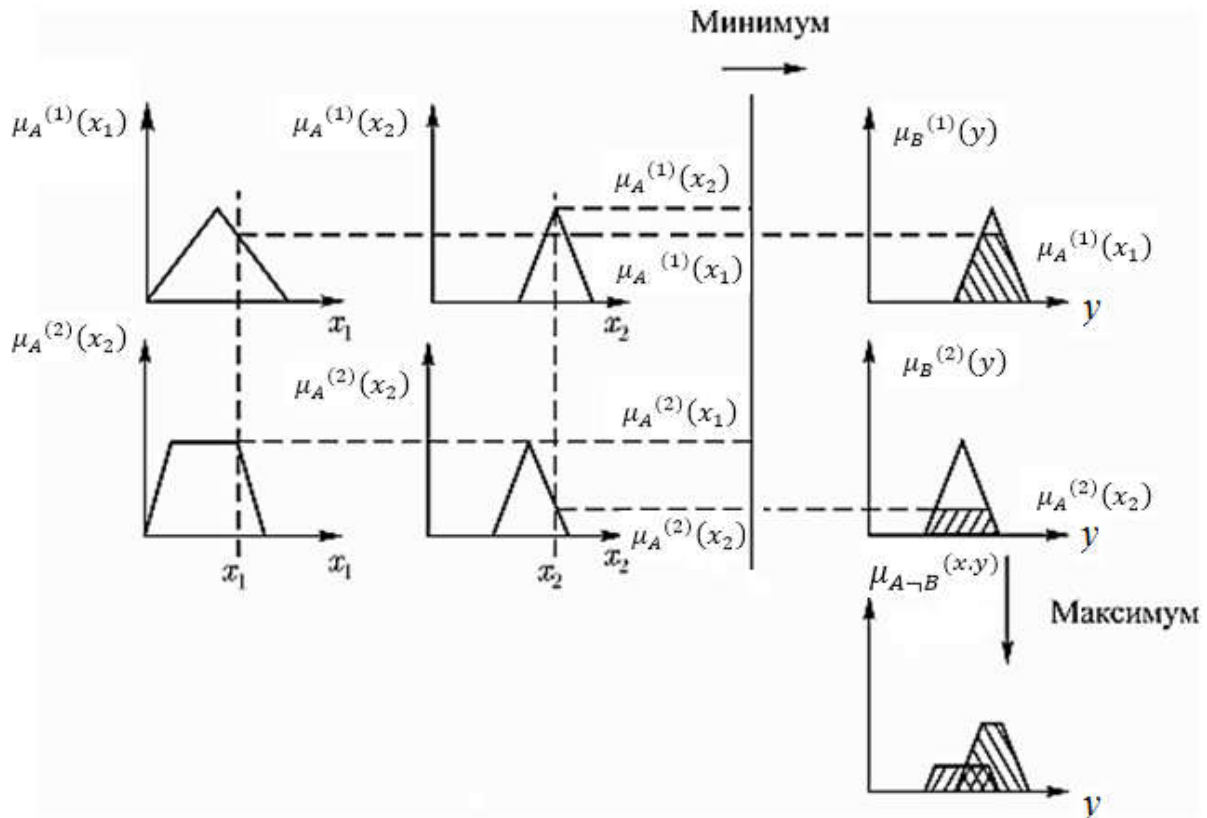


Рисунок 18. Алгоритм нечеткого вывода Мамдани

Пусть дана система управления с двумя правилами нечеткого управления:

Правило 1: IF x is A1 AND y is B1 THEN z is C1;

Правило 2: IF x is A2 AND y is B2 THEN z is C2.

Предположим, что величины x_0 и y_0 , считываемые датчиком, являются четкими входными величинами для лингвистических переменных x и y , и что заданы следующие функции принадлежности для нечетких подмножеств $A_1, A_2, B_1, B_2, C_1, C_2$ этих переменных:

$$\mu_{A_1}(x) = \begin{cases} \frac{x-2}{3} & 2 \leq x \leq 5; \\ \frac{8-x}{3} & 5 \leq x \leq 8; \end{cases} \quad \mu_{A_2}(x) = \begin{cases} \frac{x-3}{3} & 3 \leq x \leq 6; \\ \frac{9-x}{3} & 6 \leq x \leq 9; \end{cases}$$

$$\mu_{B_1}(y) = \begin{cases} \frac{y-5}{3} & 5 \leq y \leq 8; \\ \frac{11-y}{3} & 8 \leq y \leq 11; \end{cases} \quad \mu_{B_2}(y) = \begin{cases} \frac{y-4}{3} & 4 \leq y \leq 7; \\ \frac{10-y}{3} & 7 \leq y \leq 10; \end{cases}$$

$$\mu_{C_1}(z) = \begin{cases} \frac{z-2}{3} & 1 \leq z \leq 4; \\ \frac{7-z}{3} & 4 \leq z \leq 7; \end{cases} \quad \mu_{C_2}(z) = \begin{cases} \frac{z-3}{3} & 3 \leq z \leq 6; \\ \frac{9-z}{3} & 6 \leq z \leq 9; \end{cases}$$

Предположим, что в момент времени t_1 были считаны значения датчиков $x_0(t_1) = 4$ и $y_0(t_1) = 8$. Проиллюстрируем, как при этом будет вычисляться величина выходного сигнала.

Первым шагом находим α -срезы для первого и второго правила на основе заданных функций принадлежности (с учетом значений x_0 и y_0). С этой целью вычисляем величины функций принадлежности в заданных точках для первого и второго правил:

$$\mu_{A_1}(x_0=4) = 2/3 \quad \text{и} \quad \mu_{B_1}(y_0=8) = 1;$$

$$\mu_{A_2}(x_0=4) = 1/3 \quad \text{и} \quad \mu_{B_2}(y_0=8) = 2/3.$$

Затем, в соответствии с правилом вывода по Мамдани (выбор минимального значения функций принадлежности), определяем:

$$\alpha_1 = \min(\mu_{A_1}(x_0), \mu_{B_1}(y_0)) = \min(2/3, 1) = 2/3;$$

$$\alpha_2 = \min(\mu_{A_2}(x_0), \mu_{B_2}(y_0)) = \min(1/3, 2/3) = 1/3.$$

Результат применения вычисленных значений α_1 и α_2 показан на рисунке 19. Окончательный результат получается путем объединения найденных значений функций принадлежности с использованием оператора максимума (с учетом стратегии вывода по Мамдани). Результирующая функция принадлежности также представлена на рисунке 19.

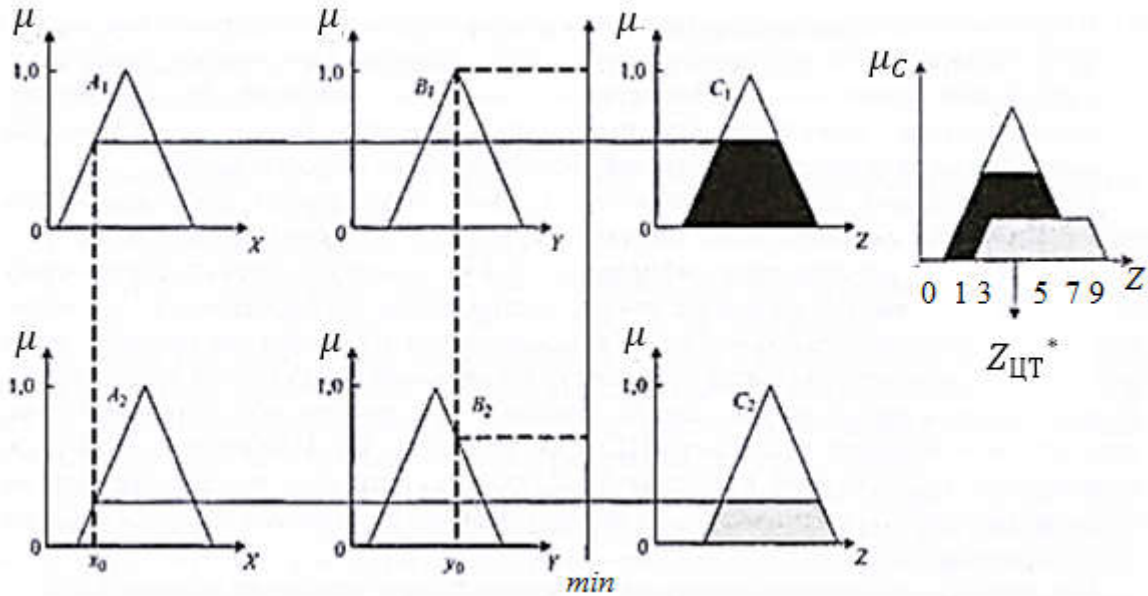


Рисунок 19. Иллюстрация нечеткого вывода по Мамдани

Для вычисления искомой выходной величины Z проводим дефазификацию нечеткой величины μ_C . По методу центра тяжести получаем

$$Z_{\text{ЦТ}}^* = \frac{2\left(\frac{1}{3}\right) + 3\left(\frac{2}{3}\right) + 4\left(\frac{2}{3}\right) + 5\left(\frac{2}{3}\right) + 6\left(\frac{1}{3}\right) + 7\left(\frac{1}{3}\right) + 8\left(\frac{1}{3}\right)}{\left(\frac{1}{3}\right) + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right) + \left(\frac{2}{3}\right) + \left(\frac{1}{3}\right) + \left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)} = 47.$$

Нечеткая логика в анализе временных рядов

Предположим, что задан процесс, состояния которого описываются n значениями одной переменной. Пусть в результате наблюдения получен временной ряд этой переменной, представляющий последовательность упорядоченных в равноотстоящие моменты времени пар $\{x_i, t_i\}$, таких, что $\forall x_i \in X, X \subset R^1, t_i \in N, i \in [1, n]$. Значение x_i – *уровень* временного ряда. Тогда введем понятие нечеткого временного ряда.

Нечетким временным рядом (НВР) называют упорядоченную в равноотстоящие моменты времени последовательность наблюдений над некоторым процессом, состояния которого изменяются во вре-

мени, если значение состояния процесса в момент t_i может быть выражено с помощью нечеткой метки \tilde{x}_i .

Нечеткая метка \tilde{x}_i может быть сформирована непосредственно экспертом в форме

$$\mu_{\tilde{x}_i}(w, x_i),$$

где $\tilde{x}_i \in \tilde{X}$, \tilde{X} – множество нечетких меток;

w – носитель (интервал на X) нечеткой метки \tilde{x}_i , $x_i \in w$;

$\mu_{\tilde{x}_i}(w, x_i) \in [0,1]$ – функция принадлежности нечеткой метки \tilde{x}_i уровню временного ряда x_i , обычно треугольной формы.

Носитель нечеткой метки \tilde{x}_i – это четкое множество $w \subseteq B$ таких точек $x_i \in w$, для которых $\mu_{\tilde{x}_i}(w) > 0$, где $B \subset X$ – базовое множество нечетких меток \tilde{X} .

Таким образом, нечеткий временной ряд формируется в результате интервального качественного оценивания уровней числового ВР. Интервалы-носители нечетких меток, образованные на множестве X , обязательно пересекаются. Качественный аспект нечеткой метке придает функция $\mu_{\tilde{x}_i}(w) \in [0,1]$.

На рисунке 20 изображен абстрактный нечеткий временной ряд, где каждой нечеткой метке \tilde{x}_i соответствует нечеткое множество, задаваемое функцией принадлежности.

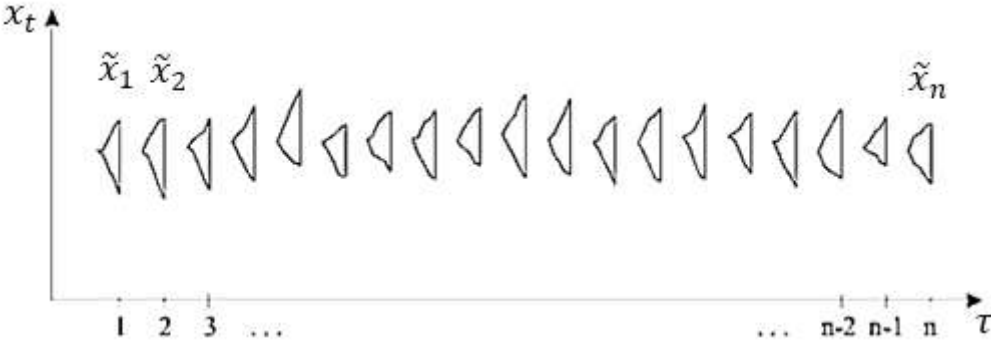


Рисунок 20. Абстрактный нечеткий временной ряд

В отличие от традиционного временного ряда значениями нечеткого ВР являются нечеткие множества, а не действительные значения уровней ВР.

В 1993 году ученые Сонг (Song) и Чиссом (Chissom) [20] предложили нечеткие модели детерминированных (time-variant) и авторегрессионных (time-invariant) временных рядов первого порядка (first-order) и применили разработанные модели для прогнозирования количества регистрирующихся студентов университета штата Алабама (США), фаззифицировав предварительно четкий временной ряд. Это было первое применение нечетких моделей при моделировании ВР и первое определение моделей нечетких временных рядов.

Пусть $X_t, (t = 1, \dots) \subset R^1$ – универсальное множество, на котором определены нечеткие множества $y_t^i, (i = 1, 2, \dots)$ и Y_t – коллекция $y_t^i, (i = 1, 2, \dots)$. Тогда Y_t называется нечетким ВР.

На практике в большинстве временных рядов последовательные наблюдения зависимы, так что:

$$R = \{(y_t, y_{t-1}), (y_{t-1}, y_{t-2}) \dots\} \subseteq Y_t \times Y_{t-1},$$

где Y_t, Y_{t-1} обозначают переменные;

y_t, y_{t-1} – наблюдаемые значения этих переменных.

Наиболее частой моделью зависимости является явная функция:

$$f: Y_{t-1} \rightarrow Y_t,$$

представленная линейной функцией:

$$y_t = f(y_{t-1}, \phi, \varepsilon) = \phi y_{t-1} + \varepsilon_t,$$

где ε_t – случайная ошибка, шум.

В случае нечеткого временного ряда в качестве модели авторегрессии используется нечеткое разностное уравнение:

$$y_t^j = y_{t-1}^i \circ R_{ij}(t, t-1),$$

$$y_t^i \in Y_t, y_{t-1}^i \in Y_{t-1}, i \in I, j \in J,$$

где \circ – обозначает операцию композиции из теории нечетких множеств;

$R(t, t-1) = \bigcup_{i,j} R_{ij}(t, t-1)$ – система нечетких отношений, которая

символически может быть записана в виде $Y_t \rightarrow Y_{t-1}$.

Систему отношений R в выражении $Y_t = Y_{t-1} \circ R(t, t-1)$ называют *моделью нечеткого временного ряда* первого порядка. Данная модель – важный частный случай общей модели порядка p :

$$Y_t = (Y_{t-1} \times Y_{t-2} \times \dots \times Y_{t-p}) \circ R(t, t-p),$$

$$R(t, t-p) = \max_p \left\{ \min_{j, i_1, i_2, \dots, i_p} \left\{ y_t^j, y_{t-1}^{i_1}, \dots, y_{t-p}^{i_p} \right\} \right\}.$$

Метод моделирования нечетких временных рядов

Моделирование нечетких временных рядов в соответствии с нечеткой моделью, предложенной в работе [20], состоит в реализации следующих шагов:

1. Определение нечетких переменных – разбиение данных на множество интервалов (носителей нечетких множеств), определение лингвистических значений нечетких множеств и их функций принадлежности.
2. Формирование логических отношений $Y_t \rightarrow Y_{t-1}$.
3. Фаззификация входных данных – определение степени принадлежности входных данных входным нечетким переменным.
4. Вычисление результата применения нечеткого правила $R_{ij}(t, t-1)$ для каждой импликации.
5. Вычисление результирующего отношения R как объединения $\bigcup_{i,j} R_{ij}(t, t-1)$.
6. Применение полученной модели к входным данным и получение выходных нечетких результатов.
7. Дефаззификация нечетких результатов.

Одной из проблем в нечетком моделировании ВР является отсутствие четких рекомендаций на первом этапе построения модели

по выбору количества и параметров нечетких множеств, моделирующих входные и выходные переменные, в частности по определению их носителей (длины интервалов). Данные задачи выполняются экспертом, и, как показывают исследования, от выбора интервалов сильно зависит результат исследования.

Пример моделирования временного ряда в нечетком подходе

Приведем пример нечеткого моделирования временного ряда для прогнозирования прямых валютных котировок ЦБ USD/RUB за июнь 2005 года, описанный в работе [21] и представляющий модификацию метода Сонга, отличающуюся (а) использованием изменений (приращений) данных прошлого вместо реальных числовых значений (регистрации или валютного курса), и (б) вычислением отношений R_j для предсказания будущих состояний.

Рассматриваемый в работе метод был изначально успешно применен к временному ряду, характеризующему количество поступающих в Алабамский университет, которые являются бенчмаркингом при сравнении методов моделирования нечетких временных рядов.

Анализ результативности предлагаемого метода по показателю точности средней относительной ошибки аппроксимации для 6 нечетких множеств (MAPE=2,42) показал, что предложенный метод превышает аналогичный показатель для этого ВР, полученный методом Сонга и Чена (рисунок 21).

Применительно к проблеме прогнозирования валютного курса USD/RUB пошаговое описание предлагаемого метода нечеткого моделирования ВР можно свести к следующему:

Шаг 1: Задание области определения (универсального множества U) проблемы, исходя из вычисленных приращений валютного курса в течение рассматриваемого интервала времени.

Имеются следующие данные. Наибольшее положительное приращение курса доллара по отношению к российскому рублю наблю-

дается в феврале 2002 года, т. е. по сравнению с январским значением рост составляет 0.3679 (более 36 копеек/месяц). В ноябре-декабре 2004 года происходит самое значительное за трехлетний период наблюдений падение котировки доллара практически на 70 копеек (-0.6949). В результате, с целью упрощения последующего разбиения на равновеликие интервалы, полученные граничные значения (-0.6949 и +0.3679) слегка корректируются. Тогда, например, в случае использования шести подынтервалов U может быть представлена отрезком $[-0.7, -0.5]$.

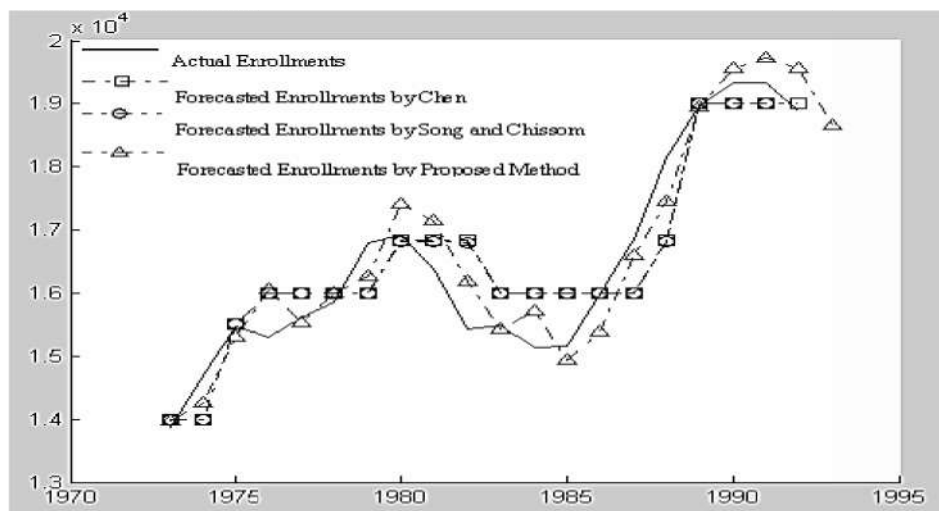


Рисунок 21. Сравнение результатов нечеткого моделирования

Шаг 2: Разбиение множества U на интервалы одинаковой длины.

Если мы оперируем с шестью нечеткими множествами, то область определения делится на 6 интервалов $u_i, i = \overline{1,6}$, $u_1 = [-0.7, -0.5], u_2 = [-0.5, -0.3], \dots, u_6 = [0.3, 0.5]$ (в действительности количество нечетких множеств не обязательно должно совпадать с числом интервалов разбиения).

Шаг 3: Определение нечетких множеств A_i .

Предположим, что лингвистическая переменная «изменение валютного курса» характеризуется терм-множеством, образуемым следующими значениями: A_1 (значительное уменьшение), A_2

(уменьшение), A_3 (без изменений/флэт), A_4 (увеличение), A_5 (значительное увеличение), A_6 (очень большое увеличение).

Для шести построенных выше интервалов $u_i, i = \overline{1,6}$, факт принадлежности каждого конкретного u_i определенному множеству $A_j, j = \overline{1,6}$ выражается действительным числом из единичного интервала $[0,1]$ (предполагается, что элементы, отсутствующие в представлении множеств A_j , характеризуются нулевой степенью принадлежности):

$$\begin{aligned} A_1 &= \{1/u_1 + 0.5/u_2\} \\ A_2 &= \{0.5/u_1 + 1/u_2 + 0.5/u_3\} \\ A_3 &= \{0.5/u_2 + 1/u_3 + 0.5/u_4\} \\ A_4 &= \{0.5/u_3 + 1/u_4 + 0.5/u_5\} \\ A_5 &= \{0.5/u_4 + 1/u_5 + 0.5/u_6\} \\ A_6 &= \{0.5/u_5 + 1/u_6\}, \end{aligned}$$

где $u_i \subset U$ – элементы универсума U , а число, стоящее в числителе каждого элемента нечеткого множества, представляет собой степень принадлежности $\mu(u_i)$ этого элемента нечеткому множеству $A_j, j = \overline{1,6}$.

Шаг 4: Фаззификация приращений, полученных на шаге 1.

Считаем, что если приращение года t есть $p \in u_i$, и существует лингвистическое значение (нечеткое множество A_j) с максимальной степенью принадлежности, приходящейся на элемент u_i , тогда p фаззифицируется как A_j . Например, приращение за март 2002 по сравнению с предыдущим месяцем составляет +0.2476 – это значение попадает в интервал u_5 , и фаззифицированное приращение становится равным A_5 . Аналогичным образом производятся попарные сравнения каждого последующего и предыдущего месяцев, приводящие к формированию последовательности A_6 (февраль 2002), A_5

(март 2002), A_5 (апрель 2002), A_4 (май 2002), A_5 (июнь 2002), A_5 (июль 2002), A_4 (август 2002) и т. д.

Шаг 5: Формирование логических отношений $A_i \rightarrow A_j$.

Для построения последовательности логических отношений, рассматриваем попарно последовательные фаззифицированные приращения (февраль – март, март – апрель, и т. д.), определенные на шаге 4. Исключая повторяющиеся комбинации, окончательный список отношений принимает вид

$$\begin{aligned} &A_1 \rightarrow A_2, A_1 \rightarrow A_4 \\ &A_2 \rightarrow A_1, A_2 \rightarrow A_2, A_2 \rightarrow A_3, A_2 \rightarrow A_4, A_2 \rightarrow A_5 \\ &A_3 \rightarrow A_2, A_3 \rightarrow A_3, A_3 \rightarrow A_4 \\ &A_4 \rightarrow A_2, A_4 \rightarrow A_3, A_4 \rightarrow A_4, A_4 \rightarrow A_5 \\ &A_5 \rightarrow A_3, A_5 \rightarrow A_4, A_5 \rightarrow A_5, A_5 \rightarrow A_6 \\ &A_6 \rightarrow A_5, A_6 \rightarrow A_4. \end{aligned}$$

Мы предполагаем, что нечеткое импликативное отношение $D = B \rightarrow C$ для произвольных векторов B и C интерпретируется как нечеткая импликация Мамдани, следовательно, элементы матрицы D вычисляются по формуле $d_{ij} = b_i^T \times c_j = \min(b_i, c_j)$, где b_i и c_j – элементы векторов B и C , соответственно.

Шаг 6: Объединение логических отношений (шаг 5), имеющих одинаковые левые части, в группы, и вычисление отношений R_i , $i = \overline{1,6}$ для каждой сформированной группы. Можно обратить внимание на то, что группы отношений уже практически построены (см. шаг 5), и выглядят они следующим образом:

$$\begin{aligned} &A_1 \rightarrow A_2, A_4 \\ &A_2 \rightarrow A_1, A_2, A_3, A_4, A_5 \\ &A_3 \rightarrow A_2, A_3, A_4 \\ &A_4 \rightarrow A_2, A_3, A_4, A_5 \\ &A_5 \rightarrow A_3, A_4, A_5, A_6 \\ &A_6 \rightarrow A_5, A_4. \end{aligned}$$

Результирующие отношения $R_i, i = \overline{1,6}$, представляют собой объединения логических отношений, попавших в i -ю группу:

$$R_1 = A_1^T \times A_2 \cup A_1^T \times A_4$$

$$R_2 = A_2^T \times A_1 \cup A_2^T \times A_2 \cup A_2^T \times A_3 \cup A_2^T \times A_4 \cup A_2^T \times A_5$$

$$R_3 = A_3^T \times A_2 \cup A_3^T \times A_3 \cup A_3^T \times A_4$$

$$R_4 = A_4^T \times A_2 \cup A_4^T \times A_3 \cup A_4^T \times A_4 \cup A_4^T \times A_5$$

$$R_5 = A_5^T \times A_3 \cup A_5^T \times A_4 \cup A_5^T \times A_5 \cup A_5^T \times A_6$$

$$R_6 = A_6^T \times A_4 \cup A_6^T \times A_5.$$

Шаг 7: Прогнозирование и дефаззификация получаемых результатов.

Вычисленные отношения R_i используются в модели прогнозирования

$$A_i = A_{i-1} \circ R_i,$$

где A_i – нечеткое множество, выражающее прогнозное приращение месяца i ;

A_{i-1} – известное приращение предшествующего $(i-1)$ -го месяца (если $A_{i-1} = A_j$, то $R_i = R_j$, $j = \overline{1,6}$);

\circ – обозначает композиционный «max-min» оператор.

Например, приращение валютного курса за февраль 2004 года при известном приращении (-0.5714) за январь месяц того же года вычисляется по формуле $F(02.2004) = A_1 \circ R_1$, где R_1 имеет вид, показанный в первой строке, а A_1 – фаззифицированное приращение января 2004 года.

Шаг 8: Вычисление прогнозных валютных котировок USD/RUB.

Этот этап предусматривает преобразование полученных на шаге 7 нечетких прогнозных приращений в целые числа. В значительной степени такой процесс зависит от особенностей рассматриваемой

задачи, и одним из критериев выбора процедуры дефаззификации является ее вычислительная простота.

После того как получено числовое приращение для рассматриваемого месяца, оно суммируется с уже имеющимся значением обменного курса предыдущего месяца. Рассмотренный метод нечеткого моделирования может быть отнесен к числу полуавтоматических процедур, поскольку большинство выполняемых шагов, включая построение универсума на основании множества исходных данных задачи, могут быть эффективно воплощены в программной форме. Однако участие аналитика (эксперта) при формировании интервалов разбиения и соответствующих нечетких множеств играет также огромную роль.

Извлечение знаний из временных рядов

Исследования временных рядов, в том числе нечетких, в последние десятилетия оформились в виде отдельного направления, называемого *интеллектуальным анализом данных*, или *DataMining*, в котором анализ временных рядов получил название *интеллектуального анализа временных рядов*, или *TimeSeries DataMining (TSDM)*.

Человеческое знание основано на образах и формулируется лингвистически. Вычисления со словами и образами дают методологию для управления информацией и для разработки систем, основанных на знаниях. *DataMining* интегрирует методы, основанные на образах, дает возможность для извлечения информации из баз данных в лингвистической форме, подходящей для их использования в методах принятия решений. Методы и технологии извлечения знаний с использованием временных рядов должны оперировать паттернами временных рядов, отыскивать ассоциации между ними и извлекать знания (рисунок 22). То есть необходимо представление результатов *DataMining* в форме, используемой в человеческих знаниях.

На основе новой методологии *DataMining* решается расширенная совокупность задач анализа временных рядов, определенных в работе [22]:

- 1) сегментация – разбиение ВР на значимые сегменты;
- 2) кластеризация – поиск группировок ВР или их паттернов;
- 3) классификация – назначение ВР или их паттернам одного из заранее определенных классов;
- 4) индексирование – построение индексов для эффективного выполнения запросов к базам данных ВР;
- 5) резюмирование (summarization) – формирование краткого описания ВР, содержащего существенные черты с точки зрения решаемой задачи;
- 6) обнаружение аномалий – поиск новых, нетипичных паттернов ВР;
- 7) частотный анализ – поиск часто проявляющихся паттернов ВР;
- 8) прогнозирование – получение очередного значения ВР на основе истории ВР;
- 9) извлечение ассоциативных правил – поиск правил, относящихся к паттернам ВР.

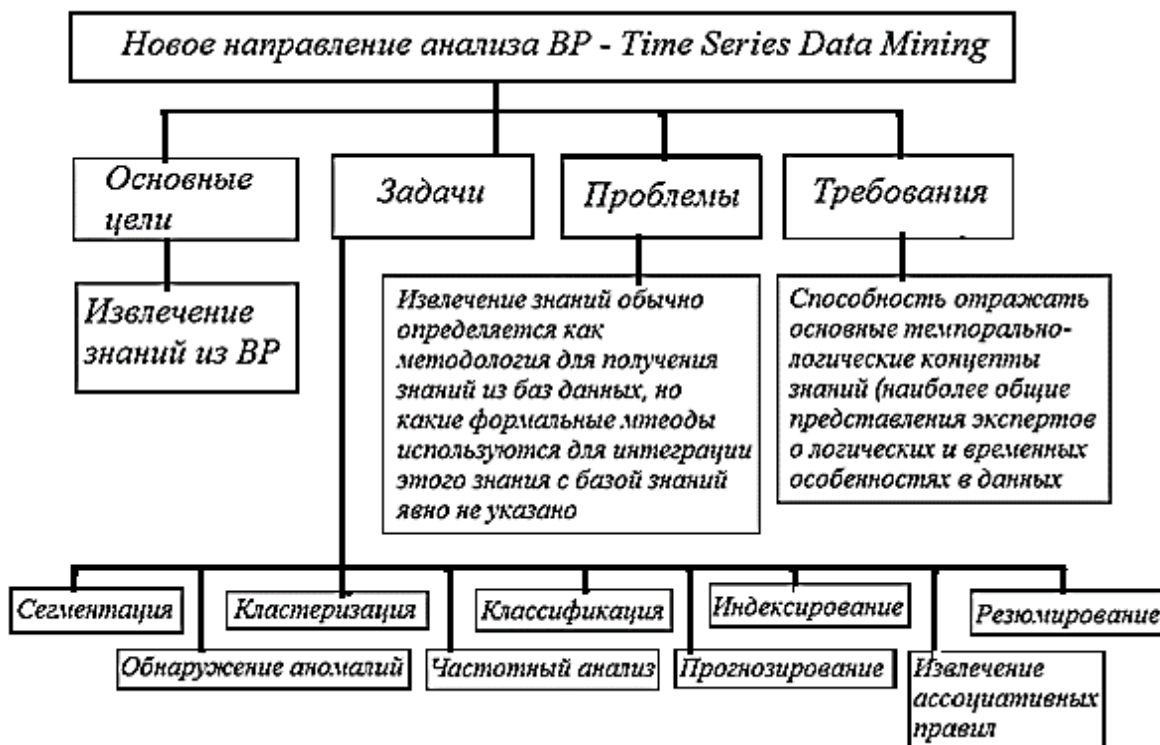


Рисунок 22. Задачи DataMining временных рядов

Традиционное выделение паттернов ВР было связано с выделением участков с постоянным знаком первой и второй производной: возрастающий и выпуклый, убывающий и гладкий и др. Различные шкалы и методы нечетких вычислений Л.Заде использовались для описания паттернов линейных трендов: рост, падение, резкий рост, медленное падение и т. д. Параметрические методы выпукло-гладкой модификации линейных функций и нечеткая грануляция выпукло-гладких паттернов позволили получить лингвистическое описание для ВР, подобное следующему: медленно убывающий и строго гладкий.

В рамках описанного выше направления TSDM акцент делается на поиск и извлечение правил из ВР, при этом полагаются на следующие основные принципы:

- Поиск правил нацелен на получение понимаемых результатов и не обязательно самых точных прогнозов;
- Важнейшим шагом на пути извлечения интерпретируемых знаний является порождение описаний фрагментов ВР в форме темпоральных образов, допускающих естественно-языковое толкование.

Требование к модели представления – способность отражать основные темпорально-логические концепты знаний (наиболее общие представления экспертов о логических и временных особенностях в данных).

Виды темпорально-логических концептов следующие:

- концепт временной продолжительности – присутствие определенного паттерна или признака ВР на определенном интервале времени;
- концепт очередности – порядок следования паттернов ВР во времени;
- концепт одновременности – совпадение во времени темпоральных событий (паттернов различных ВР);

- концепт нечеткости – нечеткость выраженности темпоральных событий и отношений.

В качестве инструментов анализа предлагается использовать нечеткие нейронные сети.

Существует множество методов прогнозирования временного ряда. Однако не создан лучший в любой ситуации метод: каждый из них имеет свои достоинства и недостатки. Поэтому одним из наиболее перспективных научных направлений стало создание комбинированных моделей.

Комбинированная модель прогнозирования – модель прогнозирования, состоящая из нескольких индивидуальных (частных) моделей, называемых базовым набором моделей. Использование комбинированной модели может повысить точность получаемого прогноза по ряду причин:

- Попытка выбрать одну модель для временного ряда с изменяющимся уровнем и динамическими свойствами приводит к выбору усредненной модели;
- При быстром изменении уровней ряда и его динамических свойств невозможно быстро производить анализ динамики и заменять одну модель прогнозирования другой;
- Любой отвергнутый из-за неоптимальности прогноз почти всегда содержит полезную независимую информацию;
- Слабые стороны одного метода прогнозирования возможно преодолеть, используя преимущества другого метода.

На кафедре «Информационные системы» Ульяновского государственного технического университета была разработана информационная система «Combination of fuzzy and exponential models». Данная система позволяет получать прогноз для временного ряда путем агрегации индивидуальных прогнозов моделей из базового набора. На момент создания системы базовый набор содержал 29 индивидуальных моделей и их модификаций, относящихся к экспоненциальным и нечетким моделям прогнозирования временных

рядов. Разработанная информационная система одержала победу на конкурсе «Computational Intelligence in Forecasting» в 2015 году [23].

Точность агрегированного прогноза напрямую зависит от выбранных из базового набора моделей. В связи с этим можно утверждать, что информационная система «Combination of fuzzy and exponential models» имеет два существенных недостатка:

- Пользователь вынужден самостоятельно выбирать модели из базового набора для каждого прогнозируемого временного ряда;
- Для всех временных рядов используется один агрегирующий метод.

Таким образом, целесообразно использовать методы машинного обучения для выбора моделей из базового набора и для выбора агрегирующего метода. Указанные задачи можно решить с помощью нейронной сети, выбирающей методы на основании значений метрик прогнозируемого временного ряда.

Для выбора метрик необходимо выбрать совокупность значимых для прогнозирования временного ряда характеристик, таких как: длина, степень выраженности тренда, степень выраженности сезонности, величина дисперсии, наличие стационарности.

При выборе используемых метрик временного ряда необходимо выполнение следующих условий:

- Метрика соответствует одной из основных характеристик временного ряда;
- Совокупность метрик максимально и разносторонне описывает временной ряд;
- Значение метрики принадлежит интервалу от 0 до 1.

При выборе методов прогнозирования из базового набора обучающая выборка нейронной сети представляет собой набор временных рядов, представленных в виде совокупности метрик, и соответствующих значений ошибки прогнозирования для каждой модели. Нейронная сеть, таким образом, обучается на основании значений метрик вычислять предполагаемые значения ошибки для каждой

модели из базового набора, что позволяет определять, какие модели из базового набора эффективнее применить для прогнозирования значений текущего временного ряда.

Для создания комбинированной модели прогнозирования на основании использования методов машинного обучения необходимо решить ряд задач:

- Определить состав моделей, входящих в базовый набор;
- Выявить ключевые для задачи прогнозирования метрики временного ряда;
- Определить структуру и способ обучения нейронных сетей выбора методов прогнозирования из базового набора и выбора метода агрегации.

Нечеткое сглаживание временного ряда

Очень часто при анализе временных рядов переходят от исследования самих значений ряда к исследованию тренда. Одним из методов его выделения является метод F-преобразования.

Нечеткое сглаживание временных рядов на основе нечеткого преобразования (F-преобразования) – методика, разработанная Перфильевой [24], которая может быть отнесена к методикам нечеткого приближения.

F-преобразование предполагает задание нечеткого разбиения универсального множества. В качестве последнего выбирается конечный интервал $[a, b]$ действительной прямой. Зафиксируем значение n ($n > 2$) узлов x_1, \dots, x_n на $[a, b]$ и предположим, что $x_1 < \dots < x_n$, причем $a = x_1, b = x_n$.

Под нечетким разбиением $[a, b]$ будем понимать совокупность n функций $A_1, \dots, A_n : [a, b] \rightarrow [0, 1]$, удовлетворяющих следующим свойствам:

- $A_k : [a, b] \rightarrow [0, 1], A_k(x_k) = 1$;

- $A_k(x) = 0$ если $x \notin (x_{k-1}, x_{k+1})$, где для единообразия обозначения мы положим $x_0 = a, x_{n+1} = b$;
- $A_k(x)$ непрерывна;
- $A_k(x), k = 2, \dots, n$ строго возрастает на $[x_{k-1}, x_k]$ и строго убывает на $[x_k, x_{k+1}]$;
- $\sum A_k(x) = 1$ для всех $x \in [a, b]$.

Функции A_1, \dots, A_n называются базисными функциями. Базисные функции A_1, \dots, A_n могут служить также функциями принадлежности нечетких подмножеств A_1, \dots, A_n (обозначения функций и множеств унифицированы). Отметим, что форма базисных функций может быть уточнена дополнительно и согласована с такими требованиями к модели, как, например, гладкость.

Следующие формулы представляют нечеткое разбиение отрезка $[x_1, x_n]$, полученное совокупностью функций:

$$A_1(x) = \begin{cases} 1 - \frac{(x - x_1)}{h_1}, & x \in [x_1, x_2] \\ 0, & \text{иначе} \end{cases}$$

$$A_k(x) = \begin{cases} \frac{(x - x_{k-1})}{h_{k-1}}, & x \in [x_{k-1}, x_k], \\ 1 - \frac{(x - x_k)}{h_k}, & x \in [x_k, x_{k+1}], \\ 0, & \text{иначе} \end{cases}$$

$$A_{n-1}(x) = \begin{cases} 1 - \frac{(x - x_{n-1})}{h_{n-1}}, & x \in [x_{n-1}, x_n] \\ 0, & \text{иначе} \end{cases}$$

где $k = 1, \dots, n-1$ и $h_k = x_{k+1} - x_k$.

Предположим, что функция f имеет своей областью определения множество $P = \{p_1, \dots, p_l\} \subset [a, b]$, где $l > n$. Множество P считается плотным относительно нечеткого разбиения A_1, \dots, A_n , если выполнено условие:

$$(\forall k)(\forall j) A_k(p_j) > 0 \quad .$$

Пусть $A_k(p_j) = a_{kj}, k = 1, \dots, n; j = 1, \dots, l$, тогда матрица $A_{n \times l} = a_{kj}$ называется матрицей нечеткого разбиения для P , для которой справедливы свойства:

$$(\forall k)(\forall j) a_{kj} \in [0, 1]$$

$$(\forall j) \sum_{k=1}^n a_{kj} = 1.$$

F-преобразованием вектора f , определяемым матрицей нечеткого разбиения A , назовем вектор $F_n[f]$, где $F_n[f] = (F_1 \dots F_n)$

$$\text{и } F_i = \frac{\sum_{j=1}^l a_{ij} f_j}{\sum_{i=1}^l a_{ij}}.$$

Координаты вектора $F_n[f]$ назовем соответственно компонентами F-преобразования. Обозначим $a_i = \sum_{j=1}^l a_{ij}, i = 1, \dots, n$; тогда $(a_1 F_1, \dots, a_n F_n)^T = A \times f$. Компоненты F-преобразования являются точками минимума функции, задающей критерий взвешенного среднеквадратичного отклонения.

Пусть $F_n[f]$ есть F-преобразование f , определяемое $A_{n \times l} = a_{kj}$. Обратным F-преобразованием $F_n[f]$ назовем вектор $f_{F,n}$, вычисляемый по формуле $f_{F,n}^T = F_n[f] \times A$. Можно доказать, что если n возрастает, тогда $f_{F,n}(p_j)$ сходится $f(p_j), j = 1, \dots, N$.

F-преобразование имеет (кроме прочих) следующие свойства, важные для использования в качестве сглаживания временных рядов:

- У него прекрасные фильтрующие свойства;
- Его легко вычислять
- F-преобразование стабильно относительно выбора точек p_1, \dots, p_N . Это означает, что при выборе других точек p_k (и, возможно, изменяя их число N), результирующая функция $f_{F,n}$ значительно не меняется.

Прогнозирование тренда и прогнозирование числового представления временного ряда производится отдельно. Для этого необходимо вычислить так называемые остатки – разность между временным рядом и его трендом:

$$R = \{f(p_j) - F_k\}, \text{ где } j : A_k(j) > 0.$$

Полученный вектор R -вектор остатков для k -й компоненты тренда. Прогноз тренда и векторов остатков реализуется по формуле линейной комбинации. Прогноз компоненты тренда:

$$F_{k+1} = \alpha F_k + \beta F_{k-1} \dots$$

и прогноз вектора остатков:

$$R_{k+1} = \alpha R_k + \beta R_{k-1} \dots$$

Для получения прогноза находится решение системы уравнений:

$$\begin{cases} F_{k-1} = \alpha F_{k-2} + \beta F_{k-3} \dots \\ F_k = \alpha F_{k-1} + \beta F_{k-2} \dots \end{cases}$$

Аналогично строится прогноз вектора остатков.

Также для построения прогноза может быть использована нейронная сеть, например, многослойный перцептрон или сеть Кохонена с выходной звездой Гроссберга. Вид нейронной сети определяет вид входных данных: на вход многослойного перцептрона подаются только абсолютные значения (F_k, F_{k-1}, \dots) , на вход сети Кохонена подаются значения точек тренда, вычисленные относительно друг друга $(F_k - F_{k-1}, F_{k-1} - F_{k-2}, \dots)$.

На качество прогноза влияет количество нейронов во внутреннем слое. После получения прогнозных значений тренда и вектора остатков возможно построение числового представления прогноза временного ряда. Для этого производится сложение прогнозной компоненты тренда и прогнозного вектора остатков.

Точность прогноза зависит от количества точек временного ряда, участвующих в обучении: чем больше их, тем меньше ошибка (рисунок 23). Как видно из рисунка, вывод справедлив и для MAPE, и для SMAPE (линии на графике практически совпадают).

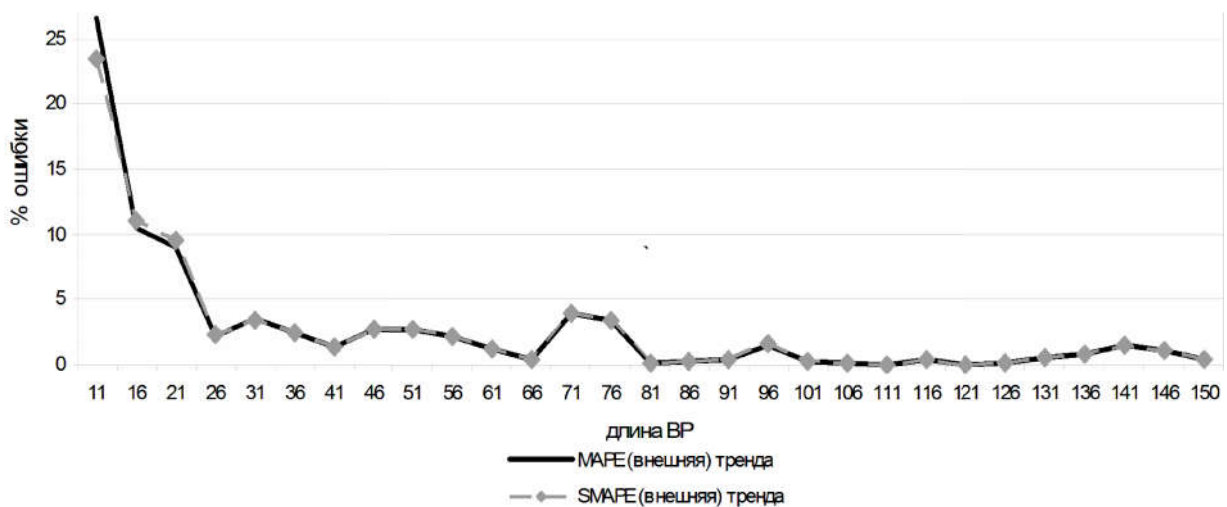


Рисунок 23. Зависимость процента ошибок от длины обучающей выборки

Процесс получения прогноза тренда и остатков зависит от некоторых параметров, которые в модели не удастся задать жестко. Выбор данных параметров в значительной степени влияет на качество прогноза. Определим их:

- Степень авторегрессии при построении прогноза тренда (область значений);
- Метод, которым производится прогноз:
 - решение системы линейных уравнений;
 - нейронная сеть с абсолютными значениями предыдущих компонент на входе;
 - нейронная сеть с разностями между предыдущими компонентами на входе;

- нейронная сеть с абсолютными значениями предыдущих компонент, а также разности между ними на входе;
- количество точек, покрываемых базисной функцией;
- сезонность (история отстоит на k точек от прогноза).

Сочетание параметров, при котором получается наилучший прогноз, получается перебором.

Нечеткая регрессия

В области прикладной статистики, анализа временных рядов и принятия решений в условиях неопределенности накоплен богатый опыт исследований и существует множество моделей, начиная от простейших линейных регрессионных моделей поиска тренда временного ряда и заканчивая сложными многоуровневыми авторегрессионными и адаптационными моделями. Регрессионный анализ, основанный на методе наименьших квадратов (Least-square), является очень удобным методом построения моделей, позволяющим численно оценивать зависимость интересующего исследователя параметра от воздействующих на него факторов. При анализе зависимости нечетких оценок от воздействующих факторов зачастую исследователям приходится иметь дело с важной информацией, которая не может быть задана точно. Некоторые наблюдения могут быть описаны только лингвистическими выражениями (типа «удовлетворительный», «хороший» и «превосходный»). Для таких данных аппаратом формализации может служить теория нечетких множеств. Были разработаны различные нечеткие регрессионные модели, основой которых является модель нечеткой линейной регрессии.

В нечеткой регрессионной модели параметры представляются треугольными нечеткими числами и являются коэффициентами в нечеткой линейной функции. Неопределенность (vagueness) системы представляется суммарным разбросом («шириной») параметров (нечетких коэффициентов).

Построение модели состоит в нахождении оптимальных в некотором смысле коэффициентов с учетом нечеткой информации об объекте и субъективных представлений исследователя.

Базовые предположения нечеткой регрессии заключаются в том, что остатки, полученные как разность между наблюдениями и их оценками, продуцируются не случайными ошибками измерения, а неопределенностями (типа нечеткость) при вычислении параметров модели.

Большинство работ, посвященных нечеткой регрессии, были основаны на следующих базовых определениях.

Пусть дано множество наблюдений: $(y_j, x_{j1}, \dots, x_{jn}), j = 1, \dots, m$, необходимо найти нечеткую модель по следующей форме:

$$\tilde{Y} = A_0 \tilde{ } + A_1 \tilde{ } x_1 + \dots + A_n \tilde{ } x_n,$$

где $A_i(a_i^c, s_i^L, s_i^R), i = 1, \dots, n$ – триангулярные нечеткие числа;

a_i^c – среднее значение $A_i \tilde{ }$;

s_i^L, s_i^R – показывают левый и правый разброс треугольной функции принадлежности соответственно.

Используются два критерия определения нечетких коэффициентов модели:

1. Для всех наблюдений принадлежность значения y_j к его нечеткой оценке $Y_j \tilde{ }$ должна быть как минимум $Y_j \tilde{ } (y_j) \geq h, j = 1, \dots, m$, где h – уровень доверия, выбранный лицом, принимающим решения.
2. Общая нечеткость предсказываемого значения зависимой переменной должна быть минимизирована. Это может быть достигнуто минимизацией суммы разбросов нечетких чисел для всех наборов данных.

Итак, проблему настройки нечеткой модели с заданными данными $(y_j, x_{j1}, \dots, x_{jn}), j = 1, \dots, m$ можно решить как эквивалентную задачу линейного программирования. То есть найти $a_-^c = (a_0^c, \dots, a_n^c)$, $s_-^L = (s_0^L, \dots, s_n^L)$, $s_-^R = (s_0^R, \dots, s_n^R)$, которые минимизируют выражение:

$$Z = m(s_0^L + s_0^R)s_0^L + \sum_{I=1}^n \left[(s_i^L + s_i^R) \sum_{j=1}^m |x_{ij}| \right].$$

Чтобы оценить качество настройки нечеткой регрессии, используют метод наименьших квадратов. Для нечеткой регрессии среднеквадратичное отклонение (MSE) определяется следующим образом:

$$MSE = \frac{1}{m} \sum_{j=1}^n [y_j - def(Y_j)]^2,$$

где $def(Y_j)$ – дефазифицированное значение зависимой переменной.

Исследователями были выделены различные варианты методов на основе классификации «вход – выход»: «четкий вход – четкий выход» метод CICO (Crisp-Input and CrispOutput), «нечеткий вход – нечеткий выход» метод FIFO (Fuzzy-Inputs and Fuzzy-Outputs) и смешанные данные – метод CIFO (Crisp-Inputs and Fuzzy-Outputs).

ACL-шкала и нечеткая кластеризация объектов

Для задания отношений между элементами нечеткого временного ряда мы можем использовать специальную лингвистическую шкалу в качестве инструмента как абсолютного, так и сравнительного нечеткого оценивания – ACL-шкала (Absolute&Comparative Linguistic) [25]. Абсолютные оценки, полученные по ACL-шкале, будут соответствовать нечетким меткам уровней ВР, а сравнительные оценки – нечетким тенденциям НВР. Опишем ACL-шкалу [25] как алгебраическую систему вида

$$Sx = \langle Name_Sx, \tilde{X}, N, X, G, P, TTend, RTend \rangle,$$

где $Name_Sx$ – имя ACL-шкалы; \tilde{X} – базовое терм-множество абсолютных лингвистических оценок (лингвистическое название градаций); N – мощность базового терм-множества шкалы; X – универсальное множество, на котором определена шкала; G – синтаксические правила вывода (порождения) цепочек оценочных высказываний (производные термов, не входящих в базовое терм-множество); P – семантические правила, определяющие функции принадлежности для каждого терма (задаются обычно экспертно); $TTend(\tilde{x}_i, \tilde{x}_j)$ – лингвистическое отношение, фиксирующее тип изменения между двумя оценками \tilde{x}_i, \tilde{x}_j шкалы; $RTend(\tilde{x}_i, \tilde{x}_j)$ – лингвистическое отношение, фиксирующее интенсивность различия между двумя оценками \tilde{x}_i, \tilde{x}_j шкалы.

Первое, что нужно сделать для построения шкалы, – определить базовое терм-множество. Например, определим множество мощностью 5, содержащее в себе оценки уровней тренда исходного временного ряда: {малое, ниже среднего, среднее, выше среднего, большое}. Мощность множества обуславливается необходимостью вербализации каждого понятия шкалы, таким образом, чтобы она была понятна для человека и семантически содержательна.

Далее необходимо выбрать способ построения ACL-шкалы. Этим способом может стать неравномерное разбиение зафиксированной области значения величины с помощью алгоритма кластеризации. Достоинством такого подхода является то, что разбиение происходит без участия эксперта (тогда как в остальных возможных способах обязательно участие эксперта).

При безэкспертном разбиении задача формулируется следующим образом: дано множество из L точек, необходимо разбить его на k групп. Здесь L – количество элементов ВР (зафиксированное множество значений), k – количество термов на ACL-шкале, в нашем случае равное N . Как мы видим по формулировке задачи, это задача кластеризации. Кластеризация – это процесс разбиения объектов на группы по степени их схожести между собой. В отличие от

классификации, где есть заданная структура групп и признаки, на основе которых объекты в эти группы помещаются, в кластеризации структура разбиения, а также характеристические признаки являются не входными, а выходными данными. Кластеризация может быть четкой и нечеткой. При четкой кластеризации предполагается, что каждый объект принадлежит только одному кластеру, при нечеткой объект принадлежит всем кластерам, но с разной степенью принадлежности. При нечеткой кластеризации мы сможем использовать собственно степени принадлежности объектов ВР кластерам для получения термов шкалы. В итоге вся зафиксированная область разобьется на N взаимопересекающихся кластеров, упорядоченных по возрастанию значений их центров. Взаимопересечение получается за счет нечеткости кластеризации.

Кластеризация может выполняться с помощью нейронных сетей, генетических алгоритмов или с помощью собственно алгоритмов кластеризации. Например, для нечеткого разбиения некоего множества точек на заданное количество кластеров можно использовать fcm-алгоритм кластеризации. Рассмотрим его подробнее.

FCM-алгоритм (Fuzzy Classifier Means) кластеризации применяется для нечеткого разбиения некоего множества объектов на заданное количество кластеров. Целью алгоритма кластеризации является автоматическое разбиение множества объектов, которые задаются векторами признаков в пространстве признаков. Этот алгоритм предполагает, что объекты принадлежат всем кластерам с определенной степенью принадлежности, которая определяется расстоянием от объекта до соответствующих кластерных центров. Данный алгоритм итерационно вычисляет центры кластеров и новые степени принадлежности объектов. При этом он основан на минимизации целевой функции по мере расстояния объектов от центров кластеров:

$$J = \sum_{i=1}^N \sum_{j=1}^C (u_{ij})^m \|x_i - c_j\| ,$$

где N – количество объектов; C – количество кластеров; u_{ij} – степень принадлежности объекта i кластеру j ; m – любое действительное число, большее 1; x_i – i -й объект набора объектов; c_j – j -й кластер набора кластеров; $\|x_i - c_j\|$ – норма, характеризующая расстояние от центра кластера j до объекта i .

Объектами кластеризации при анализе текстов, например, могут являться термины. Вектор признаков объектов кластеризации в данном случае содержит только значение какой-либо статистической метрики (например, частоты термина в тексте).

Алгоритм нечеткой кластеризации выполняется по шагам. Рассмотрим каждый из шагов подробнее, приведя необходимые модификации целевой функции для упрощения дальнейшего программирования.

Первый шаг – инициализация. На этом шаге задаются параметры кластеризации и инициализируется первоначальная матрица принадлежности объектов кластерам. К параметрам относятся следующие величины. Во-первых, степень нечеткости кластеризации – параметр m . От выбора этого параметра зависит значение функции принадлежности точки кластеру. Обычно он выбирается в пределах $\sim 1.5..2$. Чем больше его значение, тем более «нечеткая» кластеризация. В приведенных ниже примерах кода эмпирически было выбрано значение $= 1.6$. Во-вторых, выбирается мера расстояний $\|x_i - c_j\|$. Она характеризует степень близости точки к центру кластера. Если вектор характеристик состоит только из одного значения и задача, по сути, сводится к выделению значимых интервалов на прямой, то мера расстояния может иметь вид

$$\|x_i - c_j\| = |x_i - c_j|.$$

Эта мера характеризует удаленность точки от центра кластера на прямой.

Третий параметр, который выбирается при инициализации – параметр сходимости алгоритма ε (уровень точности). Когда разность значений целевых функций текущей и предыдущей итераций достиг-

нет этого уровня, считается, что кластеризация завершена. Обычно этот параметр равен 0.001.

Четвертый параметр задается во избежание зависания алгоритма при невозможности достижения уровня точности. Это количество итераций алгоритма. Например, 10 000.

После выбора параметров генерируется случайным образом первоначальная матрица принадлежности объектов кластерам и происходит переход ко второму шагу алгоритма.

На шаге 2 происходит вычисление центров кластеров следующим образом:

$$c_j = \frac{\sum_{i=1}^N u_{ij}^{1.6} \cdot x_i}{\sum_{i=1}^N u_{ij}^{1.6}},$$

где c_j – центр j -го кластера; N – количество объектов; x_i – значение i -го объекта; u_{ij} – степень принадлежности объекта i кластеру j .

На третьем шаге формируется новая матрица принадлежности с учетом вычисленных на предыдущем шаге центров кластеров:

$$u_{ij} = \frac{1}{\sum_{l=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_l\|} \right)^{3.33}},$$

где u_{ij} – степень принадлежности объекта i кластеру j ; c – количество кластеров; c_j – вектор центра j -го кластера; c_l – вектор центра l -го кластера.

При этом если для некоторого кластера j и некоторого объекта i расстояние $\|x_i - c_j\| = 0$, тогда полагаем, что степень принадлежности u_{ij} равна 1, а для всех остальных кластеров степень принадлежности этого объекта равна нулю.

Далее на четвертом шаге вычисляется значение целевой функции, и полученное значение сравнивается со значением на преды-

душей итерации. Если разность не превышает заданного в параметрах кластеризации значения ε , считаем, что кластеризация завершена. В противном случае переходим ко второму шагу алгоритма.

Мы рассмотрели методы нечеткой логики, теперь перейдем к следующей модели – искусственных нейронных сетей (ИНС).

Искусственные нейронные сети

Для подготовки данного раздела использовался материал (в том числе иллюстративный) из следующего источника: [26].

Особенности нейронных сетей

Прежде чем перейти к описанию того, что собой представляет модель искусственной нейронной сети, акцентируем внимание на особенностях данного подхода, а в частности его достоинствах и недостатках. Это основные моменты, которые обязательно необходимо понять и запомнить.

Достоинства ИНС:

1. Нелинейность модели, что дает возможность аппроксимировать любые нелинейные функции. Та же задача с помощью полиномиальной модели не всегда решается, либо же ее решение становится низким по качеству.
2. Локальности восприятия, заключающаяся в том, что каждый нейрон получает не весь входной вектор. Это позволяет ей сегментировать данные и работать с более сложными ситуациями.
3. Каскад слоев. В сочетании с пунктом 2 это дает способность воспринимать более абстрактные признаки.
4. Лучше работает с мультиколлинеарностью и комбинациями признаков.
5. Нейронная сеть дает несколько механизмов для контроля переобучения.
6. Нейронная сеть способна дообучаться при непротиворечивости новых образов.

Недостатки:

Нейронная сеть не интерпретируема. Поэтому не ясна логика преобразования входных данных в выходные. Не всегда можно убедиться в стабильности решения на всей области определения.

Математика процессов и некоторых аспектов функционирования еще недостаточно изучена. Часто количество нейронов и слоев приходится подбирать экспериментально для конкретной задачи, потому что определенной методики нет.

Нет аналитического решения, а решение численными методами градиентного спуска может приводить к попаданию в локальные минимумы ошибки.

Для ИНС очень часто требуется большая выборка. Так для распознавания речи одного языка может потребоваться от 5 до 10 тыс. часов размеченных записей. Для Глубоких сетей необходимо еще большее число объектов обучающей выборки, иначе будет постоянно происходить переобучение.

Обучение большой ИНС требует больших вычислительных ресурсов и специального оборудования GPU.

Поскольку на сегодняшний день нейронные сети очень модная тема, следует отметить, что ИНС, так же как любые обобщающие модели машинного обучения, чувствительна к противоречиям в данных и, конечно же, на текущем этапе развития ИС не могут содержать противоречивые «представления» об объекте, проводить какие-то внутренние размышления и перестановки. Иными словами, из неоткуда информацию ИНС не возьмет (самостоятельно не проделает ряд выводов).

Так одним из главных конкурентов ИНС в ряде задач являются решающие деревья и ансамбли над решающими деревьями (например, XGBoost). Исключением являются задачи машинного зрения и работы с изображениями вообще, в которых ИНС занимают сегодня лидирующие позиции.

Определение модели искусственной нейронной сети

Согласно Википедии [1]: «Искусственная нейронная сеть – математическая модель, а также ее программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма. Это понятие возникло при изучении процессов, протекающих в мозге, и при попытке смоделировать эти процессы. Первой такой попыткой были нейронные сети, созданные У. МакКаллоком и У. Питтсом. После разработки алгоритмов обучения получаемые модели стали использовать в практических целях: в задачах прогнозирования, для распознавания образов, в задачах управления и др.».

Искусственные нейроны сильно упрощены по сравнению с их биологическими прототипами. Каждый из них имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим нейронам. Способность решать довольно сложные задачи обуславливается тем, что все нейроны соединены в достаточно большую сеть с управляемым взаимодействием. Схему простой нейронной сети можно увидеть на рисунке 24. Литерой «I» обозначены входные нейроны, «S» – скрытые нейроны, «O» – выходной нейрон. Соединяющие линии – связи между нейронами, или синаптические связи. Именно синаптические связи играют важнейшую роль в модели ИНС, так как в основе обучения лежит механизм корректировок силы этих связей.

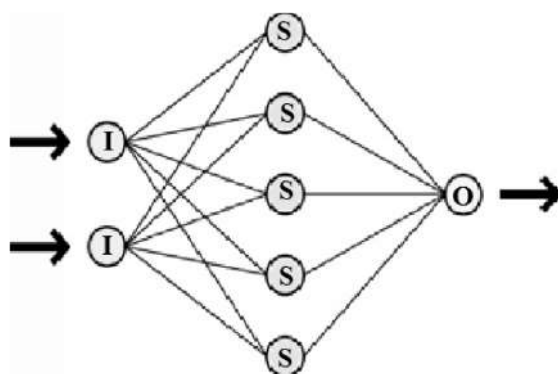


Рисунок 24. Схема простой нейросети

Нейронную сеть можно рассмотреть с разных точек зрения. Например, в машинном обучении, нейронная сеть – частный случай методов распознавания образов, дискриминантного анализа, методов кластеризации и т. п. С точки зрения математики, обученная нейронная сеть – решенная многопараметрическая задача нелинейной оптимизации. Кибернетика представляет нейронную сеть как «черный ящик» или произвольную передаточную функцию, которую можно получить при обучении сети и затем воспроизводить при использовании сети. С точки же зрения искусственного интеллекта, ИНС является основой философского течения коннективизма и основным направлением в структурном подходе по изучению возможности построения (моделирования) естественного интеллекта с помощью компьютерных алгоритмов.

Очень часто именно про нейронные сети говорится, что они не программируются в привычном смысле этого слова, они обучаются и что возможность обучения – одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Конечно же, такие трактовки и объяснения нельзя назвать точными и их можно допускать только при первом знакомстве с машинным обучением вообще.

Термин «обучение» необходимо трактовать лишь математически. В таком случае и Регрессия, и Деревья также обучаются, хотя в любой модели происходит одно и то же – корректировка весов.

Однако шумиха вокруг нейронных сетей порождает массу мифов, принижающих остальные методы, изученные нами ранее.

В чем-то эта модель напоминает уже изученные модели, такие как Регрессия и Деревья решений. В каком-то смысле Нейронные сети являются комбинацией этих методов (на уровне концепции).

В основе математической модели вычислений в ИНС лежит взвешенное суммирование, то есть, по сути, уже знакомое вам полиномиальное уравнение (в котором коэффициенты уже принято называть весами). Веса этого полинома и отражают силу синаптической связи между нейронами. Именно синаптические связи играют важнейшую роль в модели ИНС, так как в основе обучения лежит

механизм корректировок силы этих связей (подобно биологическому прототипу).

Все также вычисления завязаны на понятии ошибки и на том принципе, что корректное обобщение (отображение $X \Rightarrow Y$) строится посредством уменьшения этой ошибки. Таким образом, они являются логическим продолжением прошлых тем.

Однако стоит отметить, что в отличие от методов Байеса и Регрессии нейронные сети изначально разрабатывались исходя из кибернетического подхода к моделям и информации. То есть нейронные модели – это некая адаптивная система, которая может прийти из одного состояния в требуемое, но при этом имеется некая случайная составляющая, которая вносит нестабильность в решения. Также нет хорошей компактной формулы, описывающей работу сети, – с самого начала суть вычислений в нейронных сетях была основана на численном (постепенном\итеративном) подходе к уменьшению ошибки.

Первая формальная модель и первая реализация нейронной сети

Согласно [27]: «Первой формальной моделью нейронных сетей (НС) и нейрона вообще была модель МакКаллока-Питтса, уточненная и развитая Клини. Впервые было установлено, что НС могут выполнять любые логические операции и вообще любые преобразования, реализуемые дискретными устройствами с конечной памятью. Эта модель легла в основу теории логических сетей и конечных автоматов и активно использовалась психологами и нейрофизиологами при моделировании некоторых локальных процессов нервной деятельности. В силу своей дискретности она вполне согласуется с компьютерной парадигмой и, более того, служит ее «нейронным фундаментом».

Пусть имеется n входных величин x_1, \dots, x_n бинарных признаков, описывающих объект x . Значения этих признаков будем трактовать как величины импульсов, поступающих на вход нейрона через n входных синапсов. Будем считать, что, попадая в нейрон, импульсы складываются с весами $\omega_1, \dots, \omega_n$.

Если вес положительный, то соответствующий синапс возбуждающий, если отрицательный, то тормозящий. Если суммарный импульс превышает заданный порог активации ω_0 , то нейрон возбуждается и выдает на выходе 1, иначе выдается 0.

Таким образом, нейрон вычисляет n -арную булеву функцию

$$a(x) = \varphi(\sum_{j=1}^n w_j x^j - w_0),$$

где $\varphi(z) = [z \geq 0]$ – ступенчатая функция Хевисайда.

В теории нейронных сетей функцию φ , преобразующую значение суммарного импульса в выходное значение нейрона, принято называть функцией активации. Таким образом, модель МакКаллока-Питтса эквивалентна пороговому линейному классификатору. Схема ее вычислений показана на рисунке 25.

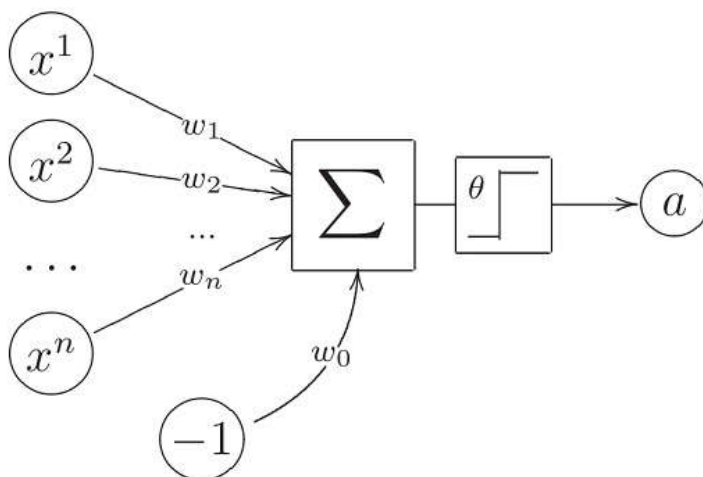


Рисунок 25. Схема вычислений в модели нейрона МакКаллока-Питтса

Получается, что нейрон производит взвешивание входных признаков (то есть рассчитывает результат с учетом их важности\веса) подобно регрессии. Однако важно понимать, что это пока всего лишь один нейрон.

Теоретические основы нейроматематики были заложены в начале 40-х годов, и в 1943 году У. МакКаллок и его ученик У. Питтс сформулировали основные положения теории деятельности головного мозга.

Ими были получены следующие результаты:

- разработана модель нейрона как простейшего процессорного элемента, выполняющего вычисление переходной функции от скалярного произведения вектора входных сигналов и вектора весовых коэффициентов;
- предложена конструкция сети таких элементов для выполнения логических и арифметических операций;
- сделано основополагающее предположение о том, что такая сеть способна обучаться, распознавать образы, обобщать полученную информацию.

Недостатком данной модели является то, что в качестве функции активации используется функция Хевисайда или униполярная пороговая функция (рисунок 26). В формализме, предложенном МакКаллоком и Питтсом, нейроны имеют состояния 0, 1 и пороговую логику перехода из состояния в состояние. Каждый нейрон в сети определяет взвешенную сумму состояний всех других нейронов и сравнивает ее с порогом, чтобы определить свое собственное состояние.

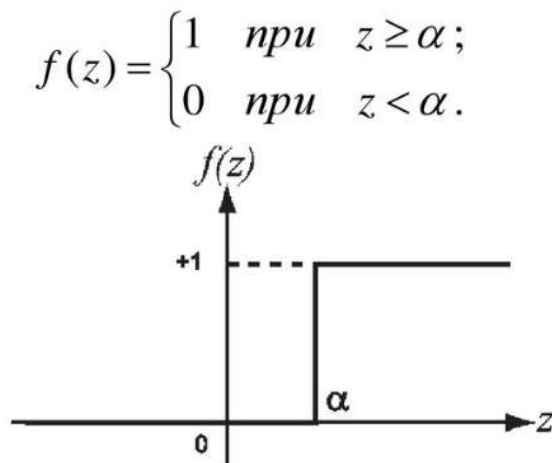


Рисунок 26. Пороговая функция, или функция Хевисайда

Пороговый вид функции не предоставляет нейронной сети достаточную гибкость при обучении и настройке на заданную задачу. Если значение вычисленного скалярного произведения, даже незначительно, не достигает до заданного порога, то выходной сигнал не

формируется вовсе, и нейрон «не срабатывает». Это значит, что теряется интенсивность выходного сигнала (аксона) данного нейрона и, следовательно, формируется невысокое значение уровня на взвешенных входах в следующем слое нейронов.

К тому же модель не учитывает многих особенностей работы реальных нейронов (импульсного характера активности, нелинейности суммирования входной информации, рефрактерности).

Несмотря на то, что за прошедшие годы нейроматематика ушла далеко вперед, многие утверждения МакКаллока остаются актуальными и поныне. В частности, при большом разнообразии моделей нейронов принцип их действия, заложенный МакКаллоком и Питтсом, остается неизменным».

Первой реализацией модели нейронной сети была созданная в 1960 электронная машина «Марк-1». Идею предложил Фрэнк Розенблатт в 1957. Согласно Википедии [1]: «Несмотря на то, что это первая модель и концепция, тем ни менее многие идеи современных нейронных сетей во многом совпадают с концепцией персептрона. Так выше была рассмотрена модель упрощенного искусственного нейрона, способного производить вычисления. Персептрон по своей сути является сетью таких нейронов. Из чего и предполагалось получить механизм для более сложных вычислений.

Несмотря на свою простоту, персептрон способен обучаться и решать довольно сложные задачи. Основная математическая задача, с которой он справляется, – это линейное разделение любых нелинейных множеств, так называемое обеспечение линейной сепарабельности. Логическая схема персептрона с тремя выходами представлена на рисунке 27.

Персептрон состоит из трех типов элементов, а именно: поступающие от датчиков сигналы передаются ассоциативным элементам, а затем реагирующим элементам. Таким образом, персептроны позволяют создать набор «ассоциаций» между входными стимулами и необходимой реакцией на выходе.

В биологическом плане это соответствует преобразованию, например, зрительной информации в физиологический ответ от двигательных нейронов. Согласно современной терминологии, перцептроны могут быть классифицированы как искусственные нейронные сети:

- с одним скрытым слоем;
- с пороговой передаточной функцией;
- с прямым распространением сигнала».

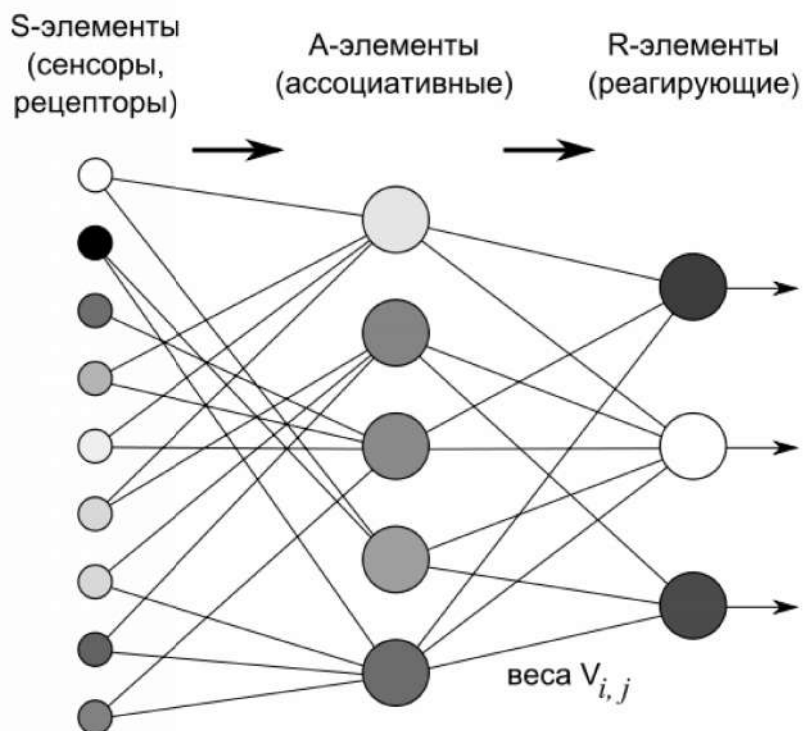


Рисунок 27. Логическая схема перцептрона с тремя выходами

Многослойный перцептрон (MLP)

Данный тип архитектуры является самой распространенной архитектурой искусственных нейронных сетей и в каком-то смысле классической. Очень часто, когда речь идет о нейронных сетях вообще, то имеется в виду именно многослойная сеть прямого распространения. Многие задачи классификации, аппроксимации и управления, которые решаются с помощью нейронных сетей, реша-

ются именно этим типом сети. Кроме того, принципы обучения, разработанные для этого типа сети, впоследствии стали применяться и для других типов. Так что в каком-то смысле это базовая архитектура для других типов сетей.

Если подходить к определению строго, то не совсем корректно называть такой тип сети Персептроном, ведь она отличается от него не только количеством слоев, но и тем, что:

- Нейроны имеют не ступенчатую функцию активации Хевисайда, а сигмоидальную функцию.
- Обучение производится не по правилу Хебба, а с помощью обратного распространения ошибки.

Поэтому такую архитектуру называют либо просто многослойной сетью, либо сетью прямого распространения или многослойной сетью прямого распространения. Но в то же время можно встретить и более удобное название MLP, что значит многослойный персептрон. Такое название определяет, что это все-таки не просто сеть. Структура MLP представлена на рисунке 28.

Замечание: выше уже упоминалось о том, что терминология в машинном обучении и в нейросетевых технологиях в частности еще не до конца систематизирована. Поэтому необходимо разбираться в сути моделей, допуская определенную вариативность названий. Однако со временем это перестает доставлять какие-либо трудности.

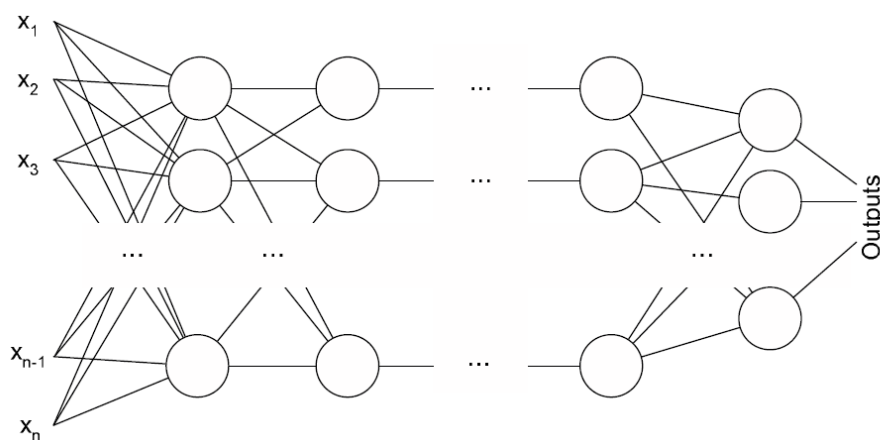


Рисунок 28. Структура многослойного персептрона

На рисунке 29 представлен график сигмоидальной функции активации.

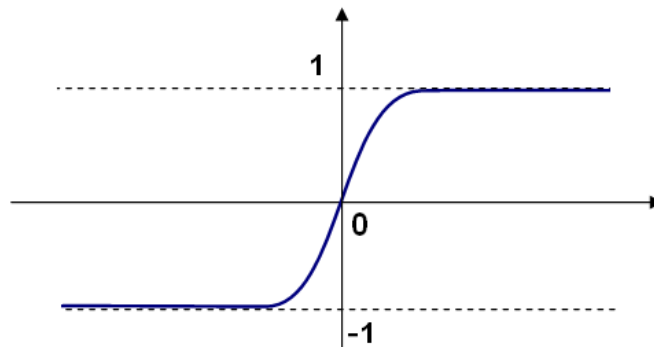


Рисунок 29. График сигмоидальной функции

Сигмоидальная функция активации обладает следующими преимуществами по сравнению с пороговой функцией Хевисайда:

Нелинейность. Обеспечивает модели возможность обрабатывать (воспринимать) нелинейные закономерности в данных.

Фиксированные ограничения выхода. Это особенность позволяет масштабировать данные от слоя к слою.

Непрерывность и Дифференцируемость. Данные свойства позволяют обойти главный недостаток пороговой функции – резкий переход. Поскольку обучение сети происходит посредством градиентного спуска, то движение по пороговой функции слишком резкое, в отличие от сигмоидальной функции – движение по градиенту которой плавное и, как следствие, нахождение минимума ошибки становится более стабильным и вероятным.

Замечание 1: Кроме того, что нелинейная функция активации позволяет модели обрабатывать нелинейные закономерности в данных, она выполняет и другую важную задачу. Нелинейные выходы одного слоя нейронов позволяют подключить к нему второй, третий и т. п. слои. Если сделать много слоев в Персептроне с линейной функцией активации или с пороговой (суть в том, что она не непрерывна), то можно свести все эти слои всего лишь к одному слою

(с математической точки зрения последовательность взвешенных сумм (полиномов) можно легко свести к одному другому полиному).

Замечание 2: Подобно регрессионным моделям обучение модели ИНС подразумевает правку только весовых коэффициентов. Никакие другие процессы или изменения в модели не происходят.

Алгоритм обратного распространения ошибки как специальный метод градиентного спуска, разработанный именно для ИНС, будет подробнее разобран далее.

Сверточные (Convolutional Neural Net) и Глубокие (DeepNet) Сети

Одна из базовых способностей людей и животных заключается в том, что мы легко распознаем визуальные образы под любым углом и в любом положении. Буква «А» останется для нас буквой «А» в какой бы области видимости наших глаз мы ее не увидели. Конечно же, при условии того, что выдержан определенный порог зашумленности, освещенности и т. п. Даже если текст расположен вверх ногами, то довольно быстро можно привыкнуть его читать. А уж небольшие наклоны или разные шрифты на рекламных щитах вообще не представляют для нас проблемы.

А вот системы по распознаванию текста, к сожалению, не обладают такими возможностями. Либо они заточены распознавать определенный тип шрифта, либо они очень чувствительны к сдвигам или наклонам, либо к масштабу, либо вообще ко всему. И несмотря на то, что нейронные сети в качестве прототипа использовали принципы работы мозга, долгое время не было хорошего решения этой проблемы. Данный недостаток называется проблемой чувствительности к пространственным искажениям. А добиться нужно так называемой инвариативности к таким искажениям трех основных типов: смещениям, поворотам, масштабированию.

По поводу пространственных искажения и появления сверточной сети, в источнике [28] сказано следующее:

«Дело в том, что работа зрительной коры гораздо сложнее: в ней происходит анализ и цвета, и текстуры, и движения; информация от

двух глаз объединяется для осуществления стереозрения; работает множество других механизмов. Высшие зрительные функции все еще остаются загадкой. И самое главное, до сих пор не известно, как происходит обучение. До конца неясно даже, что в структуре зрительной системы заложено генетически, а что формируется под влиянием опыта. Хотя структура зрительной системы у человека продолжает формироваться до 4–5 лет, это может быть, как реализацией генетической программы, лишь немного адаптирующей к окружению, так и детальным обучением, в результате которого создаются основные связи зрительного тракта.

Пытаясь разработать ИНС, которая была бы еще ближе к биологическому прототипу, японский ученый-компьютерщик Кунихика Фукусима предлагает принципиально новую модель Когнитрона в 1975 г., а в 1980 г. модификацию этой модели – «Неокогнитрон». Главное отличие этих сетей от MLP заключалось в том, что слои нейронов упаковывались не в линию, а в двумерную плоскость, чтобы информация циркулировала не только от слоя к слою, но еще сохраняла определенную пространственную ориентацию (рисунок 30). Кроме того, в этих сетях использовались принципы самоорганизации, то есть обучение происходило без учителя».

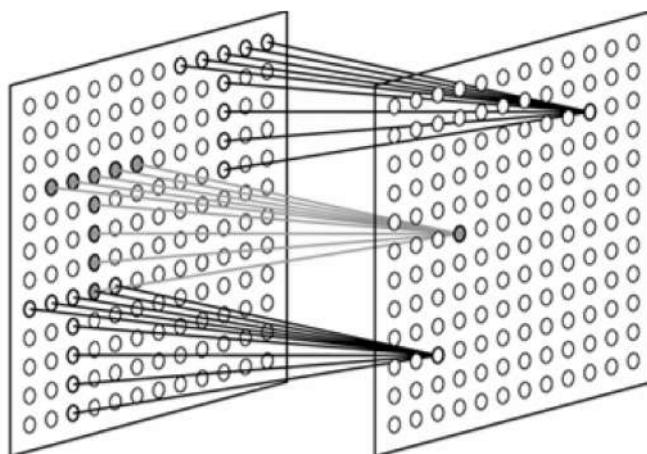


Рисунок 30. Рецептивные поля (квадратные плоскости) простых клеток, настроенных на поиск выбранного паттерна в разных позициях

В итоге подобная архитектура дала хорошие результаты при распознавании символов и даже рукописного текста относительно классических нейронных сетей. Однако по сравнению другими специализированными алгоритмами (не машинного обучения) на тот момент неокогнитрон «не дотягивал». Требовалось усложнять модель, но тогда были явные недостатки по скорости обучения и работы.

Однако общая концепция пространственно-ориентированных карт была очень ценной. И французский ученый и специалист в области обработки сигналов и компьютерного зрения Ян ЛеКун предлагает более упрощенную модель сети. Он убрал из неокогнитрона функции, которые нужны были только для того, чтобы быть похожим на реальный мозг. Он также показал, как можно использовать метод обратного распространения ошибки для обучения сетей, архитектура которых, как и у неокогнитрона, отдаленно напоминает строение коры мозга. И этот способ обучения, уже хорошо проверенный на тот момент, оказался куда эффективнее самоорганизации сетей. Так в 1995 появились нейронные сети Сверточного типа.

Структура сверточных сетей представлена на рисунке 31.

Нейроны здесь также упакованы не только в слои, но и двумерные карты. Информация циркулирует слева на право по нейронам такого же типа (как и в MLP). Но главная особенность заключается в том, что сеть не полносвязная, то есть каждый нейрон имеет свою небольшую область видимости (на верхнем рисунке область видимости показана пунктирными линиями). Такое локальное восприятие и обобщение от слоя к слою и дает решение проблемы чувствительности к пространственным искажениям, о которых говорилось выше. Иными словами, Сверточная Нейронная Сеть (СНС) способна обрабатывать пространственную топологию.

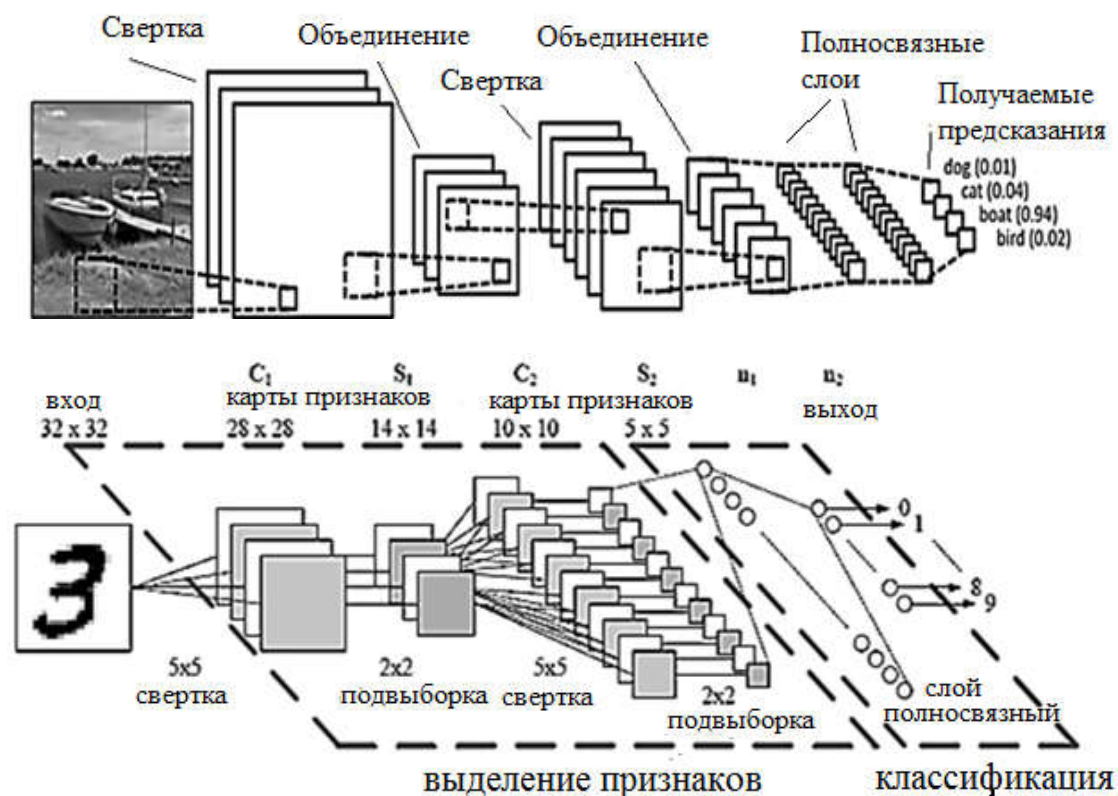


Рисунок 31. Структурная схема Сверточных нейронных сетей

Именно эта архитектура ИНС легла в основу так называемых Глубоких сетей, что породило понятие глубокого обучения (Deep Learning). Именно сеть этой архитектуры произвела такую шумиху в 2007 г., приблизив качество распознавания текста к человеческому уровню. Безусловно, данный метод не лишен недостатков, но во многих задачах именно визуального распознавания СНС является лучшим решением.

Глубокие сети, по сути, являются уже четвертой вехой развития когнитрона и их главная особенность заключается в многослойности (рисунок 32).

Но кроме этого важную роль играет разнообразное комбинирование определенных блоков карт. В глубоких сетях очень много разных модулей и ответвлений (может быть и такое, что сеть имеет несколько входов на разных уровнях).

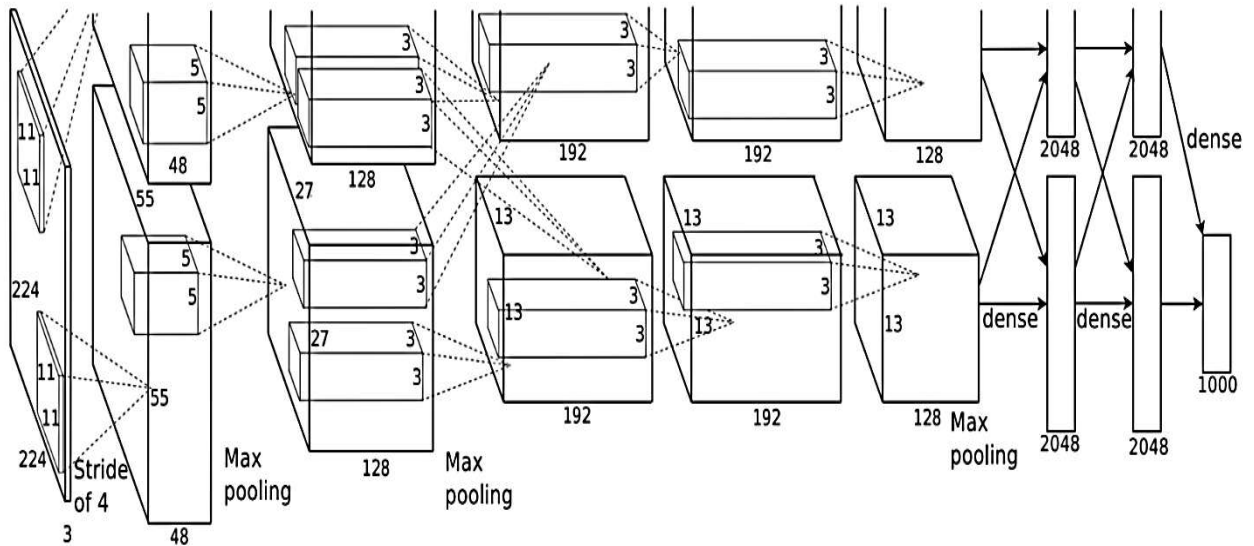


Рисунок 32. Схематичное представление Глубокой сети

Однако наращивание слоев приводит к целому ряду трудностей, как с объемом вычислений, так и с обучением сети, из чего складываются особенности, присущие именно глубокому обучению.

Карты (ART, SFAM)

Данный тип нейронных сетей разрабатывался как решение проблемы «пластичности-стабильности». Как известно, взрослый человек может понять какую-либо информацию с одного или нескольких объяснений. Речь, конечно, не идет о сложных вещах вроде курса математического анализа, но, тем не менее, мозг человека способен понять какие-то концепции даже за одно объяснение. Если переводить это на язык машинного обучения, то это означает, что нам не нужно многократное предъявление схожих примеров, нам хватит одного-двух объектов обучающей выборки. Конечно, мозг способен на такие вещи не только или, правильнее будет сказать, не столько благодаря быстро обучающимся нейронам, а потому что, воспринимая чье-то объяснение (например, когда вы читаете инструкцию к телевизору) мозг уже имеет представление о многих вещах. Иными словами, мозг уже обучен, у него есть предыдущий опыт. Кроме того,

разум способен держать контекст и т. п. Если объяснять упрощенно, то мозг дообучается, а не переобучается в таких случаях.

Но даже если не принимать во внимание контекст и априорный опыт, то возможно ли сделать обучение сети столь же пластичным, ну или близким к этому? Первая задача разработчиков сетей ART заключалась в том, чтобы сделать процесс обучения быстрым, чтобы не проходить итеративный градиентный спуск для правки весов. Эта проблема и есть проблема пластичности ИНС.

Но была и вторая задача. Опять же, вдохновляясь биологическим прототипом, разработчики понимали, что если человек узнает, что по какому-то вопросу у него на протяжении некоторого времени была неверная точка зрения, то он достаточно просто скорректирует отдельные представления о мире. У него не разрушится вся картина мира, он не сойдет с ума и не потеряет другие знания. Конечно, тут мы не берем в расчет психологические факторы, которые помешают ему это сделать.

Но если не рассматривать такие ситуации, то можно принять тот факт, что мозг очень пластичен, то есть может обучиться\дообучиться с малого количества раз и в то же время стабилен – одни знания могут быть заменены на другие без разрушения другой информации.

В то же время информация в классических ИНС (в MLP) распределяется по весам всей сети. Иными словами, нет одного четкого места, которое бы отвечало за определенный блок информации. Таким образом, если дообучение происходит на объектах тех же классов, которые использовались во время обучения, то многослойную сеть прямого распространения можно дообучить (но опять же медленно). Но если взять обученную сеть на одних классах и попытаться дообучить на новых, то вся старая информация начнет разрушаться.

Например, ИНС типа MLP была обучена для классификации квадратов и кругов по визуальным образам. Допустим, что сеть стала показывать хорошие показатели качества после предъявления различных образов квадратов и кругов в количестве 10 тыс. штук. Если

позже дообучить сеть на новых вариантах квадратов и кругов, то сеть может еще улучшить показатели. Но если начать обучать ее еще и на треугольниках, то под этот новый класс не выделится определенное место в памяти сети, а распределенная по весам информация, наоборот, начнет меняться, разрушая обученное состояние.

Это и есть «проблема стабильности ИНС». Она была второй задачей разработчиков сетей типа ARTMAP (Творческие карты, или просто Карты).

Первые варианты Карт (ARTMAP 1, ARTMAP 2) были очень сложными, избыточными и во многом были инженерным творением в виде электрических схем, а не алгоритмом.

После возвращения интереса к нейронным сетям этот тип сетей был существенно оптимизирован и модифицирован. Один из самых успешных и эффективных вариантов называется Simplified FUZZY ARTMAP (упрощенная нечеткая Карта). Под нечеткостью здесь понимается то, что вычисления основаны на нечеткой логике. Детали работы нечеткости в этом подразделе не приводятся.

Структура такой сети представлена на рисунке 33.

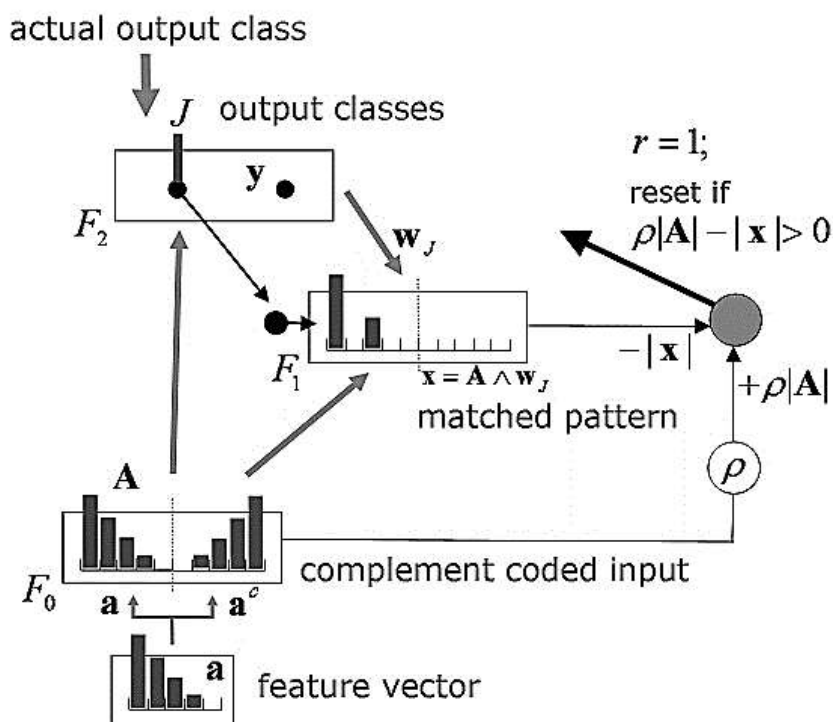


Рисунок 33. Структура основных модулей системы ARTMAP

Собственно, правильнее будет назвать это системой, содержащей нейроны, так как в структуре есть множество сторонних блоков (узлов), которые не очень красиво вписываются в биологическую концепцию мозга (хотя и не противоречат ей).

Итак, рассмотрим представленную на рисунке структуру снизу-вверх. В самом низу имеем вектор признаков «а» (как и во всех других методах ML). Данный вектор обязательно должен быть нормализован (отмасштабирован), то есть сеть не просто очень чувствительна к ненормализованным данным, а вообще не работает без нормализации. Соответственно если природа каких-то признаков неизвестна и не ясно, какой диапазон может быть в данных, то применение этой сети затрудняется (нужно специальным образом контролировать данные на вход) или вовсе отменяется.

Далее нормализованный вектор a дополняется комплементарной (обратной) парой a^c : там, где в исходном векторе было значение 0, в обратной паре будет 1, и наоборот (при условии нормализации от 0 до 1). Для этого и была необходима нормализация.

Далее, полученный вектор сравнивается со всеми шаблонами в базе нейронной сети. В данной случае это именно база, а не просто распределенная информация. Один нейрон в такой архитектуре представляет собой одну единицу информации, один конкретный шаблон\образ. То есть один нейрон выражает какой-то образ обучающей выборки (например, квадрат или круг). При обучении каждый поступающий вектор сравнивается с уже имеющейся базой образов, которые сохранены в нейронах сети.

Если близких образов нет, то образ сохраняется как новый, то есть создается новый нейрон и вектор входного образа полностью копируется в веса нового нейрона. Так с одного раза происходит полное корректное запоминание. При этом за нейроном закрепляется метка u (то есть это по-прежнему обучение с учителем, и для каждого входного вектора имеется ответ).

Если в базе уже имеется образ, похожий на входной образ, то происходит следующее. Образ в базе, выраженный нейроном и его

веса, подтягивается\приближается ко входному. Допустим, у сети в базе уже имеется некое представление о квадрате, и мы подаем еще один похожий образ квадрата, тогда после подтягивания в базе будет храниться нечто среднее, какой-то усредненный образ этих двух квадратов. Соответственно надо четко понимать, по каким принципам рассчитывается это усреднение. Ведь это все-таки не искусственный интеллект, и по одному образу человека и дельфина, сеть не усреднит их в образ млекопитающих. Надо понимать, что это лишь векторное усреднение. То есть если входной образ $(0.3, 0.7)$, а образ в базе $(0.35, 0.8)$, то итоговый будет $(0.325, 0.75)$. Что с точки зрения гипотезы компактности позволяет решать задачу обобщения.

Замечание: напомним, что гипотеза компактности говорит нам о том, что объекты одного класса обладают близкими по значению признаками (или какой-либо комбинацией признаков), из-за чего их вообще можно отделить от объектов другого класса. Если пойти еще дальше, то можно сказать, что объекты сходных классов располагаются в некотором N -мерном пространстве признаков также близко. Тут можно сказать, что объекты потому и принадлежат какому-то одному смысловому классу, потому что по какому-то признаку или группе (комбинации) похожи.

Итак, данная сеть способна сохранять образ с одного предъявления, что решает проблему пластичности. В сети сразу появляется образ с меткой класса, к которому принадлежит этот образ. Таким образом, вместо минимизации ошибки данная сеть обучается путем создания похожих прототипов и сравнением входного образа с прототипами. В рабочем режиме сеть выдает ответ Y в виде метки того нейрона, образ которого сильнее всего похож на текущий входной образ.

При этом данная сеть также решает задачу стабильности, так как теперь есть четкие хранилища информации и система организации новых нейронов. Так что даже при предъявлении новых классов сеть будет стабильно работать.

Данный тип сети хорошо зарекомендовал себя в медицине и при распознавании звуковых сигналов. Однако есть и ряд минусов данной сети:

- при равной степени обобщения сеть ARTMAP потребует существенно большее количество памяти и вычислений, чем MLP, если речь идет о входных векторах большой размерности и очень больших обучающих выборках;
- сеть не способна учитывать топологию пространства и сложные абстрактные признаки (так как слоев по сути нет).

Рекуррентные сети (Recurrent Neural Network)

Рекуррентные нейронные сети (RecurrentNeuralNetwork; RNN) – вид нейронных сетей, в которых имеется обратная связь. При этом под обратной связью подразумевается связь от логически более удаленного элемента к менее удаленному (рисунок 34). Наличие обратных связей позволяет запоминать и воспроизводить целые последовательности реакций на один стимул.

В сетях такого типа возникает эффект памяти и способности воспринимать не только статичный образ, но и динамику образов (так как есть возможность учитывать историю через обратную связь).

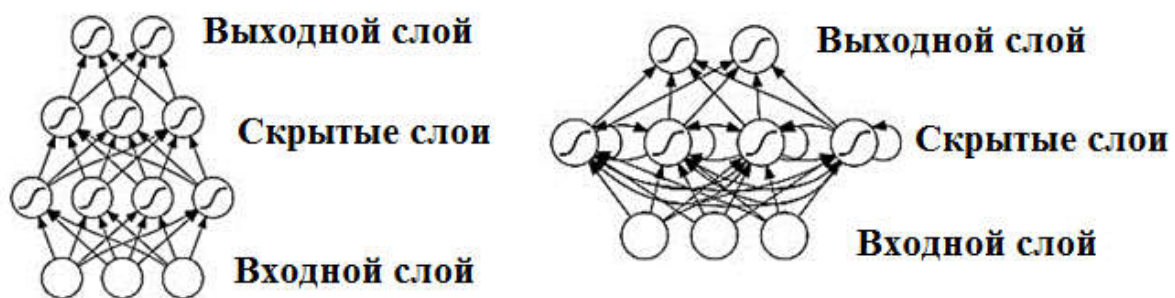


Рисунок 34. Сеть прямого распространения (слева) и рекуррентная сеть (справа)

Согласно Википедии [1]: «Долгая краткосрочная память (англ. Longshort-termmemory; LSTM) – разновидность архитектуры рекуррентных нейронных сетей (RNN), предложенная в 1997 году Сеппом Хохрайтером и Юргеном Шмидхубером. Как и большинство рекур-

рентных нейронных сетей, LSTM-сеть является универсальной в том смысле, что при достаточном числе элементов сети она может выполнить любое вычисление, на которое способен обычный компьютер, для чего необходима соответствующая матрица весов, которая может рассматриваться как программа. В отличие от традиционных рекуррентных нейронных сетей, LSTM-сеть хорошо приспособлена к обучению на задачах классификации, обработки и прогнозирования временных рядов в случаях, когда важные события разделены временными лагами с неопределенной продолжительностью и границами.

Относительная невосприимчивость к длительности временных разрывов дает LSTM преимущество по отношению к альтернативным рекуррентным нейронным сетям, скрытым Марковским моделям и другим методам обучения для последовательностей в различных сферах применения. Из множества достижений LSTM-сетей можно выделить наилучшие результаты в распознавании несегментированного слитного рукописного текста и победу в 2009 году на соревнованиях по распознаванию рукописного текста (ICDAR). LSTM-сети также используются в задачах распознавания речи, например, LSTM-сеть была основным компонентом сети, которая в 2013 году достигла рекордного порога ошибки в 17,7% в задаче распознавания фонем на классическом корпусе естественной речи TIMIT. По состоянию на 2016 год, ведущие технологические компании, включая Google, Apple, Microsoft и Baidu, используют LSTM-сети в качестве фундаментального компонента новых продуктов».

Самоорганизующиеся карты (Self-organization map, SOM)

Согласно ресурсу [29]: «Такие сети представляют собой соревновательную нейронную сеть с обучением без учителя, выполняющую задачу визуализации и кластеризации. Является методом проецирования многомерного пространства в пространство с более низкой размерностью (чаще всего, двумерное), применяется также для решения задач моделирования, прогнозирования и др. Является одной из версий нейронных сетей Кохонена. Самоорганизующиеся

карты Кохонена служат, в первую очередь, для визуализации и первоначального («разведывательного») анализа данных.

Геометрическая суть алгоритма такая, что близкие объекты в многомерном пространстве (схожие животные по тысяче признаков) будут расположены близко и на двумерной карте, где объекты будут представлены точками. Собственно, обучение сети это и есть процесс укладки таких точек (итеративного оттягивания таких точек от начальных позиций, см. рисунок 35).

Сигнал в сеть Кохонена поступает сразу на все нейроны, веса соответствующих синапсов интерпретируются как координаты положения узла, и выходной сигнал формируется по принципу «победитель забирает все», то есть ненулевой выходной сигнал имеет нейрон, ближайший (в смысле весов синапсов) к подаваемому на вход объекту. В процессе обучения веса синапсов настраиваются таким образом, чтобы узлы решетки «располагались» в местах локальных сгущений данных, то есть описывали кластерную структуру облака данных, с другой стороны, связи между нейронами соответствуют отношениям соседства между соответствующими кластерами в пространстве признаков».

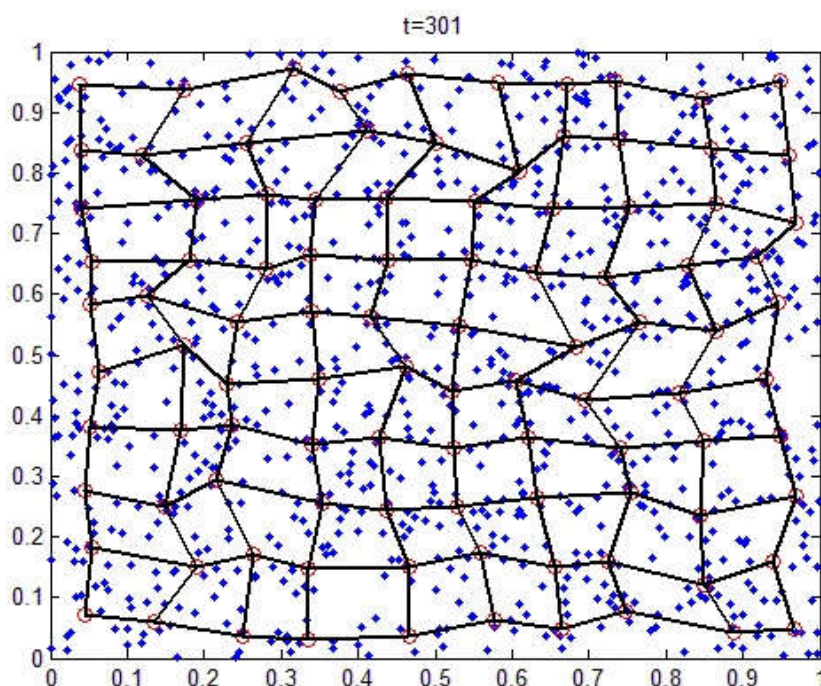


Рисунок 35. Визуализация двумерной плоскости с точками (проекциями) объектов в SOM

Автокодировщики (AutoEncoder)

Согласно Википедии [1]: «Автокодировщики (AutoEncoder) – специальная архитектура искусственных нейронных сетей, позволяющая применять обучение без учителя при использовании метода обратного распространения ошибки. Простейшая архитектура автокодировщика – сеть прямого распространения, без обратных связей, наиболее схожая с персептроном и содержащая входной слой, промежуточный слой и выходной слой. В отличие от персептрона, выходной слой автокодировщика должен содержать столько же нейронов, сколько и входной слой (см. рисунок 36).

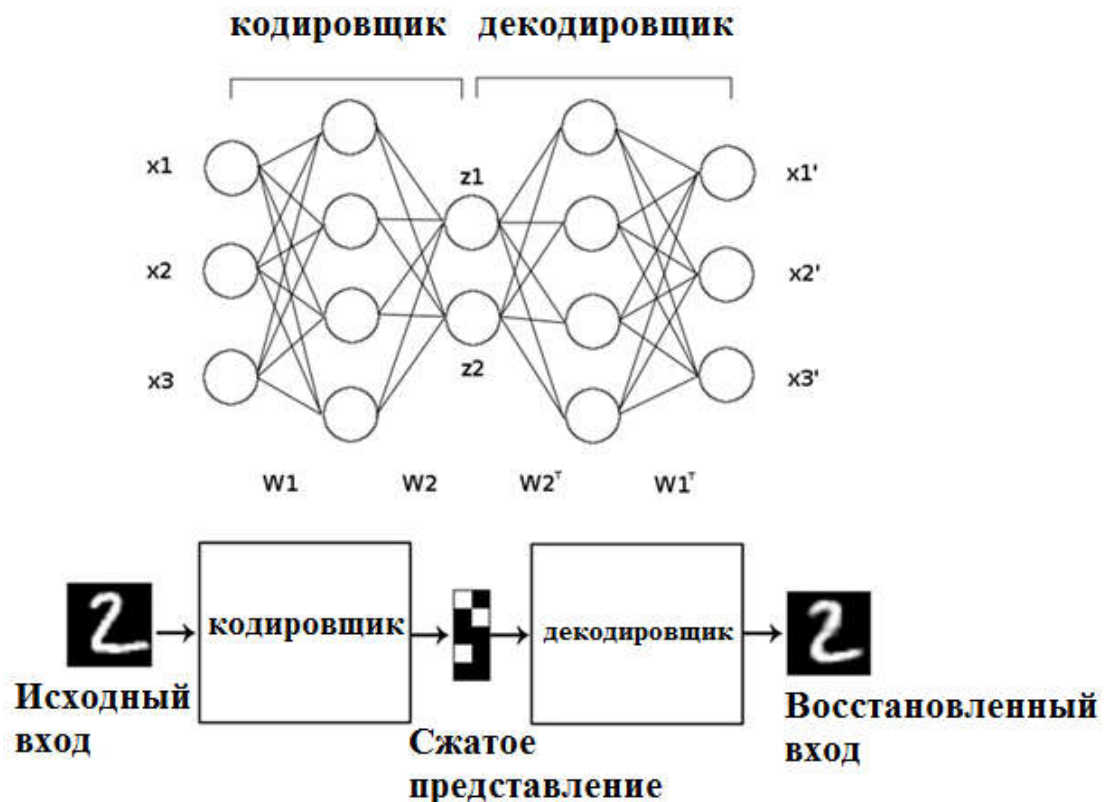


Рисунок 36. Схема сети типа Автокодировщик

Основной принцип работы и обучения сети автокодировщика – получить на выходном слое отклик, наиболее близкий к входному. Чтобы решение не оказалось тривиальным, на промежуточный слой автокодировщика накладывают ограничения: промежуточный слой должен быть или меньшей размерности, чем входной и выходной

слои, или искусственно ограничивается количество одновременно активных нейронов промежуточного слоя – разреженная активация. Эти ограничения заставляют нейросеть искать обобщения и корреляцию в поступающих на вход данных, выполняя при этом их сжатие. Таким образом, нейросеть автоматически обучается выделять из входных данных общие признаки, которые кодируются в значениях весов сети. Так, при обучении сети на наборе различных входных изображений, нейросеть может самостоятельно обучиться распознавать линии и полосы под различными углами.

Какое-то время автокодировщики применялись каскадно для обучения глубоких (многослойных) сетей, а именно для предварительного обучения глубокой сети без учителя. Для этого слои обучаются друг за другом, начиная с первых. К каждому новому необученному слою на время обучения подключается дополнительный выходной слой, дополняющий сеть до архитектуры автокодировщика, после чего на вход сети подается набор данных для обучения. Веса необученного слоя и дополнительного слоя автокодировщика обучаются при помощи метода обратного распространения ошибки. Затем слой автокодировщика отключается и создается новый, соответствующий следующему необученному слою сети. На вход сети снова подается тот же набор данных, обученные первые слои сети остаются без изменений и работают в качестве входных для очередного обучаемого автокодировщика слоя. Так обучение продолжается для всех слоев сети за исключением последних. Последние слои сети обычно обучаются без использования автокодировщика при помощи того же метода обратного распространения ошибки и на маркированных данных (обучение с учителем).

В последнее время автокодировщики мало используются для описанного «жадного» послойного предобучения глубоких нейронных сетей. После того как этот метод был предложен в 2006 г. Джеффри Хинтоном и Русланом Салахутдиновым, достаточно быстро оказалось, что новые методы инициализации случайными весами с определенным видом распределения оказываются эффективными

для улучшения сходимости. А предложенная в 2014 г. пакетная нормализация позволила обучать еще более глубокие сети, предложенный же в конце 2015 г. метод остаточного обучения позволил обучать сети произвольной глубины».

Основными практическими приложениями автокодировщиков остаются уменьшение шума в данных, а также уменьшение размерности многомерных данных для визуализации. Кроме того, с определенными оговорками, касающимися размерности и разреженности данных, автокодировщики могут позволять получать проекции многомерных данных, которые оказываются лучше тех, что дает метод главных компонент либо какой-либо другой классический метод. Помимо этого можно использовать такие сети, как алгоритм сжатия данных.

Крайне важно отметить, что если же взять архитектуру автокодировщика, но обучать его с учителем, то есть Y будет отличаться от X , то получится сеть для преобразования одного изображения в другое изображение через некоторую сложную функцию. На основе данного преобразования строятся различные фильтры изображений и видео (например, наложение различных эффектов) или же можно решать сложные задачи сегментации изображения. На сегодняшний день подобные архитектуры очень распространены и хорошо решают задачи анализа сцены для беспилотного автомобиля.

Импульсные (Спайковые) сети

Согласно Википедии [1]: «Первая научная модель импульсной нейронной сети была предложена Аланом Ходжкином и Эндрю Хаксли в 1952 году. Эта модель описывала, как потенциалы действия возникают и распространяются. Импульсы, однако, как правило, не передаются непосредственно между нейронами. Связь требует обмена химическими веществами, которые называются нейротрансмиттерами, в синаптической щели.

С точки зрения теории информации, проблема заключается в отсутствии модели, которая бы объясняла, как кодируется инфор-

мация и декодируются серии последовательностей импульсов, то есть потенциалы действия. Для нейробиологии все еще открытым является ответ на вопрос: нейроны связываются с помощью частотного или временного кодирования? С помощью временного кодирования один импульсный нейрон может заменять сотни скрытых элементов частотной нейронной сети.

Что же касается отличия спайковых сетей от классических, то тут необходимо отметить следующее. Если основное отличие сверточных сетей от классических заключается в структуре и организации связей, то импульсные сети, напротив, по структуре похожи на MLP. Их главное отличие заключается в том, что нейроны обмениваются короткими (у биологических нейронов – около 1-2 мс) импульсами одинаковой амплитуды (у биологических нейронов – около 100 мВ). Является самой реалистичной, с точки зрения физиологии, моделью ИНС (см. рисунок 37)».

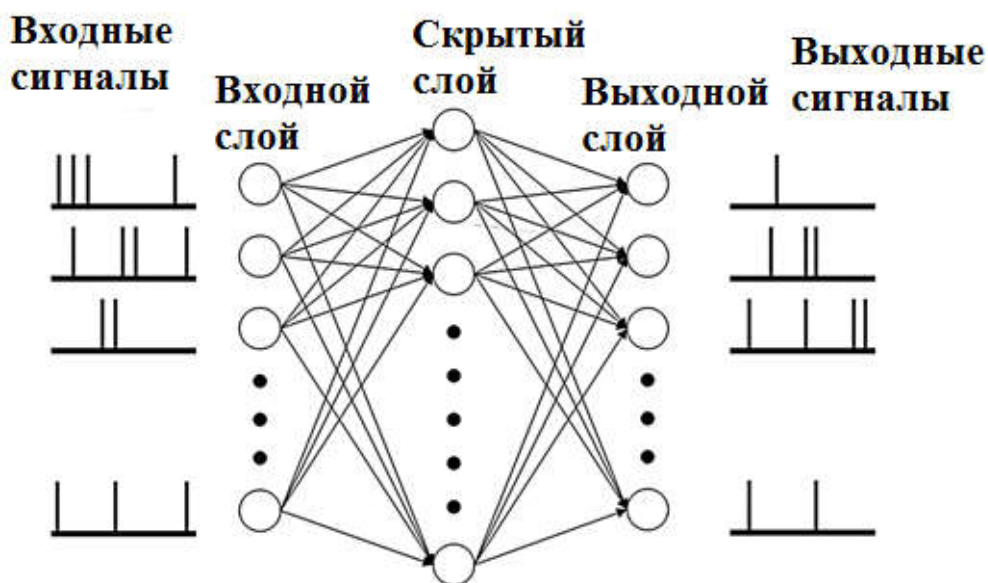


Рисунок 37. Схематичное представление структуры Импульсной нейронной сети

Данный тип сети еще находится на стадии активного исследования. Тем ни менее ИмНС уже успешно применялась для динамического управления мобильным роботом. В качестве основных ее преимуществ можно назвать:

- Перспективность в обработке динамической информации (например, видеопотока, временных рядов и т. п.).
- Более плотное и эффективное кодирование информации по сравнению с MLP.
- Более эффективное расходование электроэнергии при физической реализации.
- Высокая степень эффективности параллельного выполнения при физической реализации.

К недостаткам можно отнести слабую изученность, сложность математической модели и вычислительную сложность при выполнении на обычном железе. А главный недостаток заключается в отсутствии хорошего алгоритма обучения.

Причины бурного развития ИНС сегодня

Казалось бы, раз ИНС такой мощный инструмент\метод и первая нейронная сеть была предложена так давно, то почему только сейчас появляется столько технологий на нейронных сетях? Тому есть две причины:

Чтобы исследовать глубокие ИНС, нужно проводить эксперименты, а они могли длиться по месяцу, а если и меньше, то нужно было все равно проводить десятки экспериментов. Таким образом, исследовать сети было сложно.

Поэтому появление соответствующих параллельных вычислителей и общее увеличение мощности компьютеров изменило эту ситуацию. Кроме того, появились обычные видео карты (GPU), на которых любой мог запускать сложные вычисления. Поэтому появилось много исследователей, которые проводили эксперименты. То есть, как ни странно, развитие тормозилось не из-за принципиальных возможностей ИНС, а из-за того, что экспериментировать было долго.

Функция активации «Relu» упростила функцию сетей в целом и сделала вычисления Глубоких сетей еще быстрее.

Психологический фактор. Извилистая история нейронных сетей и долгое пребывание в состоянии заморозки создали дурную славу нейронным сетям, и исследователи очень неохотно занимались этой сферой.

Возрастание интереса сразу сложило разные кусочки мозаики, и Сверточные сети получили вторую жизнь, а их успех дал еще больший толчок.

Стоит также отметить, что сегодня наблюдается обратный эффект. Пресса и Интернет часто преувеличивают возможности ИНС или допускают неточности, из-за чего может сложиться впечатление, что нейронные сети – это искусственный интеллект.

Борьба с переобучением в ИНС

Тема переобучения была подробно рассмотрена ранее, поэтому просто перечислим способы борьбы с этим эффектом, подходящие для большинства типов ИНС:

- Кросс-валидация. Данный метод не зависит от модели, это общий подход к тренировке моделей, помогающий контролировать и бороться с переобучением особенно на малых выборках.
- Уменьшение числа слоев\нейронов. Уменьшение слоев и нейронов явно влияет на сложность модели, однако необходимо искать компромисс между обобщающей способностью нейронов и степенью абстрактности признаков.
- Добавление L2 регуляризации. Аналогично Регрессии этот дополнительный метод штрафует высокие веса модели.
- Локальность восприятия. Задавая рецептивное поле, восприятие нейрона можно также уменьшить не только количество параметров или сложность модели, но, и более того, – определить влияющие факторы, что позволит нейрону решать локальные задачи и складывать их в глобальное решение. Это крайне важная особенность ИНС.

Обратное распространение ошибки

Итак, как и у любой другой модели в машинном обучении, у ИНС есть 2 этапа работы:

- обучение;
- использование, моделирование или получение отклика, прогноза (все это подразумевает расчет выхода по обученной модели).

При этом использование сети называют прямым распространением (forwardpropagation), потому что сигнал со входа сети распространяется к выходу (рисунок 38), то есть слева направо.

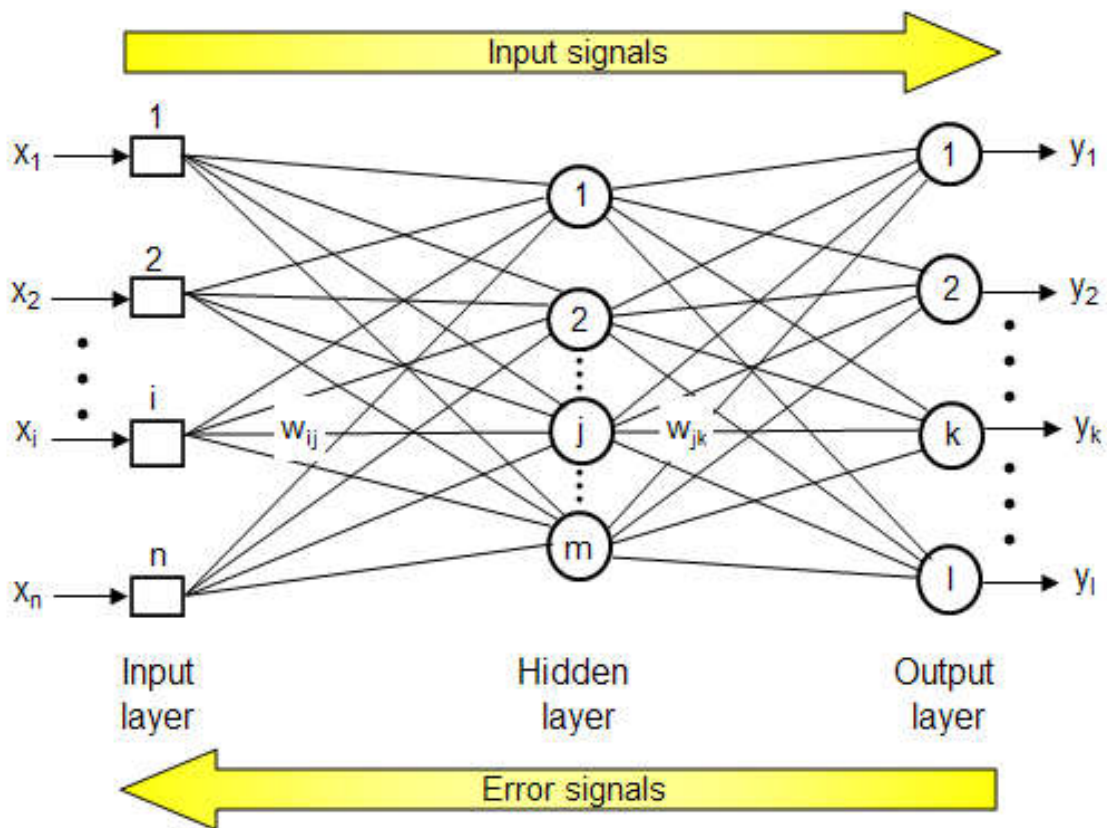


Рисунок 38. Обобщенная схема распространения сигналов по MLP

На рисунке 39 изображена принципиальная схема прямого распространения сигнала со входа на один нейрон. В случае нескольких нейронов в слое схожие вычисления будут проводиться для каждого нейрона. После чего сигнал двинется к следующему слою, и если в нем также несколько нейронов, то там будут сходные вычисления

для каждого нейрона и так далее. Каждый нейрон берет взвешенную сумму своих входов, считает функцию активации и выдает выход для следующего слоя и т. д.

Обучение называют обратным распространением (backpropagation) ошибки, потому что информация идет наоборот, с выхода сети, и информация эта – об ошибке, а не о входном образе, как при прямом распространении.

В модель нейрона на рисунке 39 включен пороговый элемент (bias), который обозначен символом b_k . Эта величина отражает увеличение или уменьшение входного сигнала, подаваемого на функцию активации. По сути, это свободный коэффициент при некотором признаке в нулевой степени (как и у Регрессии).

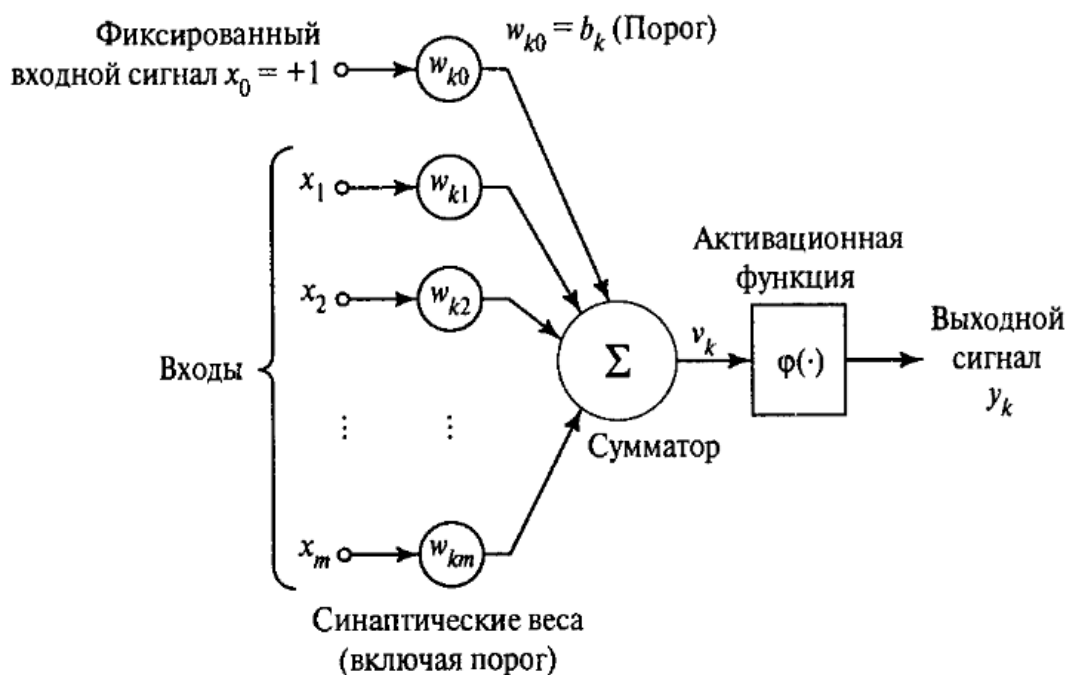


Рисунок 39. Принципиальная схема прямого распространения сигнала

В математическом представлении функционирование нейрона k можно описать следующей парой уравнений:

$$U_k = \sum_i^m w_{ki} x_i,$$

$$y_k = \varphi(U_k),$$

где U_k – индуцированное локальное поле, или потенциал активации.

Серия таких уравнений, вычисленных последовательно, в конце концов даст результат всей сети. Если в последнем слое сети один нейрон, то соответственно получим скаляр или число, а если нейронов несколько, то получим вектор значений для каждого входного образа X .

Теперь рассмотрим алгоритм обратного распространения ошибки для обучения искусственной нейронной сети. Он разработан в первую очередь для сетей MLP-типа. Но также подходит для сверточных сетей, автокодировщиков и, при определенных модификациях, для рекуррентных сетей.

Для начала введем определения:

Обучить сеть – минимизировать ошибку, а именно минимизировать разницу между выходом сети (реакцией на вход) и требуемой реакцией, то есть это обучение с учителем.

Минимизация ошибки производится путем итеративных правок\корректировок весов ИНС. Корректируются в ИНС только веса и больше ничего.

Минимизация ошибки происходит для каждого отдельно взятого входного образа. Повторяя такую минимизацию для каждого образа много раз, получаем общую минимизацию. Количество проходов по тренировочной выборке называется эпохами.

Начнем рассмотрение алгоритма с последнего (выходного) слоя сети (рисунок 40).

Сперва предположим, что выходной слой содержит только один нейрон. Тогда:

E, e – ошибка выхода сети для текущей пары (X, Y) ;

Y, y – требуемый выход для соответствующего входного образа X ;

$o = \varphi(U)$ – выход нейрона, рассчитывается как сигмоида от U ;

U – результат взвешенной суммы, или индуцированное локальное поле;

W_{ij} – вес, соединяющий некоторый нейрон слоя i с некоторым нейроном слоя j .

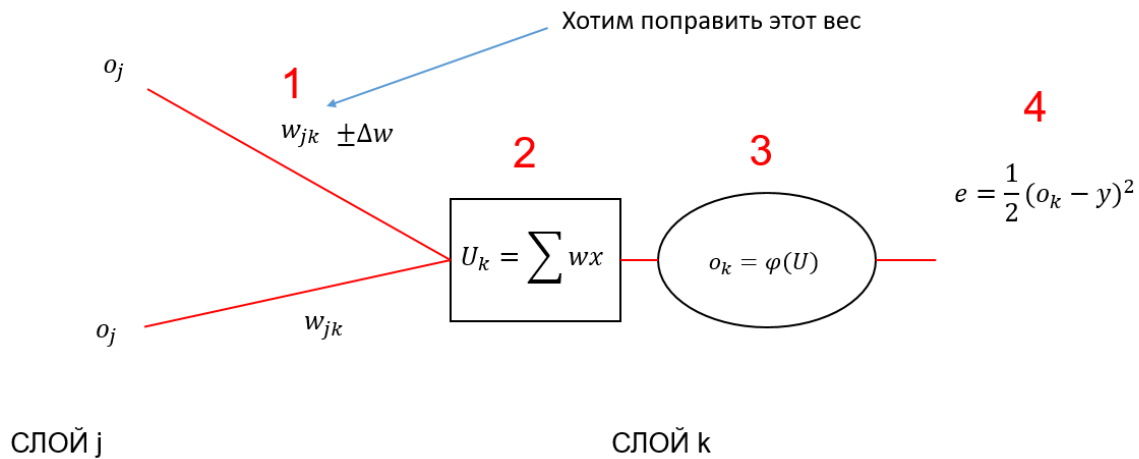


Рисунок 40. Представлена схема последнего слоя сети MLP

Индексы при символах означают индексы слоев. Так как в каждом слое имеется несколько нейронов, то о каком нейроне идет речь, когда пишется o_j ? О любом произвольном в этом слое, так как формула обобщенная и не вносит конкретики о номере нейрона внутри слоя.

Первая идея алгоритма\метода заключается в том, что мы движемся по градиенту ошибки, как и во всех численных градиентных методах. Мы не знаем, где точное решение, но если итеративно понемногу уменьшать ошибку для каждого входного образа, то мы рано или поздно придем к некому равновесному состоянию. Не обязательно минимуму ошибки, но к такому моменту, когда дальнейшие правки уже не будут коренным образом менять ситуацию, а система войдет либо в полную заморозку (изменений вообще не будет), либо в некий колебательный процесс около некоторой матрицы весов.

Собственно, двигаясь по антиградиенту ошибки $-\nabla F$, мы как раз и производим корректировки весов на некие дельта (в данном случае у весов нет никаких индексов, потому что это общая концепция правок):

$$w^{t+1} = w^t - \alpha \Delta w,$$

$$w^{t+1} = w^t - \alpha \nabla F(w^t),$$

$$\nabla F(w^t) = \frac{\partial E}{\partial w^t},$$

где w^{t+1} – вес в следующий момент времени;

w^t – вес в текущий момент времени;

Δw – правка веса на текущем шаге, в сторону уменьшения ошибки на текущем шаге;

α – размер шага, скорость движения по антиградиенту или сила правок весов.

Еще раз о том, почему используется именно градиент. Потому что нельзя рассчитать точное значение правок для всех весов сразу, ведь мы не знаем вклад каждого веса в ошибку, мы лишь знаем значение ошибки и направление. Чтобы понять суть идеи, представьте, что некая правка $\pm \Delta w$ некоего веса w даст нам увеличение или уменьшение ошибки на выходе сети (см. рисунок 40). Получается, что небольшие правки весов приводят к некоторым небольшим изменениям конечной ошибки сети. Возможно ли найти функциональную зависимость между этими небольшими изменениями? Что есть отношение небольшого изменения ошибки к изменению какого-либо веса? Это производная $\frac{\partial E}{\partial w}$.

Таким образом, найдя аналитическое выражение производной ошибки по любому весу, мы определим характер воздействия правок этого веса на ошибку. А в конкретных точках этой функции (при конкретном входном образе и всех других параметрах сети) мы сможем точно подсчитать значение функции производной, то есть сможем оценить знак и значение dE , на которое будет меняться ошибка при наших правках dw . А значит, мы можем выбрать такую правку, чтобы уменьшить эту самую E . При этом все остальные веса, кроме того который правится в данный момент, замораживаем.

Итак, для последнего уровня все относительно просто. Чтобы найти производную $\frac{\partial E}{\partial w}$, необходимо взять серию производных по каждой вложенной функции (по цепному правилу дифференцирования):

$$\begin{aligned} w_{jk}^{t+1} &= w_{jk}^t - \alpha \frac{\partial E}{\partial w_{jk}} = \\ &= w_{jk}^t - \alpha \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial U_k} \frac{\partial U_k}{\partial w_{jk}} = w_{jk}^t - \alpha (o_k - y) o_k (1 - o_k) o_j. \end{aligned}$$

Но что делать с внутренними слоями? Дело в том, что для нейронов внутренних слоев мы не можем явно рассчитать не то что значение правки, но и значение ошибки. Ведь чтобы рассчитать значение ошибки, надо знать, какое значение выхода должно быть у каждого нейрона внутреннего слоя. А мы не можем этого знать, ведь не знаем конкретный вклад каждого нейрона в ошибку. Мы знаем производную, то есть ближайший характер изменений ошибки от правок весов, но не можем посчитать нужные значения для произвольных точек.

Но что если пойти таким же путем, как и раньше? Допустим, что некая правка $\pm \Delta w$ некого веса w (теперь уже на скрытом слое) даст нам увеличение или уменьшение ошибки на выходе сети (см. рисунок 41). Получается, что небольшие правки этого веса тоже приводят к некоторым небольшим изменениям конечной ошибки сети при условии, что все остальные веса сети константны в рамках этого временного среза.

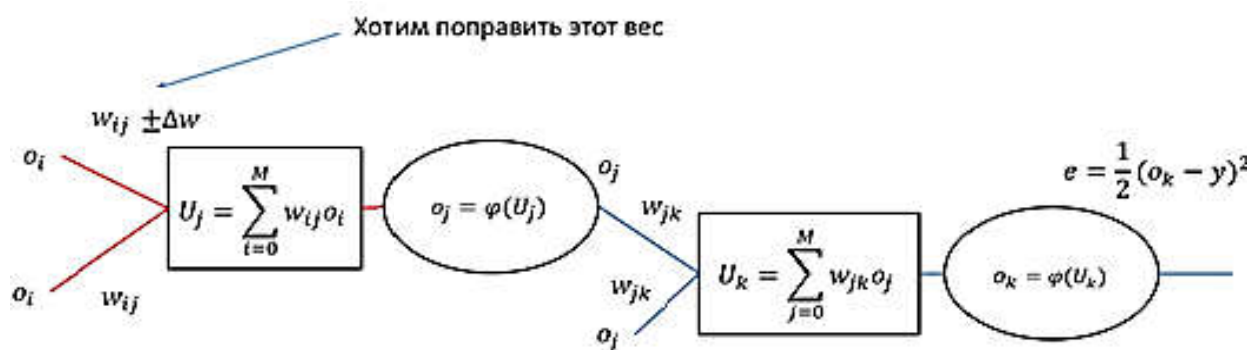


Рисунок 41. Схема последних двух слоев сети

Поэтому можно взять производную выходной ошибки по этому нейрону. Это будет такая же цепочка производных (только вес w_{ij} теперь будет константой, а не переменной, поэтому он и останется после взятия производной по o_j , а берем мы именно по o_j , чтобы пройти (протиснуть информацию) дальше).

$$w_{ij}^{t+1} = w_{ij}^t - \lambda \frac{\partial E}{\partial w_{ij}} = w_{ij}^t - \lambda \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial U_k} \frac{\partial U_k}{\partial o_j} \frac{\partial o_j}{\partial U_j} \frac{\partial U_j}{\partial w_{ij}}$$

Теперь можно подставить каждую производную и получить формулу корректировки веса в скрытом слое. Однако выше мы делали серьезное допущение, что в последнем слое будет только один нейрон. А это может быть не так. И нас интересует именно этот общий случай и для любого слоя. Таким образом, из этой формулы надо выделить закономерность, которую можно использовать для произвольного слоя:

$$w_{ij}^t - \lambda \left| \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial U_k} \frac{\partial U_k}{\partial o_j} \right| \left| \frac{\partial o_j}{\partial U_j} \right| \left| \frac{\partial U_j}{\partial w_{ij}} \right| .$$

1 2 3

В формуле градиента ошибки по весу скрытого слоя присутствуют 3 составляющих. 1 – это составляющая следующего слоя, 2 – это составляющая текущего слоя и 3 – составляющая предыдущего слоя. То есть изменение выходной ошибки складывается из влияния этих составляющих, если все остальные элементы среды заморозить. А раз так, то можно сделать два вывода:

Первая составляющая в случае нескольких нейронов в следующем слое будет не одна. Понятно, что текущий нейрон j -го слоя распространяет свою ошибку на все связанные с ним нейроны слоя k . А значит, надо просуммировать вклад, который оказывает корректировка веса по всем путям. Или, иными словами, просуммировать производные по разным связям нейронов. Тогда первый блок можно будет переписать так:

$$\sum_{k=0}^M \frac{\partial E}{\partial o_k} \frac{\partial o_k}{\partial U_k} \frac{\partial U_k}{\partial o_j} = (o_k - y) \varphi'_k w_{jk} .$$

Для сколь угодно далекого слоя можно не считать полный градиент, а брать вклад следующего слоя за основу. Ведь любой произвольный слой в первую очередь делает вклад именно в следующий слой. Таким образом, учитывая промежуточные вклады или точнее влияние всех промежуточных вкладов, мы сможем оценить итоговое влияние на ошибку. Опять же речь идет о небольших локальных приращениях $\pm \Delta w$ по конкретному w . Значит, можно выделить первую составляющую формулы в отдельную передаточную

величину вклада ошибки δ . Также ее называют производной ошибки по взвешенной сумме, по локальному индуцированному полю (по U). Тогда для последнего слоя:

$$\delta_k = (o_k - y) o_k (1 - o_k).$$

А для любого другого скрытого слоя:

$$\delta_j = (\sum \delta_k w_{jk}) o_j (1 - o_j).$$

Получается, что сигма рассчитывается как совокупность всех вкладов в ошибку (начиная с конца). И именно эта величина является той информацией, которая выражает ошибку от слоя к слою и распространяется обратно. Учитывая оба пункта одновременно, формулы для корректировок можно переписать следующим образом:

$\Delta w = -a \delta_k o_j = -a (o_k - y) o_k (1 - o_k) o_j$ – если корректируется вес последнего слоя.

$\Delta w = -a \delta o_i = a (\sum \delta_k w_{jk}) o_j (1 - o_j) o_i$ – если корректируется вес скрытого слоя.

Общий алгоритм можно представить так:

- Инициализировать веса случайным образом;
- Рассчитать выход сети прямым распространением сигнала;
- Рассчитать ошибку на выходе;
- Рассчитать корректировки весов текущего слоя по ошибке следующего;
- Обновить веса.

Условия завершения алгоритма обратного распространения ошибки:

- Требуемая величина ошибки;
- Максимальное количество итераций;
- Последние N проходов веса не изменились больше чем на T.

Замечание 1: шаг градиентного спуска или скорость градиентного спуска крайне важный параметр и нужен не только ради формальности. Дело в том, что при $a = 1$ движение по градиенту ошибки будет слишком быстрым, т.к. каждая корректировка отдельно взятого веса производится с учетом заморозки всех остальных весов.

А следующий вес корректируется на основании исходного значения ошибки (вычисленного еще до правки первого веса). В противном случае движение по градиенту будет плавным, но это очень расточительно с точки зрения вычислительных ресурсов (такой алгоритм будет очень медленным). Кроме того, точный пересчет в рамках одного образа и не нужен, так как полное уменьшение ошибки для одного образа не гарантирует факт того, что уменьшится общая ошибка. А уменьшить необходимо именно общую ошибку, а значит в рамках одного образа корректировки должны быть небольшими. Уточним, что на практике часто принимают $\alpha = 0.01$.

Замечание 2: конечно, рассмотренный алгоритм не лишен недостатков. Некоторые из них очевидны и их возможно устранить. Существует множество модификаций градиентного спуска, например: Adadelta, Adam и т. п. Есть и другие методики улучшения обучения. Один из самых мощных и простых – это Batching. Но все эти методы выходят за рамки пособия, так как при необходимой фундаментальной подготовке вы можете изучить их особенности самостоятельно.

Замечание 3: алгоритм обратного распространения ошибки линеен с точки зрения вычислительной сложности.

Замечание 4: более подробное рассмотрение модификаций алгоритма обратного распространения выходит за рамки этого пособия, как и рассмотрение альтернативных функций активации (таких как ReLU, LeakyReLU, Parametric ReLU, RandomizedReLU). Поэтому вам предлагается изучить эти вопросы самостоятельно.

Нечеткие нейронные сети

Нечеткой нейронной сетью (НС) обычно называют четкую нейросеть, которая построена на основе многослойной архитектуры с использованием специальных «И»-, «ИЛИ»-нейронов.

Нечеткая нейросеть функционирует стандартным образом на основе четких действительных чисел, нечеткой является только интерпретация результатов.

Нечеткие нейронные сети осуществляют выводы на основе аппарата нечеткой логики, а параметры функций принадлежности настраиваются с использованием алгоритмов обучения НС. Поэтому для подбора параметров таких сетей применим метод обратного распространения ошибки, изначально предложенный для обучения многослойного персептрона. Нечеткая нейронная сеть, как правило, состоит из четырех слоев: слоя фаззификации входных переменных, слоя агрегирования значений активации условия, слоя агрегирования нечетких правил и выходного слоя.

Наибольшее распространение в настоящее время получили архитектуры нечеткой НС вида ANFIS и TSK [18]. Доказано, что такие сети являются универсальными аппроксиматорами. Быстрые алгоритмы обучения и интерпретируемость накопленных знаний – эти факторы сделали сегодня нечеткие нейронные сети одним из самых перспективных и эффективных инструментов мягких вычислений.

Преимущества нечетких нейронных сетей

Основным преимуществом технологии нейрокомпьютинга служит возможность выразить зависимость «выход-вход» без предварительной аналитической работы по выявлению правил, а на основе обучения на примерах. Недостатком нейросетей является невозможность объяснить выходной результат, так как значения распределены по нейронам в виде значений коэффициентов весов. Основной трудностью в применении нечетких экспертных систем служит необходимость явно сформулировать правила проблемной области в форме правил. В нечетких экспертных системах легко построить объяснение результата в форме протокола рассуждений. Поэтому в настоящее время создаются гибридные технологии.

Примером гибридной технологии служит реализация базы нечетких правил на основе нейросети. База нечетких правил для двух входных переменных имеет следующую структуру:

$R_i: \text{if } x_{1i} \text{ is } A_{1i} \text{ and } x_{2i} \text{ is } A_{2i} \text{ then } z_i \text{ is } C_i.$

Простой реализацией базы нечетких правил служит интерпретация базы правил как таблицы определения некоторой функции, то есть базу правил можно представить обучающей выборкой:

$$\{((A_{1i}, A_{2i}), C_i)\}.$$

Например, обучающая выборка в нечетких терминах может быть сформулирована следующим образом:

$$\{((\text{малое}, \text{большое}), \text{около нуля})\}.$$

Чтобы совместить две технологии: технологию нечетких систем и технологию нейрокомпьютинга, необходимо предложить способ четкого дискретного представления непрерывных функций принадлежности. Чтобы представить в четких данных непрерывные функции принадлежности, выберем максимально большой интервал $[x_1, x_2]$, в котором представлены все нечеткие множества условных частей правил. Разбиваем интервал с равным шагом, тогда любое нечеткое значение представляется четким вектором. Другой способ представления нечеткого понятия в виде четких данных состоит в представлении нечеткого множества в виде совокупности α -срезов.

α -срезом называется четкое множество, включающее все элементы x некоторого нечеткого множества X , принадлежность которых больше равна α .

$$\mu_X(x) \geq \alpha.$$

Каждое α -подмножество представляется двумя числами – левой и правой границей α^L, α^R , то есть α -срезы четко представляют непрерывную функцию принадлежности.

Изменение элемента нейросети для адаптации к нечетким системам может касаться выбора функции активации, реализации операций сложения и умножения, так как в нечеткой логике сложение моделируется любой треугольной конормой ($\max, a + b - a \times b \dots$), а операция умножения треугольной нормой ($\min, a \times b, \dots$).

И-нейроном (AND-нейроном) называется нейрон, в котором умножение веса на вход моделируется конормой $S(w, x)$, а сложение – нормой $T(w, x)$.

ИЛИ-нейроном (OR-нейроном) называется нейрон, в котором умножение веса и входа моделируется нормой $T(w,x)$, а сложение взвешенных весов конормой $S(w,y)$ $Y = S(T(w_1,x_1),T(w_2,x_2))$.

Пример: $(\max(\min(w_1,x_1),\min(w_2,x_2)))$.

В качестве функции активации обычно используют функцию

$$F(x) = 1/(1+\exp(b(x-a))) .$$

Нечеткой нейросетью называют четкую нейросеть, которая построена на основе многослойной архитектуры с использованием «И-», «ИЛИ-нейронов».

Нечеткая нейросеть функционирует стандартным образом на основе четких действительных чисел. Нечеткой является только интерпретация результатов. При создании гибридной технологии, кроме объединения систем по данным, можно использовать нейрокомпьютинг для решения частной подзадачи нечетких экспертных систем, а именно настройки параметров функции принадлежности. Функции принадлежности можно сформировать двумя способами: методом экспертной оценки или на основе статистики. Гибридные технологии предлагают третий способ: в качестве функции принадлежности выбирается параметризованная функция формы (например, параметризованная Гауссова кривая), параметры которой настраиваются с помощью нейросетей. Настройка параметров может быть получена в рамках алгоритма обратного распространения ошибки.

Пусть задана следующая система нечетких правил:

If x_1 is A_{1i} and ... x_n is A_{ni} then z_i is C_i ,

где A_i – нечеткие числа; C_i – действительные числа. Значение $\alpha_j = \prod_i \alpha_i$ – сила или достоверность правила, $i = 1, \dots, n$ – номер входной переменной, $j = 1, \dots, m$ – количество правил.

$$Z = \frac{\sum_{j=1}^m a_j z_j}{\sum_{j=1}^m a_j} ,$$

где Z – вычисленное значение выхода.

Допустим, что разработана нейросеть с n входами и одним выходом. Каким образом такая НС может аппроксимировать базу нечетких правил? Любая совокупность нечетких правил может рассматриваться как нелинейное соответствие, заданное таблицей определения $\{(x,y)\}$, где x – вектор входа; y – желаемое значение выхода; а z – значение, вычисляемое нейросетью. Тогда можно определить текущую ошибку:

$$E^K = 1/2 \times (z^K - y^K)^2.$$

То есть можно применить стандартный алгоритм коррекции ошибки на основании данного определения

$$Z(t+1) = Z(t) - \zeta \times (\partial E^K / \partial Z),$$

где ζ – уровень обучения; $\partial E^K / \partial Z$ – направление градиента снижения ошибки.

$$Z(t+1) = Z(t) - \zeta \times (Z^K - Y^K) \times \alpha_j / (\alpha_1 + \alpha_2 + \dots + \alpha_m).$$

При применении стандартного алгоритма обратного распространения ошибки для того, чтобы настроить выход, необходимо изменить параметры функции принадлежности условных частей, то есть обучение сети позволит настроить функцию принадлежности с точки зрения обучающей выборки. При практической реализации системы нечетких правил важным является вопрос о типичных представителях нечетких значений в правилах. Большинство нечетких понятий, представленных лингвистическими переменными, выражает свои значения с помощью количественных нечетких множеств: NB – отрицательное большое; NM – отрицательное среднее; NS – отрицательное малое; ZE – около нуля; PS – положительное малое; PM – положительное среднее; PB – положительное большое.

Пример использования нечеткой нейронной сети

Рассмотрим нечеткие нейронные сети на примере нечеткого регулятора для стиральной машины [18], то есть построим fuzzy-neuro контроллер. Выберем для демонстрации технологии нечетких нейронных сетей архитектуру ANFIS (Adaptive Network Based Fuzzy Inference System). Основа интеграции нейронных сетей и систем нечеткого вывода заключается в том, что оба метода представляют

нелинейное отношение в пространстве входы-выходы. Важная задача нечеткого моделирования – это настройка функций принадлежности, являющаяся по существу задачей оптимизации. Как нейронные сети, так и генетические алгоритмы используются для ее решения. Самый простой подход требует назначить определенную параметризованную функцию формы в качестве функции принадлежности и подобрать параметры на основе обучения нейронной сети. Рассмотрим простой пример с тремя нечеткими правилами вывода в базе знаний.

R_1 : ЕСЛИ <количество_белья> is <много>

И <температура_воды> is <высокая> И <загрязненность> is <высокая>

ТО <длительность> is <высокая>;

R_2 : ЕСЛИ <количество_белья> is <много>

И <температура_воды> is <высокая> И <загрязненность> is <низкая>

ТО <длительность> is <низкая>;

R_3 : ЕСЛИ <количество_белья> is <мало>

И <температура_воды> is <низкая> И <загрязненность> is <низкая>

ТО <длительность> is <малая>;

Зададим следующие функции формы для высказываний. Пусть высказывание <количество_белья> is <мало> соответствует функции $L_1(x)$ и высказывание <количество_белья> is <много> – функции $H_1(x)$.

$$L_1(x) = 1/(1 + \exp(b_1(x - c_1))),$$

$$H_1(x) = 1/(1 + \exp(-b_1(x - c_1))),$$

причем $L_1(x) + H_1(x) = 1$.

Аналогично для высказывания <температура_воды> is <высокая> будем использовать функцию $L_2(t)$ и для <температура_воды> is <низкая> – функцию $H_2(t)$.

$$L_2(t) = 1/(1 + \exp(b_2(t - c_2))),$$

$$H_2(t) = 1/(1 + \exp(-b_2(t - c_2))),$$

причем $L_2(t) + H_2(t) = 1$.

Для высказывания <загрязненность> is <высокая> определим функцию $L_3(z)$ и для <загрязненность> is <низкая> – функцию $H_3(z)$.

$$L_3(x) = 1/(1 + \exp(b_3(z - c_3))),$$

$$H_3(x) = 1/(1 + \exp(-b_3(z - c_3))),$$

$$L_3(z) + H_3(z) = 1.$$

Для выходов <длительность> is <высокая> и <длительность> is <малая> аналогично определим функции

$$L_4(y) = 1/(1 + \exp(b_4(y - c_4))),$$

$$H_4(y) = 1/(1 + \exp(-b_4(y - c_4))),$$

$$\text{то есть } L_4(x) + H_4(x) = 1.$$

Для четких значений <количество_белья>, <температура_воды> и <загрязненность>: A_1, A_2, A_3 определим релевантность (силу) правил α :

$$\alpha_1 = H_1 \wedge H_2 \wedge H_3,$$

$$\alpha_2 = H_1 \wedge H_2 \wedge L_3,$$

$$\alpha_3 = L_1 \wedge L_2 \wedge L_3.$$

Выходы по каждому из правил определяется с помощью обратных функций принадлежности правых частей правил.

$$Y_1 = H_4^{-1}(\alpha_1),$$

$$Y_2 = H_4^{-1}(\alpha_2),$$

$$Y_3 = L_4^{-1}(\alpha_3).$$

Общий выход из системы нечетких правил определяется как

$$y_0 = (\alpha_1 \times y_1 + \alpha_2 \times y_2 + \alpha_3 \times y_3) / (\alpha_1 + \alpha_2 + \alpha_3).$$

Далее построим нечеткую нейронную сеть, идентичную системе нечеткого вывода, и обучим функции принадлежности анцедента и консеквента правил (рисунок 42).

Слой 1. Выходы узлов – это степени, в которых заданные входы удовлетворяют функциям принадлежности, ассоциированным с этими узлами.

Слой 2. Каждый узел вычисляет силу правила. Выход верхнего нейрона $\alpha_1 = H_1 \wedge H_2 \wedge H_3$, выход среднего нейрона $\alpha_2 = H_1 \wedge H_2 \wedge L_3$, а выход нижнего – $\alpha_3 = L_1 \wedge L_2 \wedge L_3$. Все узлы помечены Т, так как

можно выбрать любую T-норму для моделирования логического И. Узлы этого слоя называются узлами правил.

Слой 3. Каждый узел помечен N, чтобы показать, что узлы нормализуют силу правил $\beta_i = \alpha_i / (\alpha_1 + \alpha_2 + \alpha_3)$.

Слой 4. Выход нейронов – это произведение нормализованной силы правила и индивидуального выхода соответствующего правила.

$$\beta_1 Y_1 = \beta_1 H_4^{-1}(\alpha_1),$$

$$\beta_2 Y_2 = \beta_2 H_4^{-1}(\alpha_2), \quad \beta_3 Y_3 = \beta_3 L_4^{-1}(\alpha_3).$$

Слой 5. Одиночный выходной нейрон вычисляет выход сети

$$y_0 = \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 Y_3.$$

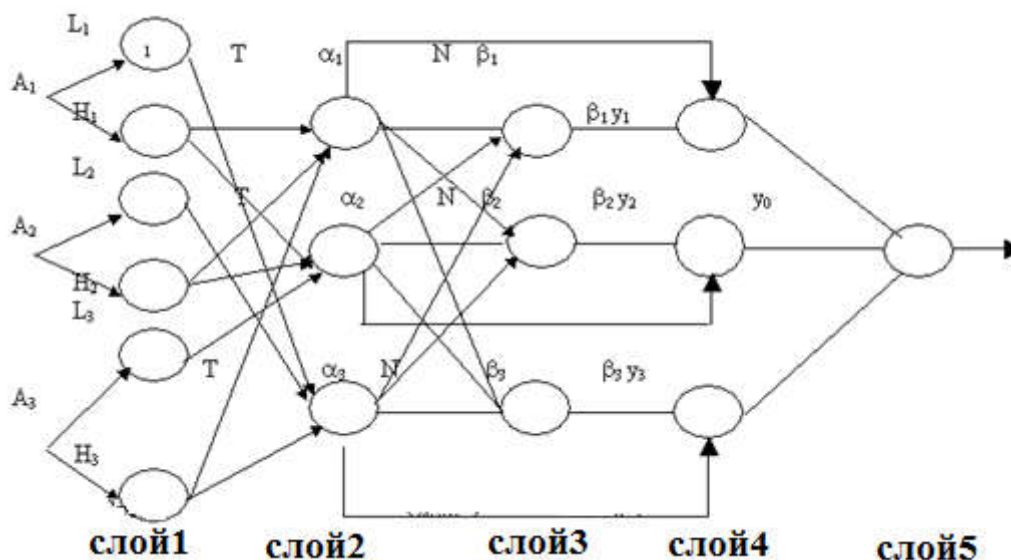


Рисунок 42. Пример нечеткой нейронной сети

Алгоритм обучения для нечеткой нейронной сети примера

Пусть задана четкая обучающая выборка

$\{(x_1, t_1, z_1, y_1), \dots, (x_k, t_k, z_k, y_k)\}$, где x, t, z – входные условия количества

белья, температуры воды и загрязненности, а y – длительность стирки. Ошибку на k -м образце определим как обычно:

$$E^k = 1/2(O_i - Y_i)^2.$$

Используем традиционный градиентный метод для обучения параметров левой и правой частей нечетких правил. Покажем, как можно настроить параметры функции формы.

$$b_4(t+1) = b_4(t) - \zeta \times (\partial E / \partial b_4),$$

$$c_1(t+1) = c_1(t) - \zeta \times (\partial E / \partial c_1)$$

.....

$$c_4(t+1) = c_4(t) - \zeta \times (\partial E / \partial c_4).$$

Используя данные соотношения как правила изменения весов в алгоритме обратного распространения ошибки, можно настроить параметры функций принадлежности в ходе обучения нечеткой нейронной сети.

Генетические алгоритмы

Генетические алгоритмы представляют собой адаптивные методы поиска, используемые для решения задач функциональной оптимизации. В их основе лежит идея природной эволюции, когда популяции развиваются в течение нескольких поколений и подчиняются законам естественного отбора. Лучшая особь выбирается по принципу «выживает наиболее приспособленный». Как мы помним, он был открыт Чарльзом Дарвином. Согласно этим принципам генетические алгоритмы способны «развивать» решения реальных задач при соответствующем формальном описании. Причем в отличие от эволюции, генетические алгоритмы моделируют лишь существенные для развития процессы.

В природе действует такой механизм, что наиболее приспособленные особи имеют больше шансов на воспроизведения потомства. Причем комбинация наиболее хороших характеристик может привести к появлению максимально приспособленного потомка. Генетические алгоритмы используют прямую аналогию с этим механизмом. Множество возможных решений проблемы – это популяция. Каждое решение оценивается по степени его «приспособленности» для решения поставленной задачи.

Сфера применения генетических алгоритмов весьма обширна. Они могут использоваться при создании таких вычислительных структур, как автоматы или сети сортировки. В машинном обучении их

можно использовать для проектирования нейронных сетей или в управлении роботами. Их можно использовать для моделирования процессов в различных областях (биологические, социальные, когнитивные системы). Однако наиболее популярное приложение – оптимизация многопараметрических функций, когда нужно найти оптимальное значение, зависящее от некоторых входных параметров.

В своей работе генетический алгоритм работает с хромосомами, которые состоят из генов. Чаще всего ген – это 0 или 1, говорящая о включении или не включении признака, характеризующего решение, в хромосому. Соответственно, хромосома – это битовая строка, описывающая решение. Но иногда возможны и другие кодировки в зависимости от задачи.

При работе с генетическим алгоритмом нам необходимо определить следующее:

Кодировку хромосомы (что будет представлять собой ген, и как гены будут участвовать в характеристике решения).

Пространство гипотез (популяцию), из которых мы должны выбрать лучшую.

Функцию приспособленности, оценивающую хромосомы.

Набор и вид генетических операций (скрещивание, мутацию).

Критерий остановки алгоритма (либо желаемое оптимальное значение, либо количество шагов эволюции популяции).

Для примера рассмотрим решение задачи с помощью генетического алгоритма. Формулировка задачи: необходимо составить наиболее сбалансированный рацион питания, удовлетворяющий определенным медицинским требованиям. Известен перечень продуктов из N наименований, каждый из которых имеет такие характеристики, как содержание жиров, белков, углеводов, калорий, а также известны рамки допустимого их содержания. Подходящим считается рацион, который лучше всего удовлетворяет определенным медицинским требованиям. Например, если рекомендуемое содержание жиров – 50, белков – 60, углеводов – 70, клетчатки – 80, витамина В – 90, витамина С – 100, то лучшим будет считаться рацион, чье суммарное

отклонение минимально, но чья десятая часть, например, укладывается в рамки от 40 до 80.

Первое, что нам необходимо сделать, – выбрать кодировку хромосомы. Так как основная наша задача – подобрать список продуктов, то хромосомой может быть весь перечень из N продуктов, а 0 или 1 в ней будут говорить о включении или не включении продуктов в рацион. Для лучшего понимания рассмотрим конкретный пример. Пусть у нас есть 4 продукта, каждый из которых можно описать набором из 6 числовых параметров-характеристик. Тогда по сути получим массив чисел:

Список_объектов={Объект₁={20,30,40,50,60,70},
 Объект₂={40, 40, 40, 40, 40, 40},
 Объект₃={30, 30, 30, 30, 30, 30},
 Объект₄={20, 30, 40, 40, 30, 30}}.

Тогда хромосома может иметь вид:

1010

Это говорит о том, что в рацион включаются первый и третий объекты. Тогда популяция может иметь вид:

0100

1010

1100

0101

Теперь определим функцию приспособленности для оценки каждой хромосомы в нашей задаче. Пусть у нас есть набор эталонных значений для каждой из характеристик объектов:

Эталон= {50,60,70,80,90,100}.

Тогда функция приспособленности будет иметь вид:

$$F = \sum_{j=1}^N \left(\frac{\sum_{i=1}^m \text{Эталон}_i - \text{Объект}_{ij}}{10} \right) \times \text{Hrom}_j ,$$

где Hrom_j – соответствующий ген в хромосоме.

Для оценки приспособленности хромосомы необходимо также проверить, укладывается ли значение функции в рамки от 40 до 80 (добавлением соответствующих условий), а для выявления наиболее приспособленной особи необходимо найти минимальную подходящую F в популяции.

Обратите внимание!

- *Определение вида функции приспособленности – чрезвычайно важная задача, потому что от правильности ее определения будет во многом зависеть успешность решения.*

Теперь рассмотрим генетические операции мутации и скрещивания. Если говорить формально, то мутация – это изменение одного или нескольких генов хромосомы вследствие случайного влияния. В контексте нашей задачи это может быть принудительное включение некоего продукта в набор, изменение значения вхождения продукта в набор на противоположное, принудительное исключение некоего продукта из набора и т. д. То есть, по сути, мутация – это изменение значения одного или нескольких генов хромосомы на противоположный или четко заданный. Вид и критерий применения мутации выбираются эмпирически. Допустим, для нашей задачи мы решили, что мутация будет изменять случайный ген на противоположный. Тогда, если исходная хромосома имела вид 0010, то мутировавшая может быть такой: 1010.

Операция скрещивания, или кроссовер, – операция, которая получает из двух хромосом одну, используя заданную маску. По сути из каждой хромосомы «вырезается» кусок, который помещается в новую. Существует несколько видов кроссовера. Одноточечный кроссовер: «разрез» хромосомы происходит только в одной точке, и новая особь получается путем соединения первой части первой хромосомы и второй части второй хромосомы. Двухточечный кроссовер: есть две точки «разреза», и новая хромосома получается из двух частей первой хромосомы и одной части второй.

Одноточечный кроссовер:

Хромосома₁ = 1001, Хромосома₂ = 1100. Точка разреза = 2. Маска потомка: 1 часть_Хромосома₁ + 2 часть_Хромосома₂. Потомок = (10)(00).

Двухточечный кроссовер:

Хромосома₁ = 1001, Хромосома₂=1100. Точки разреза = 1,3.
Маска потомка: 1часть_Хромосома₁ + 2часть_Хромосома₂ +
+3часть_Хромосома₁. Потомок=(1)(10)(1).

Теперь разберемся с таким вопросом, как отбор хромосом для воспроизведения потомства и выживание в популяции. Обычно в популяцию выбирается с наиболее приспособленных особей, которые и дают в ней потомство. Для того чтобы выбрать хромосомы в популяцию, можно использовать разные методы: турнирный, ранговый или метод рулетки. Рассмотрим эти методы подробнее.

Турнирный метод: задаем некую фиксированную вероятность выживаемости хромосомы, обозначив ее p . Случайно выбираем две особи. С вероятностью p выживает наиболее приспособленная, а $1-p$ – менее приспособленная.

Метод рулетки: вычисляем удельную приспособленность каждой особи относительно суммарной приспособленности популяции. Эту величину используем в качестве значения вероятности выживания.

Ранговый метод: выживает с наиболее приспособленных особей.

Таким образом, общую схему генетического алгоритма можно описать так:

Генерация начальной популяции из N особей.

Оценка приспособленности особей.

Пока не сработало условие выхода, делаем следующее:

 Выберем с особей для новой популяции и кроссовера.

 Выполним кроссовер и мутацию.

 Оценим приспособленность итоговой популяции.

Нечеткие системы с генетической настройкой

Настройка функций принадлежности с помощью нейронной сети или генетического алгоритма (ГА) устраняет принципиальную слабость теории нечетких систем – субъективность функций принадлежности. Применение нейронных сетей к настройке функций

принадлежности позволяет рассматривать окончательную форму функции как аппроксимацию обучающей выборки. Такую же задачу можно решить с помощью ГА, как метода стохастической оптимизации. Генетической нечеткой системой называют нечеткую систему, функции принадлежности и база правил которой спроектирована с помощью генетического алгоритма.

Применить ГА – это значит выбрать единицу кодирования, то есть хромосому; уточнить эволюционные операторы рекомбинации, мутации и селекции; сформировать функцию адаптивности (fitness-function, performanceindex). В настоящее время ГА используют либо для настройки функций принадлежности (базы данных), либо для формирования базы правил, либо для одновременного формирования и функций принадлежности и правил (а именно, базы знаний). В соответствии с объектами оптимизации выбирают единицу кодирования – хромосому. Для настройки функций принадлежности (ФП) за хромосому выбирают одно правило (Мичиганский подход), для настройки базы правил за хромосому выбирают вариант базы правил (Питтсбургский подход, подход итеративного обучения правил). В соответствии с кодированием уточняются правила генерации новых хромосом. Функция адаптивности представляет собой механизм нечеткого вывода, который для каждого варианта базы правил строит либо управление для нечеткого контроллера, либо экспертное заключение для диагностической экспертизы. Как видно из механизма нечеткого вывода, все нечеткие правила вносят вклад в окончательный результат, то есть правила сотрудничают. Но при отборе правил (хромосом) для генерации новых правил ГА накапливает правила, внесшие максимальный вклад в общий результат, то есть хромосомы конкурируют. В этом случае имеет место проблема «конкуренции и кооперации» в генетических нечетких системах (acompetition vs. a cooperation). Решение проблемы в каждом конкретном случае генетической нечеткой системы (ГНС) строится эвристически, например, при подходе итеративного обучения правил используют два этапа оптимизации. На первом шаге правила конку-

рируют за право войти в базу правил, а на втором – взаимодействуют при формировании общего результата.

Нечеткие нейронные сети с генетическим проектированием

Рассмотрим возможности генетических вычислений как средства структурной оптимизации нечетких нейронных сетей. Генетические вычисления можно применить на этапе проектирования нейронной сети. Возможности нейронных сетей интерполировать значения временных рядов могут быть широко использованы для оценки поведения макроэкономических показателей, в том числе индексов деловой активности или уровня ценных бумаг. Задача построения нейронного предиктора связана с принятием решений во многих точках проектирования. Необходимо исследовать входные данные и решить, по какому отрезку входных данных рационально делать предсказания, сколько точек в будущем разумно предсказать. Пространство перебора решений организации входных и выходных данных огромно для развитых рынков ценных бумаг, накопивших статистические сведения за десятки лет. Следующей точкой решения служит выбор типа нейронной сети: многослойный персептрон, радиально базисная сеть, вероятностная нейронная сеть, сеть регрессии и т. д. Для выбранного типа НС необходимо определить количество скрытых слоев, количество нейронов в них, виды функций активации и другие. Для каждой сети необходимо выбрать алгоритм обучения и его параметры, например, использование моментов (уровень моментов), уровень обученности, целевой уровень накопленной ошибки и т. д. Для нечеткой нейронной сети необходимо определить архитектуру сети, параметры функций принадлежности. В результате пространство решений при формировании нейронной сети становится необозримым для исследователя, и целесообразно применить ГА как средство эволюционного проектирования.

Генетическая оптимизация F-преобразования временных рядов

Задачи переборного типа с меньшими затратами могут быть решены путем применения эволюционных алгоритмов. Предлагается следующий классический алгоритм генетической оптимизации:

Имеется функция $F(x_1, \dots, x_4)$ – оценка построения прогноза тренда (должна быть минимизирована), где

x_1 – степень авторегрессии при построении прогноза тренда;

x_2 – метод, которым производится прогноз;

x_3 – количество точек, покрываемых базисной функцией;

x_4 – прогноз на основании истории из предыдущих N точек.

Для выявления сезонности следует строить прогноз, отодвигая историю на t точек назад:

$$F_k = \alpha F_{k-t} + \beta F_{k-t-1} + \dots$$

Алгоритм генетической оптимизации

1. Хромосомы имеют битовое представление (кодируем в коде Грея для получения отличия соседних хромосом в 1 бите). При этом в начале производится нормировка значений параметров к интервалу $[0, 1]$.

2. Оператор скрещивания (возможен случайный выбор между данными вариантами оператора скрещивания).

3. Сформировать случайно начальную популяцию, состоящую из k особей $B_0 = \{A_1, A_2, \dots, A_k\}$. При этом k определить эмпирически.

4. Вычислить приспособленность каждой особи $F_{A_i} = \text{fit}(A_i)$, $i=1 \dots k$.

5. Выбрать двух особей A_c из популяции. $A_c = \text{Get}(B_t)$.

6. Произвести скрещивание (выбрав один из предложенных вариантов) и получить новую особь.

7. Произвести мутацию генов с некоторой вероятностью.

8. Оценить полученную особь функцией приспособленности и добавить в популяцию взамен худшего родителя (таким образом сохраняется количество особей популяции).
9. Выполнить пункты 5-8 m раз.
10. Увеличить счетчик эпох.
11. Проверить условие останова (предельное количество эпох или предельное количество особей одного генотипа в популяции).

РАЗРАБОТКА ПРИЛОЖЕНИЙ В СФЕРЕ МАШИННОГО ОБУЧЕНИЯ

Для реализации полноценного решения в сфере ПО необходимо учесть аппаратную и программную составляющую. Под аппаратной составляющей подразумевается следующее. Машинное обучение в ряде случаев требует подбор специального оборудования (или конфигурирование специальных серверов), рассчитанного на массивные вычислительные нагрузки или на тяжелые параллельные вычисления. Причем отличительной чертой можно назвать то, что мощное оборудование может потребоваться не только в производственной стадии проекта, но даже на ранних стадиях разработки для проведения экспериментов и соответствующего исследования моделей. Без мощного оборудования требуемые модели могут работать в 10-60 раз медленнее или просто не запуститься ввиду нехватки памяти (или видеопамяти).

В случае разработки встраиваемых решений может потребоваться сильная оптимизация кода и квалификация в написании кода под специальное оборудование (например, программируемые логические интегральные схемы). В случае разработки распределенных систем потребуются также применение специальных программных средств для организации распределенных вычислений.

Эти специализированные темы в данном пособии не рассматриваются.

Однако если речь не идет о разработке специализированного программно-аппаратного комплекса или встраиваемого решения, то наращивание вычислительных узлов на сегодняшний день является не столько инженерной проблемой сколько экономической.

Что же касается программной составляющей, то здесь стоит упомянуть следующее. На сегодняшний день огромное число различных алгоритмов уже реализовано в виде библиотек и пакетов, поэтому нет необходимости писать алгоритм нейронной сети или

алгоритм решающих деревьев. За исключением опять же специализированных задач или исследовательских проектов.

Теоретически разработку систем МО можно вести на любом языке программирования, но при отсутствии специализированных библиотек сложность разработки может возрасти. Однако необходимо учесть, что специализированные средства создания моделей МО могут быть сложно применимы при реализации полноценных приложений. Поэтому нередко используют комбинацию технологий, когда предобработку данных и саму модель реализуют на одном языке программирования, а затем работающий прототип встраивают в приложение, написанное с использованием другой технологии. Одна из возможных архитектур реализации представлена на рисунке 43.

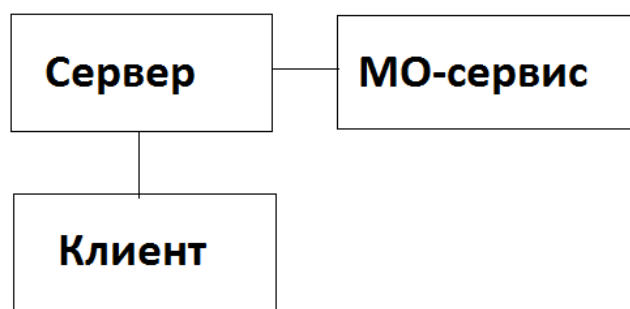


Рисунок 43. Возможная архитектура системы МО

Язык Python является одним из самых популярных языков для разработки в сфере машинного обучения и анализа данных ввиду очень простого синтаксиса, большого числа библиотек и развитого сообщества. Несмотря на то, что это интерпретируемый язык и выполнение кода на нем довольно медленное, большая часть библиотек реализована на быстрых языках типа Си, С++ и т. п. Поэтому на Python приходится писать лишь высокоуровневое обращение к API библиотек. Что делает разработку на Python удобной, а решения – не уступающими по скорости со специальной реализацией на Си. В силу вышеизложенного будем работать именно с Python.

Однако прежде чем перейти непосредственно к Python, необходимо сказать, что при разработке систем МО, как правило, приходится пройти несколько типовых этапов. Таким образом, можно сформулировать общий алгоритм создания систем МО, который может быть изменен в силу особенностей конкретного проекта.

Алгоритм состоит из следующих шагов:

1. Сбор и предобработка данных;
2. Выбор метода решения, создание и настройка модели(подбор параметров, реализация прототипа);
3. Тестирование модели и совершенствование ее до достижения необходимой точности. При невозможности достижения точности – возврат к шагу 2;
4. Сохранение модели и интеграция ее в оболочку взаимодействия с пользователем;
5. Эксплуатация и сопровождение полученной системы и ее модернизация (при необходимости).

Основы работы с Python

Теперь перейдем к рассмотрению возможностей, предоставляемых Python для разработки систем МО. Общие сведения об этом языке можно посмотреть здесь: [30], а справочную информацию по синтаксису – здесь: [31]. В нашей же книге мы рассмотрим лишь то, что касается возможностей Python в плане создания приложений в сфере машинного обучения. Все примеры кода подготовлены на Python версии 2.7, но они могут быть адаптированы и под более старшие версии языка.

Как было сказано выше, сейчас Python является одним из наиболее распространенных языков программирования. Одним из его преимуществ является большое количество пакетов, решающих самые разные задачи. В данном пособии мы рекомендуем использовать библиотеки Pandas, NumPy и SciPy, которые существенно упрощают чтение, хранение и обработку данных. Вы также познако-

митесь с пакетом Scikit-Learn, в котором реализованы многие алгоритмы машинного обучения.

Необходимо отметить, что основными отличиями Python являются:

- Отсутствие завершающих символов в конце строки (точек, запятых и т. п.), что делает написание линейных конструкций очень «приятным» и быстрым занятием (а большинство программ для машинного обучения – это все-таки линейная математика, а не сложные многоуровневые системы).
- Нестрогая типизация. Не нужно объявлять тип данных при объявлении переменной, что опять же ускоряет процесс разработки модели.
- Язык Python интерпретируемый, кросс-платформенный и обладает хорошими средствами отладки.

Для разработки приложений нам нужно следующее:

- Язык, среда для разработки кода (IDE) и исполняемая среда. В нашем случае это Anaconda и PyCharm.
- Набор основных библиотек: Scikit-learn, Numpy, Pandas, Matplotlib, Theano, Keras.

Ниже представлены инструкции по установке и настройке компонент. Важный момент заключается в том, что все компоненты для работы в своем исходном состоянии «не родные» для систем Windows, поэтому в первую очередь эти инструкции относятся к пользователям Windows.

Инструкции по установке основных компонент:

1. Скачайте и установите AnacondaEnvironment (среда свободно распространяется на официальном сайте [32]). Anaconda необходима по большей части не как IDE, а как среда, в которой будет сразу установлено множество необходимых библиотек (развернутых соответствующим образом под Windows), таких как pip, numpy, matplotlib и еще пара десятков других. Также будет установлен и сам Python 2.7.

Обратите внимание!

- *Для корректной работы всех компонент аккаунт пользователя Windows, из под которого будет проводиться установка, должен содержать только латинские символы в своем имени.*

2. Рекомендуется установить PyCharm как основную IDE для разработки (хотя можно пользоваться самой SpiderAnaconda, которая будет установлена на предыдущем шаге) или JupyterNotebook. Однако если вы новичок в программировании или в Python, то рекомендуется именно PyCharm, так как эта среда сильно облегчает программирование и навигацию по коду. CommunityEdition распространяется бесплатно и скачать его можно здесь: [33].

3. Для проверки того, что все компоненты установлены корректно, запустите IDE, создайте новый файл (команда Alt+Insert или через меню File и пункт New) под именем testc кодом на Python, как показано на рисунке 44. Вставьте в файл код из листинга 1 и выполните его через пункт меню Run, как показано на рисунке 45.

Листинг 1

```
from sklearn import preprocessing
import numpy as np
X = np.array([[ 1., -1., 2.], [ 2., 0., 0.], [ 0., 1., -1.]])
X_scaled = preprocessing.scale(X)
print(X_scaled)
```

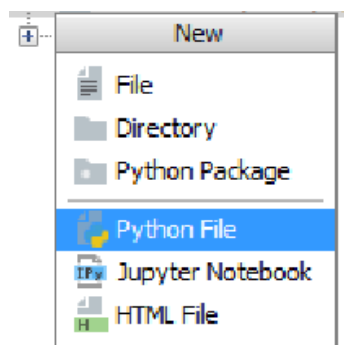


Рисунок 44. Новый файл Python

Обратите внимание!

- Если в первом пункте меню *Run* нет имени вашего файла, то выберите третий пункт и в открывшемся окошке найдите имя вашего файла.

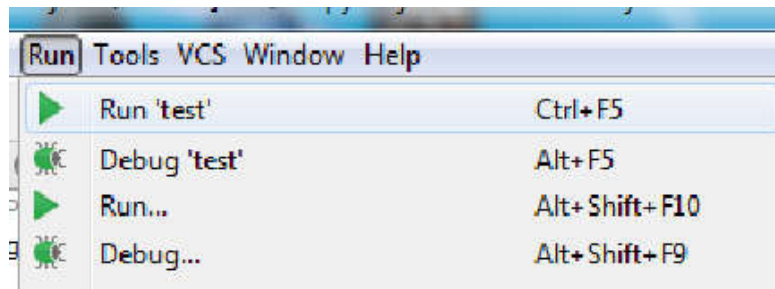


Рисунок 45. Запуск кода на выполнение

Если все установлено верно, то в консоли вы должны увидеть результат, представленный на рисунке 46.

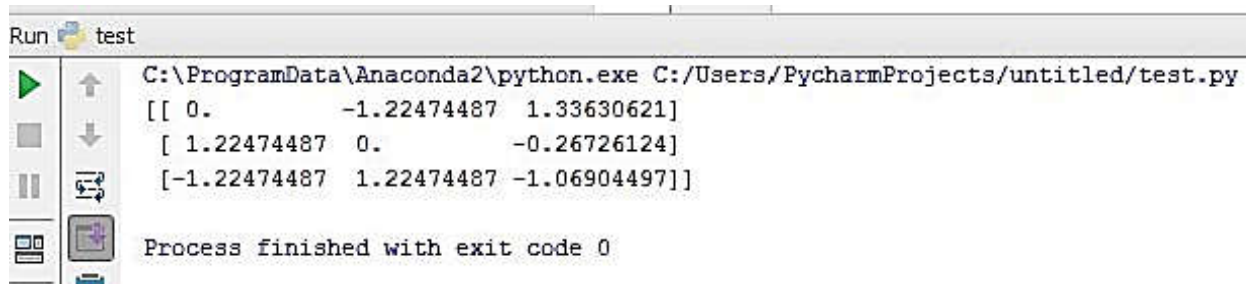


Рисунок 46. Корректная работа кода

Кроме описанного выше, нам понадобится установить и настроить еще ряд компонентов и библиотек. Инструкции по их установке следующие:

1. Скачайте и установите TDM-GCC (специальное средство компиляции для ОС Windows);
2. Откройте консоль Anacondaprompt или обычную командную строку Windows через команду `cmd`;
3. Выполните в консоли `conda update conda`;
4. Выполните в консоли `conda update --all`;
5. Выполните в консоли `conda install mingw libpython`;

6. Выполните в консоли `conda install Theano`;

7. Выполните в консоли `pip install keras`;

Обратите внимание!

- *Keras может быть настроен на tensorflowbackend, тогда надо внести изменения в файл `C:/Users/AccountName/.keras/keras.json` и в строке `backend="tensorflow"` написать `theano`;*

8. Запустите IDE, создайте новый файл с кодом на Python и выполните код из листинга 2 для проверки того, что все компоненты установлены корректно.

Листинг 2

```
import numpy as np
import sklearn.linear_model as lin
from keras.layers.core import Dense, Activation
from keras.models import Sequential
from keras.optimizers import RMSprop

x_train = np.array([[0.1, 0.3], [0.2, 0.2], [0.7, 0.8], [1.0, 0.9]])
y_train = np.array([0, 0, 1, 1]).reshape(4, 1)
x_test = np.array([[0, 0], [0.3, 0.3], [0.6, 0.7], [1, 1]])
model = Sequential()
model.add(Dense(2, init='lecun_uniform', input_shape=(2,)))
model.add(Activation('relu'))
model.add(Dense(1, init='lecun_uniform'))
model.add(Activation('linear'))
rms = RMSprop(lr=0.01)
model.compile(loss='mse', optimizer=rms)
epochs = 200
model.fit(x_train, y_train, batch_size=1, nb_epoch=epochs, verbose=0)
y_predict = model.predict(x_test, batch_size=1)
print(y_predict)
```

В результате в Output консоли не должно быть ошибок (код выхода равен 0), и должна будет распечататься информация, представленная в листинге 3.

Листинг 3

```
[[ 0.00339771]
 [ 0.14396997]
 [ 0.67507285]
 [ 1.1803354 ]]
```

Мы установили основные, необходимые нам, компоненты. Теперь перейдем непосредственно к рассмотрению конкретных приемов работы с ними.

В данном пособии будет приведено описание только тех операторов и функций, которые непосредственно используются при решении конкретной задачи. Для получения прочей информации по Python рекомендуется воспользоваться справочником или же поисковой системой Google.

Обратите внимание!

- *Если у вас есть конкретный вопрос, например, о том, как выбрать последний элемент в одномерном или двумерном массиве, или о том, как отсортировать массив и т. п., то лучшее решение – написать этот вопрос в Google на английском;*
- *Если ваш уровень английского недостаточен, то просто сформулируйте вопрос на русском максимально коротко, затем вставьте в GoogleTranslate и затем – в поисковик. С вероятностью 90% на такие конкретные вопросы вы найдете очень конкретные примеры кода на сайте StackOverflow.*

Элементарные операции с данными

Рассмотрим работу по преобразовке данных и поиску простых закономерностей в них средствами Python, а именно средствами пакетов Numpy и Pandas. Более подробно с их функциями можно ознакомиться, например, здесь: [34].

Для импорта модуля NumPy необходимо написать следующую строку кода (листинг 4).

Листинг 4

```
import numpy as np
```

Обратите внимание!

- *NumPy импортируется с псевдонимом np, через который в дальнейшем будет обращение к модулю.*

Основными единицами данных в машинном обучении являются векторы и матрицы. С точки зрения программирования такие структуры представляют собой одномерные и двумерные массивы или специальные объекты-таблицы.

Обратите внимание!

- *Если в тексте написано просто – вектор или матрица, то значит идет речь о одномерном или двумерном массиве NumPy. В других случаях будет специально написано в каком формате представлены данные (например, DataFrame, о чем речь пойдет ниже).*

Для того чтобы рассмотреть способы работы с данными, нам необходимо получить данные для работы. Так как модели, разрабатываемые на Python, в 99% случаев используют полученные откуда-то данные или же генерируют тестовые, то мы опишем три способа получения данных: прямое объявление, случайная генерация и загрузка из csv-формата.

Начнем с первого. Объявить матрицу можно, используя код из листинга 5. Как видите, ничего сложного здесь нет: ни предварительного объявления данных, ни выделения памяти, ни объявления типа данных.

Листинг 5

```
Z = np.array([[4, 5, 0],  
             [9, 9, 9]])
```

Теперь рассмотрим второй способ: сгенерируем случайную матрицу, состоящую из 6 строк и 5 столбцов. Элементы ее будут являться случайными числами из нормального распределения. Функция для генерации таких чисел: `np.random.normal`.

Ее параметры:

- `loc`: среднее нормального распределения (в нашем случае 1);
- `scale`: стандартное отклонение нормального распределения (его значение в нашем случае равно 10);
- `size`: размер матрицы (в нашем случае (6, 5)).

Код генерации матрицы `X` и ее распечатки представлены в листинге 6.

Листинг 6

```
X = np.random.normal(loc=1, scale=10, size=(6, 5))  
print X
```

Далее рассмотрим способ загрузки данных из файла в csv-формате, который предназначен для хранения табличных данных. Как правило, в файлах этого формата столбцы разделяются запятой, а первая строка содержит их имена.

Допустим, у нас есть файл «titanic.csv», содержащий данные в виде, представленном на рисунке 47.

Обратите внимание!

- *Небольшие файлы csv (до 1 Гб) можно достаточно удобно посмотреть в Excel. Чтобы корректно загрузить CSV файл, нужно сначала открыть Excel, создать пустую книгу, а затем вызвать соответствующего мастера загрузки через вкладку «Данные»=>«Из текста»..*

A	B	C	D	E	F	G	H	I	J	K	L
PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, M	male	22	1	0	A/5 21171	июл.25		S
2	1	1	Cumings, I	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, N	female	35	1	0	113803	53.1	C123	S

Рисунок 47. Файл с данными

Для загрузки этих данных нам необходимы средства библиотеки pandas. Код ее импорта, а также код загрузки и распечатки загруженных данных представлены в листинге 7. Там же представлен код записи данных в csv-файл.

Листинг 7

```
# -*- coding: utf-8 -*-
import csv
import pandas
#чтение из файла
data = pandas.read_csv('titanic.csv', index_col='PassengerId')
print data
#запись в файл данных в исходном виде
data.to_csv('q.csv')
#запись в файл данных как массива значений
with open('test.csv', 'w') as csvfile:
    spamwriter = csv.writer(csvfile, delimiter=',', quotechar='|', quoting=csv.QUOTE_MINIMAL)
    spamwriter.writerow (data.as_matrix())
```

Обратите внимание!

- Символ # - знак комментария.
- Для того чтобы закомментировать (или раскомментировать) несколько строк, необходимо выделить их и нажать сочетание клавиш Ctrl+правый слеш.
- Для того чтобы писать комментарии в Python на русском языке, необходимо добавить следующую строку в начало файла с кодом: # -*- coding: utf-8 -*-

При выполнении кода из листинга 7 данные будут загружены в виде DataFrame, с помощью которого можно удобно работать с ними. Параметр `index_col='PassengerId'` означает, что колонка `PassengerId` задает нумерацию строк данного DataFrame.

DataFrame – первичная структура данных pandas, представляющая собой двумерную, изменяемую по размеру, потенциально гетерогенную структуру табличных данных с маркированными осями (строками и столбцами).

В виде DataFrame над данными удобно производить различные манипуляции сортировки, выборки, применения функций построчно или по колонкам и т. п. Однако DataFrame – это комплексный объект высокого уровня и его нельзя использовать напрямую для обучения моделей. Перед этим его нужно перевести в обычную цифровую матрицу, например, следующим образом (листинг 8), и дальше работать с ним как с обычным двумерным массивом данных.

Листинг 8

```
x_test= data[['ColumnName_1', 'ColumnName_2']].as_matrix()
```

Вернемся к листингу 7. В нем представлены два способа записи в файл. Первый – через метод самого DataFrame, второй – через метод библиотеки csv. Необходимо отметить, что второй способ может быть использован не только для работы с DataFrame, но и с любыми данными, которые надо записать в csv-файл.

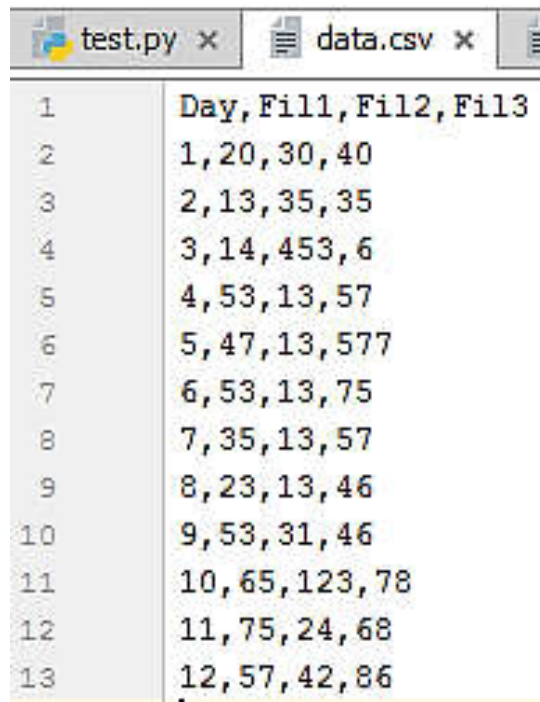
Мы рассмотрели основные способы получения данных, теперь перейдем к работе с ними и продемонстрируем некоторые возможности библиотеки Numpy по операциям с матрицами на конкретных задачах из практики.

Допустим, перед нами стоит следующая задача: есть данные о ежедневной выручке по филиалам компании за первые 12 дней месяца, записанные в виде таблицы (рисунок 48) и сохраненные в файле «data.csv». Выведем номер филиала, преодолевшего по прибыли порог, равный 1000.

Первое, что нам необходимо сделать, – загрузить данные. Однако, чтобы pandas корректно обработал файл, нам нужно убрать из него кириллицу и убедиться, что разделители в csv действительно запятые. Для этого мы можем поместить файл в папку проекта ruCharm, двойным щелчком мыши на нем (в окне проекта) открыть его и привести к виду, показанному на рисунке 49. Теперь для чтения данных, преобразованию их в матрицу и распечатки полученного массива мы можем использовать код из листинга 9.

A	B	C	D
День месяца	Филиал1	Филиал2	Филиал3
1	20	30	40
2	13	35	35
3	14	453	6
4	53	13	57
5	47	13	577
6	53	13	75
7	35	13	57
8	23	13	46
9	53	31	46
10	65	123	78
11	75	24	68
12	57	42	86

Рисунок 48. Файл с данными для задачи



```
test.py x data.csv x
1 Day, Fil1, Fil2, Fil3
2 1, 20, 30, 40
3 2, 13, 35, 35
4 3, 14, 453, 6
5 4, 53, 13, 57
6 5, 47, 13, 577
7 6, 53, 13, 75
8 7, 35, 13, 57
9 8, 23, 13, 46
10 9, 53, 31, 46
11 10, 65, 123, 78
12 11, 75, 24, 68
13 12, 57, 42, 86
```

Рисунок 49. Корректный файл с данными

Обратите внимание!

- Если исходный файл создан через Excel, то разделителями могут быть точки с запятой. Заменить их на запятые можно при редактировании файла в PyCharm через цепочку пунктов меню «Edit»=>«Find» =>«Replace».

Листинг 9

```
import pandas
data= pandas.read_csv('data.csv', index_col='Day')
x_test= data[['Fil1', 'Fil2', 'Fil3']].as_matrix()
print x_test
```

Если все сделано верно, в консоли должен появиться результат, показанный на рисунке 50.

```
test
C:\ProgramData\Ar
[[ 20  30  40]
 [ 13  35  35]
 [ 14 453   6]
 [ 53  13  57]
 [ 47  13 577]
 [ 53  13  75]
 [ 35  13  57]
 [ 23  13  46]
 ---  --  ---
```

Рисунок 50. Загруженные данные

Обратите внимание!

- При преобразовании к матрице мы указываем в кавычках имена столбцов с данными по филиалам и не указываем столбец Day, так как содержащиеся в нем данные для задачи не нужны.

Теперь перейдем непосредственно к решению поставленной задачи. С точки зрения работы с матрицей, для получения ответа нам

необходимо посчитать сумму по каждому столбцу и вывести номера тех, чья сумма превысит 1000.

Функция для подсчета суммы: `np.sum`. В качестве параметров она принимает матрицу, для которой необходимо посчитать сумму и измерение (строки или столбцы), которое необходимо суммировать. Измерение (`axis`) задается цифрой (для двумерной матрицы 0 – строки, 1 – столбцы). Если этот параметр не задать, то результат функции будет рассчитан для всей матрицы целиком. Результатом выполнения операции будет массив с соответствующими суммами. Библиотека `Numpy` предоставляет возможности применения к матрицам (и массивам) логических операций, причем применяемых поэлементно. Соответственно, результатом такой операции будет матрица такого же размера, в ячейках которой будет записано либо `True`, либо `False` (удовлетворяет текущий элемент условию или нет). Индексы элементов со значением `True` можно получить с помощью функции `np.nonzero`. Функция в качестве параметра принимает матрицу, в которой необходимо отыскать ненулевые элементы. Заметим, что в нашем случае мы можем сразу передать логическое выражение, составленное из массива сумм и самого условия, тогда элементы со значением `False` будут интерпретироваться как нулевые. Код решения представлен в листинге 10. Результат работы кода – на рисунке 51. Ответ к задаче: третий филиал.

Листинг 10

```
import numpy as np
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
x_test= data[['Fil1', 'Fil2', 'Fil3']].as_matrix()
#print x_test
r = np.sum(x_test, axis=0)
print (r)
print np.nonzero(r> 1000)
```

```
C:\ProgramData\Anaconda2\py
[ 508  803 1171]
(array([2], dtype=int64),)
```

Рисунок 51. Решение задачи

Теперь решим следующую задачу: определим, в какой день суммарная выручка по филиалам превысила 500. Код решения представлен в листинге 11, а результат – на рисунке 52.

```
C:\ProgramData\Anaconda2\python.exe C:/Users/Pychi
[ 90  83 473 123 637 141 105  82 130 266 167 185]
(array([4], dtype=int64),)
Process finished with exit code 0
```

Рисунок 52. Решение задачи

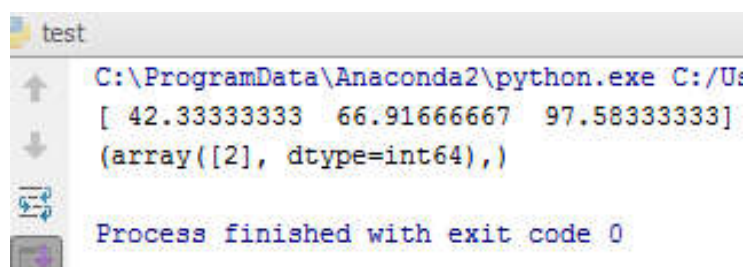
Листинг 11

```
import numpy as np
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
x_test= data[['Fil1', 'Fil2', 'Fil3']].as_matrix()
r = np.sum(x_test, axis=1)
print (r)
print np.nonzero(r>500)
```

Код нахождения филиала с максимальной средней выручкой представлен в листинге 12, а результат показан на рисунке 53.

Листинг 12

```
import numpy as np
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
x_test= data[['Fil1', 'Fil2', 'Fil3']].as_matrix()
r = np.mean(x_test, axis=0)
print (r)
print np.nonzero(r== np.max(np.mean(x_test, axis=0)))
```

```
test
C:\ProgramData\Anaconda2\python.exe C:/U:
[ 42.33333333 66.91666667 97.58333333]
(array([2], dtype=int64),)

Process finished with exit code 0
```

Рисунок 53. Номер филиала с максимальной средней выручкой

Номер филиала с максимальной величиной стандартного отклонения прибыли можно получить, используя код из листинга 13.

Далее нам необходимо более подробно рассмотреть работу со структурой DataFrame, но перед этим упомянем еще две полезные функции NumPy, которые могут понадобиться в последующем: функцию генерации единичной матрицы (`np.eye`) и функцию вертикальной стыковки матриц (`np.vstack`). Первая в качестве параметра принимает количество строк (оно же количество столбцов), вторая – матрицы, которые нужно объединить. В листинге 14 показан пример генерации и объединения двух единичных матриц.

Листинг 13

```
import numpy as np
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
x_test= data[['Fil1', 'Fil2', 'Fil3']].as_matrix()
r = np.std(x_test, axis=0)
print (r)
print np.nonzero(r== np.max(np.std(x_test, axis=0)))
```

Листинг 14

```
import numpy as np
A = np.eye(3)
B = np.eye(3)
print A
print B
AB = np.vstack((A, B))
```

Работа с DataFrame

Мы рассмотрели основные моменты работы с матрицами, поэтому теперь вернемся к другой структуре данных: DataFrame. Как мы помним, она является таблицей, где колонки имеют заголовки, а строки, кроме номеров, могут иметь дополнительные индексы в виде цифр или имен (по сути, индексы представляют собой отдельную, специальную колонку в таблице DataFrame, которая не относится к данным). Для того чтобы посмотреть, что представляют собой данные, можно воспользоваться несколькими способами.

Если мы хотим распечатать первые 5 строк данных, то можем воспользоваться кодом из листинга 15, в случае, если указан только один индекс, или же воспользоваться методом `head()` DataFrame, как показано в листинге 16.

Листинг 15

```
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
print data[:5]
```

Листинг 16

```
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
print data.head()
```

Если же нас интересует содержимое конкретного столбца, то можно использовать квадратные скобки и название столбца, как показано в листинге 17.

Листинг 17

```
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
print data['Fil1']
```

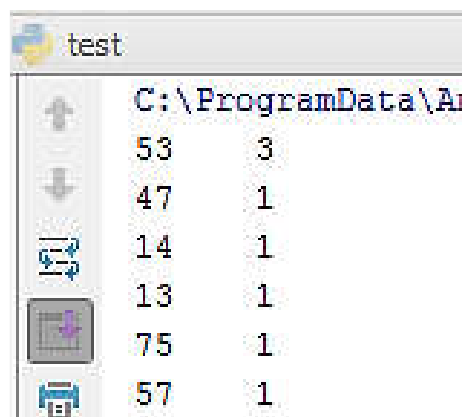
Обратите внимание!

- *Код из листинга 16 не применим к индексному столбцу.*

Для подсчета некоторых статистик (количества, среднее, максимум, минимум) можно также использовать методы объекта DataFrame. В листинге 18 приведен пример кода, считающего количество повторов каждого из значений в столбце с именем 'Fill'. Результат работы кода показан на рисунке 54. Значения суммы, минимума, максимума, среднего, количества и среднеквадратического отклонения вычисляются аналогично с использованием соответствующих функций (sum, min, max, mean, count, std).

Листинг 18

```
import pandas
data = pandas.read_csv('data.csv', index_col='Day')
print data['Fill'].value_counts()
```



53	3
47	1
14	1
13	1
75	1
57	1

Рисунок 54. Результат работы кода из листинга 18

Теперь рассмотрим еще несколько более сложных функций работы с DataFrame на примере некоторой таблицы, представленной на рисунке 55.

В листинге 19 представлены интересующие нас фрагменты кода с комментариями. Как можно заметить, возможности DataFrame весьма обширны, поэтому здесь приведены лишь основные моменты, а с остальным вам предлагается ознакомиться самостоятельно.

	first_name	company_name	address	city	county	postal	phone1	phone2
last_name								
Villamarin	Antonio	Combs Sheetmetal	353 Standish St #8264	Little Parndon and Hare Street	Hertfordshire	CM20 2HT	01559-403415	01388-777812
Glasford	Antonio	Saint Thomas Creations	425 Howley St	Gaer Community	Newport	NP20 3DE	01463-409090	01242-318420
Heilig	Antonio	Radisson Suite Hotel	35 Elton St #3	Ipplepen	Devon	TQ12 5LL	01324-171614	01442-946357

Рисунок 55. Структура таблицы с данными

Обратите внимание!

- Для лучшего понимания кода листинга 19 вам рекомендуется почитать про Лямбда функции в Python, например, здесь: [35];
- Более подробно со списком методов DataFrame можно познакомиться в документации: [36];
- Ссылка на уроки по Pandas: [37].

Листинг 19

```
# *- coding: utf-8 -*-
# Выбор элементов DataFrame с использованием iloc
# Строки:
data.iloc[0] # выбирает первую строку таблицы.
data.iloc[1] # выбирает вторую строку таблицы
data.iloc[-1] # выбирает последнюю строку таблицы

# Столбцы:
data.iloc[:,0] # выбирает первую колонку таблицы
data.iloc[:,1] # выбирает вторую колонку таблицы
data.iloc[:,-1] # выбирает последнюю колонку таблицы

# Множественный выбор
data.iloc[0:5] # выбирает первые пять строк таблицы
data.iloc[:, 0:2] # выбирает первые две колонки таблицы со всеми строками

# Выбирает строки с индексами 'Andrade' и 'Veness' (в случае текстовых индексов),
# со всеми колонками между колонок с именем 'city' и 'email'
data.loc[['Andrade', 'Veness'], 'city':'email']
```

Окончание листинга 19

```
# Выбирает те же строки, только с колонками 'first_name', 'address' и 'city'
data.loc['Andrade':'Veness', ['first_name', 'address', 'city']]

# Выбирает строки с first_name равным Antonio и выбирает только колонки между
# колонками 'city' и 'email'
data.loc[data['first_name'] == 'Antonio', 'city':'email']

# Выбирает строки со всеми колонками, где в колонке email встречаются ячейки,
# заканчивающиеся на 'hotmail.com'
data.loc[data['email'].str.endswith("hotmail.com")]

# Выбирает строки, где first_name равно одному из следующих значений
data.loc[data['first_name'].isin(['France', 'Tyisha', 'Eric'])]

# Выбирает строки, где в колонке first_name встречается Antonio и в колонке email
# встречается gmail.com
data.loc[data['email'].str.endswith("gmail.com") & (data['first_name'] == 'Antonio')]

# Выбирает строки, у которых в колонке id есть значения от 100 до 200, и возвращает
# только колонки 'postal' и 'web'
data.loc[(data['id'] > 100) & (data['id'] <= 200), ['postal', 'web']]

# Лямбда функция, которая возвращает True/False переменную.
# Выбирает строки, где ячейка в колонке company_name имеет 4 слова.
data.loc[data['company_name'].apply(lambda x: len(x.split(' ')) == 4)]

# Распечатка заголовков
print list(data)
```

Предобработка данных. Стандартизация и нормализация

Мы рассмотрели основные моменты работы с данными, теперь перейдем к такому вопросу, как предобработка данных. Прежде чем обучать и использовать большинство моделей, необходимо привести исходные данные к единообразному виду (к единому диапазону). Единый диапазон определяется через математическое ожидание выборки и разброс (дисперсию). То есть нужно, например, чтобы все данные были в диапазоне от 0 до 1 или от -1 до 1. Как было сказано

в одном из предыдущих разделов, достигается такое приведение посредством двух процессов\методов: стандартизации данных и нормализации.

Можно провести стандартизацию самостоятельно, используя NumPy. Для этого вычтите из каждого столбца его среднее значение, а затем поделите на его стандартное отклонение. Или же можно использовать библиотеку `scikit-learn`, как показано в листинге 20.

Однако, чаще всего, кроме нормировки одной матрицы, требуется нормировать несколько матриц по одним и тем же шкалам (для каждого столбца\признака будут свои коэффициенты смещения и масштабирования).

Так если тестовая\рабочая выборки поступают в систему онлайн, то их потребуется нормировать по тем же коэффициентам, что и тренировочную (рассчитанным на тренировочной выборке). Иначе нормировка будет некорректной.

Листинг 20

```
from sklearn import preprocessing
import numpy as np
X = np.random.normal(loc=1, scale=10, size=(6, 5))
X_scaled = preprocessing.scale(X)
print X
print X_scaled
```

Вместо того чтобы хранить матрицы коэффициентов нормировки, гораздо удобнее воспользоваться классом `StandardScaler` из `scikit-learn`. Сперва нужно «обучить» его на тренировочных данных (методом `fit`), а потом преобразовывать разные матрицы одинаковым образом. Работа с этим классом показана в листинге 21.

Листинг 21

```
from sklearn import preprocessing
import numpy as np
X = np.random.normal(loc=1, scale=10, size=(6, 5))
X_test = np.random.normal(loc=1, scale=10, size=(6, 5))
```

Окончание листинга 21

```
scaler = preprocessing.StandardScaler().fit(X)
train_data_scaled = scaler.transform(X)
test_data_scaled = scaler.transform(X_test)
print train_data_scaled
print test_data_scaled
```

Как было сказано в одном из предыдущих разделов, стандартизация смещает значения векторов (выравнивает относительно единого центра в нуле), а также производит выравнивание разброса. Однако значения разных векторов не будут в одинаковом диапазоне (от -1 до 1 , например). Они будут лишь иметь стандартный разброс в рамках вектора\столбца\признака. Для того чтобы достичь одинакового масштаба всех векторов, необходима нормализация.

Средствами `scikit-learn` нормализацию можно выполнить следующим образом (листинг 22).

Листинг 22

```
#-*- coding: utf-8 -*-
from sklearn import preprocessing
import numpy as np
X=[[1., -1.,2.],
   [2., 0., 0.],
   [0., 1., -1.]]

#нормализация max-нормой
X_normalized_max = preprocessing.normalize(X, norm='max', axis=0)

#нормализация нормой l1
X_normalized_max = preprocessing.normalize(X, norm='l1', axis=0)

#нормализация нормой l2
X_normalized_max = preprocessing.normalize(X, norm='l2', axis=0)
```

Более подробную информацию о методах библиотеки `scikit-learn` по преобработке данных можно найти здесь: [38].

Работа с деревьями решений

Мы рассмотрели методы предобработки данных. Теперь перейдем к следующему – к деревьям решений. Как было сказано в одном из предыдущих разделов, решающие деревья относятся к классу логических методов. Их основная идея состоит в объединении определенного количества простых решающих правил, благодаря чему итоговый алгоритм является интерпретируемым. Как следует из названия, решающее дерево представляет собой бинарное дерево, в котором каждой вершине сопоставлено некоторое правило вида «*j*-й признак имеет значение меньше *b*». В листьях этого дерева записаны числа-предсказания. Чтобы получить ответ, нужно стартовать из корня и делать переходы либо в левое, либо в правое поддерево в зависимости от того, выполняется правило из текущей вершины или не выполняется.

Одна из особенностей решающих деревьев заключается в том, что они позволяют получать важности всех используемых признаков. Важность признака можно оценить на основе того, как сильно улучшился критерий качества благодаря использованию этого признака в вершинах дерева.

Рассмотрим данные о пассажирах «Титаника» (их можно скачать здесь: [39]). Решим на них задачу классификации, в которой по различным характеристикам пассажиров требуется найти у выживших пассажиров два наиболее важных признака (из четырех рассматриваемых: пол, класс, возраст, цена билета).

В библиотеке `scikit-learn` решающие деревья реализованы в классах `sklearn.tree.DecisionTreeClassifier` (для классификации) и `sklearn.tree.DecisionTreeRegressor` (для регрессии). Обучение модели производится с помощью функции `fit`. Найти важность признаков можно, имея уже обученный классификатор: его поле `feature_importances_` содержит массив «важностей» признаков. Индекс в этом массиве соответствует индексу признака в данных. Стоит обратить внимание, что данные могут содержать пропуски.

Pandas выгружает такие значения как `nan` (not a number). Для того чтобы проверить, является ли число `nan`'ом, можно воспользоваться функцией `numpy.isnan`.

Перейдем к решению задачи. Первое, что нам необходимо сделать, – внимательно посмотреть на данные перед загрузкой. Ранее мы уже предположили, что среди них могут быть пропуски, и определились с методом решения этой проблемы, но при изучении информации мы должны увидеть еще одну: признак `Sex` имеет строковые значения, тогда как все остальные – числовые. Следовательно, нам необходимо привести и его к числовому виду. Например, заменим `male` на `0`, остальное – на `1`.

Для этого определим функцию, которую применим к нашему массиву данных. В Python функции начинаются с ключевого слова `def`, и их операторы (тело функции) обязательно имеют отступ от начала строки. Применение функции к объекту выполняется оператором `apply`. Код функции представлен в листинге 23.

Листинг 23

```
def Sex_to_bool(sex):
    if sex == "male":
        return 0
    return 1
```

Алгоритм решения задачи будет следующим:

1. Загрузить выборку из файла `titanic.csv` с помощью пакета `Pandas`.
2. Оставить в выборке четыре признака: класс пассажира (`Pclass`), цену билета (`Fare`), возраст пассажира (`Age`) и его пол (`Sex`). Привести пол к числовому виду и убрать из выборки пустые значения.
3. Выделить целевую переменную (она записана в столбце `Survived`).

4. Обучить решающее дерево с параметром `random_state=241` и остальными параметрами по умолчанию (речь идет о параметрах конструктора `DecisionTreeClassifier`).
5. Вывести важности признаков.

Код решения задачи (с комментариями) – в листинге 24. Результат работы – на рисунке 56.

PassengerId	Pclass	Fare	Age	Sex	Importance
1	3	7.2500	22.0	0	0.14751816
2	1	71.2833	38.0	1	0.29538468
3	3	7.9250	26.0	1	0.25658495
4	1	53.1000	35.0	1	0.30051221
5	3	8.0500	35.0	0	

Рисунок 56. Важности признаков

Листинг 24

```

-*- coding: utf-8 -*-
import pandas
from sklearn.tree import DecisionTreeClassifier
import numpy as np

data = pandas.read_csv('titanic.csv', index_col='PassengerId')
#функция для приведения пола к числу
def Sex_to_bool(sex):
    if sex == "male":
        return 0
    return 1
#приведение пола к числу
data['Sex'] = data['Sex'].apply(Sex_to_bool)
#выгрузка непустых данных
data=data.loc[(np.isnan(data['Pclass'])==False) & (np.isnan(data['Fare'])==False)
&(np.isnan(data['Age'])==False) & (np.isnan(data['Sex'])==False)&
(np.isnan(data['Survived'])==False)]
#отбор нужных столбцов
corr = data[['Pclass', 'Fare', 'Age', 'Sex']]
#респечатка первых 5 строк данных
print corr.head()

```

Окончание листинга 24

```
#определение целевой переменной
y = data['Survived']
#создание и обучение дерева решений
clf = DecisionTreeClassifier(random_state=241)
clf.fit(corr, y)
#получение и распечатка важностей признаков
importances=clf.feature_importances_
print importances
```

Таким образом мы видим, что наиболее важными признаками будут пол и цена билета. Однако по описанному решению стоит сделать одно важное замечание. При преобразовании пола к числовому виду мы исказили характер данных: ранее объекты столбца «пол» не могли быть математически сравнимы между собой, а после наших преобразований эта характеристика у них появилась. Это привело к неоднозначности итогового решения: при `male=0` массив ответа имеет вид: [0.14751816 0.29538468 0.25658495 0.30051221], а при `female=0` – [0.14000522 0.30343647 0.2560461 0.30051221]. Конечно, в нашей ситуации изменение незначительное и не влияет на итоговый ответ, но в другой ситуации это может быть не так. Поэтому вам предлагается самостоятельно подумать, как избежать подобного искажения данных.

Работа с линейной регрессией

Мы рассмотрели работу с решающими деревьями, теперь же перейдем к рассмотрению регрессии. Начнем с линейных моделей, приведя в качестве примера решение задачи прогнозирования уровня зарплаты по данным вакансий.

Линейные методы хорошо подходят для работы с разреженными данными. К таковым относятся, например, тексты. Это можно объяснить высокой скоростью обучения и небольшим количеством параметров, благодаря чему удается избежать переобучения.

Линейная регрессия имеет несколько разновидностей в зависимости от того, какой регуляризатор используется. В рамках предлагаемой задачи мы будем использовать гребневую регрессию, где применяется квадратичный, или L2-регуляризатор. Эта модель реализована в классе `sklearn.linear_model.Ridge`. Более подробно о реализации гребневой регрессии можно почитать здесь: [40].

Исходные данные объемом 60 000 записей хранятся в файле «salary_train.csv», тестовые данные (2 записи) хранятся в файле «salary_test_mini.csv» (скачать их можно отсюда: [41]). Оба файла имеют вид, представленный на рисунке 57.

В данном случае столбцы A-C будут содержать значения входных признаков, а последний столбец – целевой переменной.

	A	B	C	D
1	FullDescription	LocationNormalized	ContractTime	SalaryNormalized
2	International Sales Manager London ****k ****k Uncapped Commission Digital Marketing/	London	permanent	33000
3	An ideal opportunity for an individual that has gained experience in the field of M A to furth	London	permanent	50000
4	Online Content and Brand Manager// Luxury Retail // Up to ****k InstinctaБ™s Central Lond	South East London	permanent	40000
5	A great local marketleader is seeking a permanent Payroll and Purchase Ledger Assistant to	Dereham	permanent	22500
6	Registered Nurse / RGN Nursing Home for Young Adults Location; Sutton Coldfield Part Tim	Sutton Coldfield		20355
7	Sales and Marketing Assistant will provide administrative support to the staff of the Sales/S	Crawley		22500
8	Vacancy Ladieswear fashion Area Manager / Regional Manager Northern Ireland where will	UK	permanent	32000
9	Reference: LR/JAN/**** Our client is one of the largest Multi Disciplinary Engineering Consu	Bristol	permanent	30000
10	Sponsorship Manager London The Company A market leading event and conference busines	Central London	permanent	31500
11	About Barclays Barclays moves, lends, invests and protects money for customers and clients	South East London	permanent	42499
12	Counter Sales Person с****k dependent on experience Banchory Temporary MADE TO MEAS Banchory	Banchory	contract	16000
13	Central London Competitive Salary Bonus & Benefits Summary Aegis Media Limited (AML) i	Gipsy Hill	permanent	47500
14	We are recruiting for a dynamic organisation based near pontefract in Wakefield. They requi	Wakefield	permanent	20000
15	Composite Design Engineer required for a permanent role for Oxfordshire. Need relevant d	Abingdon	permanent	34000
16	Working managing a team to maintain the Home by cleaning, dusting, vacuuming and polish	Northampton		15360
17	HCL Nursing are actively recruiting skilled registered general nurses to work in various hospi	Bradford	permanent	39811
18	HGV / Commercial vehicle Customer service advisor My client based in North London is curr	Northamptonshire	contract	23000
19	Break into Banking .NET Developer (C/Agile/WinForms/WebForms/SQL)*****K This is a	London	permanent	47500
20	Registered Nurse (RGN) is required to work within the elderly nursing home environment. T	Tiverton		24500
21	Can you drive effective analytical solutions for the European Corporate compensation and b	Uxbridge		32000
22	nbsp; My client is an small, but high profile international charity who are looking to looking	London	permanent	45000
23	Specialist Clinical Negligence Practice based in Manchester City Centre has a new opportuni	Lancashire	permanent	20000
24	Job description: вЂў Answer contacts promptly and professionally вЂў Log/Validate all cont	UK	permanent	21000
25	YEAR **** TEACHER ROLE Servoca Education is currently looking for a Year **** class teacher	Surrey	contract	34800

Рисунок 57. Исходные данные

Как можно заметить, первый столбец содержит объемный текст, который необходимо предобработать для загрузки в модель. Например, извлечь TF-IDF-признаки, воспользовавшись классом `sklearn.feature_extraction.text.TfidfVectorizer`, который преобразует массив текста в матрицу, содержащую TF-IDF-признаки. Более

подробно об этом классе можно почитать здесь: [42]. Использование его для решаемой задачи показан в листинге 25.

Листинг 25

```
vectorizer = TfidfVectorizer(min_df=10)
train_text_feature_matrix = vectorizer.fit_transform(data_train['FullDescription'])
idf = vectorizer.idf_
print dict(zip(vectorizer.get_feature_names(), idf))
```

В данном случае поле `idf_` класса `TfidfVectorizer` содержит вектор с глобальными весовыми коэффициентами. Строка кода `print dict(zip(vectorizer.get_feature_names(), idf))` выведет в консоль массив данных вида: `{u'pre': 5.6339293213815242, u'paperless': 9.0576775285655771, u'peoplewithchemistry': 9.6042212349336467, u'rfps': 9.1117447498358519, u'yellow': 8.957594070008593...}`.

Теперь, что касается оставшихся столбцов-признаков: `LocationNormalized` и `ContractTime` являются строковыми, и поэтому с ними нельзя работать напрямую. Применим к ним one-hot-кодирование. Для рассматриваемого нами языка Python оно реализовано в классе `sklearn.feature_extraction.DictVectorizer`, который преобразует списки сопоставленных пар признак-значение в массивы, с которыми может работать Numpy, или в разреженные матрицы (`scipy.sparse-matrices`) для использования с классами-«измерителями качества» `scikit-learn` (`scikit-learnestimators`). Когда значения признаков строковые, этот трансформер выполняет бинарное one-hot-кодирование: для каждого из возможных строковых значений признака строится признак с логическим значением, говорящий, принимает данный признак указанное строковое значение или нет. Например, пусть на входе есть признак «с», который может принимать значения «кот» и «дог». На выходе будет два признака: один – «с = кот», другой – «с = дог». Признаки, которые не представлены в обучающей выборке, будут иметь нулевое значение в результирующей матрице или результирующем массиве.

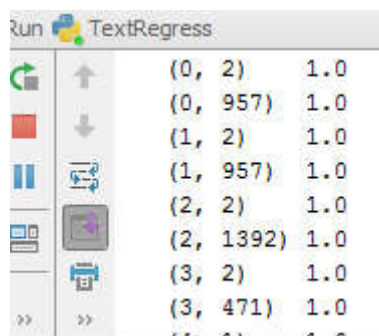
Подробнее о DictVectorizer можно прочитать здесь: [43]. Пример использования этого класса для указанных данных приведен в листинге 26.

Листинг 26

```
# -*- coding: utf-8 -*-
from sklearn.feature_extraction import DictVectorizer
.....
enc = DictVectorizer()
train_dic = data_train[['LocationNormalized', 'ContractTime']].to_dict('records')
test_dic = data_test[['LocationNormalized', 'ContractTime']].to_dict('records')
X_train_categ = enc.fit_transform(train_dic)
X_test_categ = enc.transform(test_dic)
```

При этом считается, что в `data_train` загружены данные обучающей выборки, а в `data_test` – данные, на которых будет проверяться работа модели.

В результате работы кода из листинга 26 `train_dic` (также как и `test_dic`) будет содержать множество записей, таких как: `{'LocationNormalized': 'UK', 'ContractTime': 'permanent'}`, `{'LocationNormalized': 'Bradford', 'ContractTime': 'permanent'}`. `X_train_categ` будет содержать данные, представленные на рисунке 58.



Row	Col 1	Col 2	Value
0	2	1.0	1.0
0	957	1.0	1.0
1	2	1.0	1.0
1	957	1.0	1.0
2	2	1.0	1.0
2	1392	1.0	1.0
3	2	1.0	1.0
3	471	1.0	1.0

Рисунок 58. Преобразованные данные

Интерпретировать их можно следующим образом. `X_train_categ` – это по сути матрица, но только не в классическом формате. Если точнее, то `X_train_categ` – это разреженная матрица, которая в сокращенном виде описывает некую полную матрицу (назовем ее `X_Full`).

В `X_train_categ` указано, в каких позициях `X_Full` стоят единицы. Так `(0, 2) 1.0` – означает, что в нулевой строке и втором столбце стоит единица. Во всех остальных ячейках матрицы `X_Full`, которые не обозначены в `X_train_categ`, стоят нули. `X_Full` – это матрица размера `M` на `N`, где `M` – число строк, которое определяется количеством объектов в выборке, а `N` – число уникальных слов в столбцах `['LocationNormalized', 'ContractTime']` исходной таблицы.

Теперь вернемся к предобработке данных. Так как в строковых данных есть пропуски, то нам потребуется заменять их на специальные строковые величины (например, `'nan'`). Для этого можно использовать код из листинга 27.

Листинг 27

```
data_train['LocationNormalized'].fillna('nan', inplace=True)
```

С учетом изложенного выше, алгоритм решения поставленной задачи будет следующим:

1. Загрузить данные об описаниях вакансий и соответствующих годовых зарплатах из файла `salary-train.csv`.
2. Провести его предобработку:
 - 2.1. Привести тексты к нижнему регистру (`text.lower()`).
 - 2.2. Заменить все, кроме букв и цифр, на пробелы для облегчения дальнейшего разделение текста на слова. Для такой замены в строке `text` подходит следующий вызов: `re.sub('[^a-zA-Z0-9]', ' ', text)`. Также можно воспользоваться методом `replace` у `DataFrame`, чтобы сразу преобразовать все тексты, например, так: `train['FullDescription'] = train['FullDescription'].replace('[^a-zA-Z0-9]', ' ', regex = True)`
 - 2.3. Применить `TfidfVectorizer` для преобразования текстов в векторы признаков. Оставить только те слова, которые встречаются хотя бы в 5 объектах. Для этого использовать параметр `min_df` у `TfidfVectorizer`.

- 2.4. Заменить пропуски в столбцах LocationNormalized и ContractTime на специальную строку 'nan' (листинг 27).
- 2.5. Применить DictVectorizer для получения one-hot-кодирования признаков LocationNormalized и ContractTime.
- 2.6. Объединить все полученные признаки в одну матрицу «объекты-признаки».

Обратите внимание!

- *Матрицы для текстов и категориальных признаков являются разреженными. Для объединения их столбцов нужно воспользоваться функцией `scipy.sparse.hstack`.*

- 2.7. Обучить гребневую регрессию с параметрами `alpha=1` и `random_state=241`, помня, что целевая переменная записана в столбце SalaryNormalized. Обучение производится вызовом метода `fit`.
3. Построить прогноз для двух примеров из файла `salary-test-mini.csv`. Прогноз строится методом `predict`.
- Код решения задачи (с комментариями) – в листинге 28. Результат работы – два прогнозных числа для двух записей из тестовой выборки: [56876.4573163 37557.00551031].

Листинг 28

```
# -*- coding: utf-8 -*-
import pandas
from sklearn.feature_extraction import DictVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import Ridge
from scipy.sparse import hstack

#Загрузка данных
data_train = pandas.read_csv('salary-train.csv')
data_test = pandas.read_csv('salary-test-mini.csv')

# Приведение текста к нижнему регистру
data_train['FullDescription'] = data_train.FullDescription.str.lower()
data_test['FullDescription'] = data_test.FullDescription.str.lower()
```


Окончание листинга 28

```
# Удаление ненужных символов
data_train['FullDescription'] = data_train['FullDescription'].replace('[^a-zA-Z0-9]', '',
regex=True)
data_test['FullDescription'] = data_test['FullDescription'].replace('[^a-zA-Z0-9]', '',
regex=True)
# Преобразование текста в вектор признаков, используя TfidfVectorizer из sklearn
vectorizer = TfidfVectorizer(min_df=10)
train_text_feature_matrix = vectorizer.fit_transform(data_train['FullDescription'])
test_text_feature_matrix = vectorizer.transform(data_test['FullDescription'])
# Заполнение пустых ячеек
data_train['LocationNormalized'].fillna('nan', inplace=True)
data_train['ContractTime'].fillna('nan', inplace=True)
data_test['LocationNormalized'].fillna('nan', inplace=True)
data_test['ContractTime'].fillna('nan', inplace=True)
# Преобразование категориальных признаков в числовые
enc = DictVectorizer()
train_dic = data_train[['LocationNormalized', 'ContractTime']].to_dict('records')
test_dic = data_test[['LocationNormalized', 'ContractTime']].to_dict('records')
X_train_categ = enc.fit_transform(train_dic)
X_test_categ = enc.transform(test_dic)
# Здесь по горизонтали объединяется разреженная матрица с признаками текста (из
# столбца FullDescription) и матрица с закодированными категориями (из столбцов
# Location Normalized и Contract Time)
x_train = hstack((train_text_feature_matrix, X_train_categ))
y_train = data_train['SalaryNormalized'].values
x_test = hstack((test_text_feature_matrix, X_test_categ))
# Создание и обучение регрессии
ridge_regression = Ridge(alpha=1, random_state=241)
ridge_regression.fit(x_train, y_train)
# Получение и распечатка прогнозных значений
y_test = ridge_regression.predict(x_test)
print y_test
```

Сохранение и загрузка обученной модели

Как вы могли заметить, работа программы идет довольно долго. Большую часть этого времени занимает процесс обучения, поэтому при работе с обученной моделью используют практику сохранения и загрузки ее параметров в\из файл(а). Рассмотрим следующий

пример. Пусть у нас есть некоторый массив данных о двух факторах и зависимой переменной, сохраненные в файле «data-logistic.csv» (скачать можно отсюда: [41]). Выполним следующее: обучим на этих данных модель линейной регрессии, сохраним ее в файл, создадим еще одну новую модель линейной регрессии, обучим ее на части данных, затем загрузим ранее обученную модель из файла и сравним результаты предсказания обеих моделей. Сохранение и загрузку мы будем выполнять с помощью модуля pickle, и для удобства вынесем код загрузки и сохранения в отдельные функции LoadFromObject и SaveToObject. Решение задачи показано в листинге 29, а ее результат – на рисунке 59.

Листинг 29

```
# -*- coding: utf-8 -*-

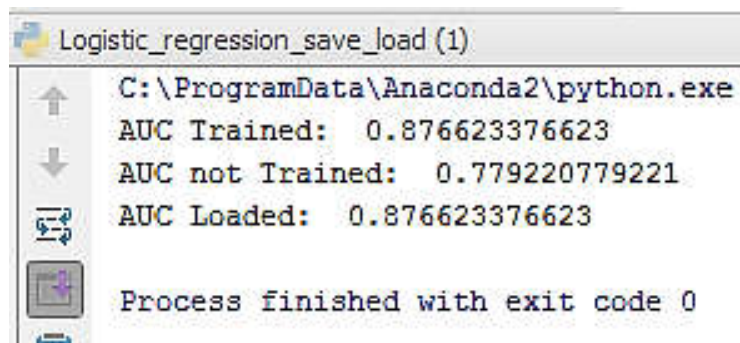
import pandas
import pandas.core.series as Series
import numpy as np
from scipy.sparse import hstack
from sklearn.metrics import roc_auc_score
from sklearn.metrics import accuracy_score
import sys
import matplotlib.pyplot as plt
from sklearn import linear_model
import pickle

#Определяем функцию сохранения
def SaveToObject(obj, filename):
    with open(filename, 'wb') as output:
        pickle.dump(obj, output)

#Определяем функцию загрузки
def LoadFromObject(filename):
    f = open(filename, 'rb')
    loaded_obj = pickle.load(f)
    f.close()
    return loaded_obj
```

Окончание листинга 29

```
# Загружаем данные
data_train = pandas.read_csv('data-logistic.csv')
X = data_train.ix[:,1:3].values
y = (data_train.ix[:,0].values + 1)/2
#Для лучшего понимания представления данных распечатайте формы массивов:
print X.shape
print y.shape
# Создание, обучение модели и получение ее прогноза
regression = linear_model.LogisticRegression()
regression.fit(X, y)
ans = regression.predict(X)
# Оценка качества созданной модели
metric = roc_auc_score(y, ans)
print "AUC Trained: ", metric.real
#Сохранение и загрузка модели в файл с именем model
SaveToObject(regression, "model")
loaded_model = LoadFromObject("model")
#Создание, обучение новой модели, получение ее прогноза и оценка качества
not_fitted = linear_model.LogisticRegression()
not_fitted.fit(X[0:2:], y[0:2])
ans = not_fitted.predict(X)
metric = roc_auc_score(y, ans)
print "AUC not Trained: ", metric.real
#получение прогноза и оценки качества загруженной модели
ans = loaded_model.predict(X)
metric = roc_auc_score(y, ans)
print "AUCLoaded: ", metric.real
```



```
Logistic_regression_save_load (1)
C:\ProgramData\Anaconda2\python.exe
AUC Trained: 0.876623376623
AUC not Trained: 0.779220779221
AUC Loaded: 0.876623376623
Process finished with exit code 0
```

Рисунок 59. Результат работы кода из листинга 29

Как видно из рисунка 59, модель, обученная на части данных, дала более плохой результат, тогда как качество загруженной модели не изменилось. То есть можно говорить о том, что сохранение и загрузка прошли корректно.

Работа с логистической регрессией

Мы рассмотрели применение регрессии для решения задачи предсказания, теперь перейдем к задаче классификации и познакомимся с еще одной моделью – логистической регрессией. Это один из видов линейных классификаторов. Ее особенность заключается в том, что результатом является вероятность принадлежности объекта к классу N , тогда как большинство линейных классификаторов могут выдавать только номера классов.

Логистическая регрессия использует достаточно сложную функцию для оценки качества, которая не допускает записи решения в явном виде (в отличие от, например, линейной регрессии). Тем не менее логистическую регрессию можно настраивать с помощью градиентного спуска.

Для примера мы выполним следующее: сгенерируем некоторую случайную выборку и каждому ее элементу поставим в соответствие число 0 или 1, которое будет обозначать принадлежность к первому или второму классу. Затем обучим логистическую модель на этих данных и линейную модель. После – построим график и сравним результаты. Для генерации случайной выборки из 100 элементов используем следующий код из листинга 30.

Листинг 30

```
import numpy as np
.....
n_samples = 100
np.random.seed(0)
X = np.random.normal(size=n_samples)
```

Обратите внимание!

- *В дальнейших листингах операторы импорта библиотек будут появляться только при первом вызове импортируемых элементов, так как предполагается, что все представленные ниже листинги формируют единый код, который вам предлагается собрать в единый файл самостоятельно.*

Для формирования множества откликов Y определим функцию, реализующую правило: $Y=1$ для всех $X>0$ и $Y=0$ – в остальных случаях. Код функции можно увидеть в листинге 31.

Листинг 31

```
Y = (X > 0).astype(np.float)
```

Следующим шагом нам необходимо выполнить преобразования нашего одномерного массива X : во-первых, добавить небольшой разброс и смещение, чтобы избавиться от прямой функциональной зависимости. Во-вторых, нужно преобразовать его в двумерный массив. Последнее выполняется для функции обучения регрессии: на вход она требует данные в формате двумерного массива. Преобразование выполняется кодом, представленным листинге 32.

Листинг 32

```
X[X > 0] *= 4  
X += .3 * np.random.normal(size=n_samples)  
X = X[:, np.newaxis]
```

Теперь мы можем объявить и обучить модели регрессии, как показано в листинге 33.

Листинг 33

```
from sklearn import linear_model  
log_reg = linear_model.LogisticRegression()  
log_reg.fit(X, Y)  
lin_reg = linear_model.LinearRegression()  
lin_reg.fit(X, Y)
```

Подробнее о реализации логистической модели можно прочитать здесь: [44], а линейной – здесь: [45].

Следующим шагом нам необходимо сгенерировать тестовые данные и выполнить предсказание. Последнее будет выполняться методом `predict`. Он возвращает бинарные метки классов (0,1). Кроме этого существует еще метод `predict_proba` (для логистической регрессии). Он возвращает вероятность принадлежности объекта к классу 0 или 1 соответственно. Поскольку у нас два класса, то в результате `predict_proba` получается двумерный массив (один столбец отражает принадлежность к классу 0, а другой к классу 1). Для отображения мы выберем лишь первый столбец.

Обратите внимание!

- *Сумма элементов массивов (столбцов, векторов) с одним индексом равна единице по правилу полной вероятности.*

Код генерации тестовых данных и вызовы методов прогноза представлен в листинге 34.

Листинг 34

```
X_test = np.linspace(start=-5, stop=10, num=300)
X_test = X_test[:, np.newaxis]
y_log_label = log_reg.predict(X_test)
y_lin = lin_reg.predict(X_test)
y_log_probabilty = log_reg.predict_proba(X_test)[:,1]
```

Следующим шагом нам необходимо вывести график для отображения результатов. Создавать полотно для рисования графиков и отображать их мы будем с помощью библиотеки `Matplotlib`, как показано в листинге 35.

Листинг 35

```
# -*- coding: utf-8 -*-
import matplotlib.pyplot as plt
....
```

Окончание листинга 35

```
# рисуем исходные точки
plt.figure(1, figsize=(4, 3))
plt.scatter(X.ravel(), y, color='black', zorder=20)
# рисуем график меток логистической регрессии
plt.plot(X_test, y_log_label, color='red', linewidth=3)
# рисуем график линейной регрессии
plt.plot(X_test, y_lin, linewidth=1)
# добавляем линию уровня, по которому точки будут относиться к 1 или 2 классу
plt.axhline(.5, color='green')
# рисуем логистическую кривую для вероятностей принадлежности объектов к
#классам
plt.plot(X_test, y_log_probabilty, color='blue', linewidth=3)
# настраиваем оси и выводим график
plt.ylabel('y')
plt.xlabel('X')
plt.xticks(range(-5, 10))
plt.yticks([0, 0.5, 1])
plt.ylim(-.25, 1.25)
plt.xlim(-4, 10)
plt.legend(('Label Logistic Regression Model', 'Linear Regression Model',
'Probabilty Log Regression'), loc="lowerright", fontsize='small')
plt.show()
```

В результате после выполнения кода появится график, как показано на рисунке 60.

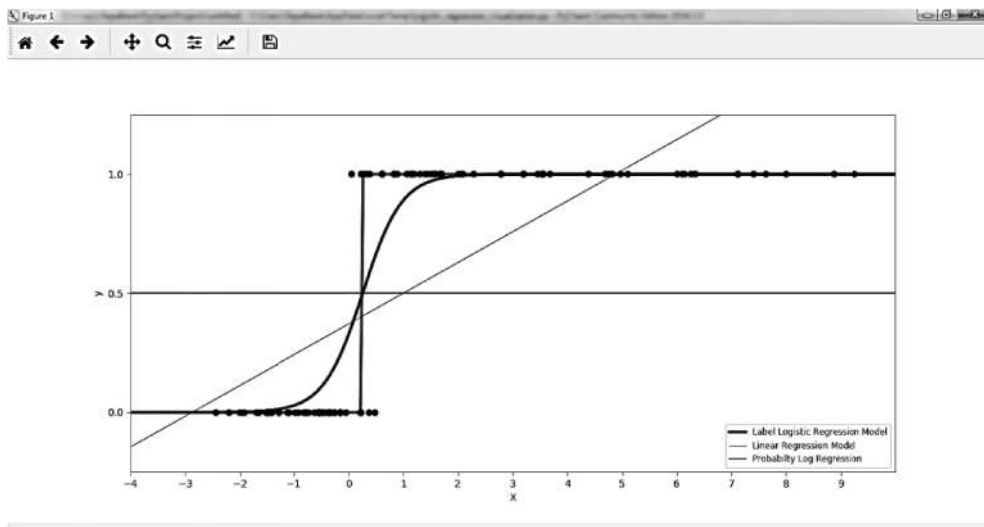


Рисунок 60. Графики регрессий

Для лучшего понимания вам рекомендуется воспроизвести код и посмотреть на график, который вам выведет программа. На нем должно быть видно, что красная кривая (результат логистического предсказания\классификации) лучше приближает распределение точек (лишь несколько точек вылетает). Красная линия построена лишь по меткам, полученным от `log_reg.predict`, поэтому она имеет такой резкий переход. Такой «округленный» результат легко использовать в дальнейшем, хотя в таком случае мы не получаем информацию о вероятностях. Линейная регрессия (голубая линия) приближает данные не так хорошо.

Важно понимать, что линейная регрессия отнесет все точки ($x < 1$) к классу 0, так как результат функции линейной регрессии будет меньше 0.5. А все точки ($X > 1$) будут отнесены к классу 1, так как результат функции линейной регрессии будет больше 0.5. Для этого на графике выведена разделяющая линия $Y = 0.5$. Граница разделения классов по логистической регрессии проходит примерно в точке $X = 0.25$, что разделяет классы гораздо точнее. Толстая синяя линия отображает вероятностную логистическую кривую, которая дает больше детальной информации о предсказании. Поскольку результат представляет собой не бинарные метки классов, а вероятность принадлежности к выбранному классу, то кривая получается гладкой.

Для точной оценки качества моделей воспользуемся функциями `score` и `accuracy_score` в `scikit-learn`. Поскольку линейная регрессия выдает вещественные значения, а `y_test` сгенерирована как бинарный массив, то для оценки качества нужно использовать функцию `model.score()`, которая способна читать и бинарные, и вещественные ответы. Для подсчета качества логистической регрессии можно воспользоваться измерителем качества классификации `accuracy_score`. Но перед этим нам необходимо получить реальную принадлежность данных тестовой выборки к классам. Сделать это можно с помощью определенной нами функции для обучающей выборки, как показано в листинге 36.

Листинг 36

```
y_test = (X_test > 0).astype(np.float)
```

Теперь можно вывести значения качества классификации для обеих моделей, как показано в листинге 37. Для линейной регрессии это выполняется ее методом `score`, а для логистической – методом `accuracy_score` из `sklearn.metrics`.

Листинг 37

```
from sklearn.metrics import accuracy_score
....
# оцениваем точность решений
lin_score = lin_reg.score(X_test, y_test)
print "Linear regression accuracy: ", lin_score
log_score = accuracy_score(y_log_label, y_test)
print "Logistic regression accuracy: ", log_score
```

В итоге в консоль будут выведены следующие результаты: `Linearregression accuracy=0,52442`, `Logisticregression accuracy=0,98333`. Они подтверждают, что логистическая регрессия выполнила классификацию более качественно.

Решение задачи ранжирования признаков

Следующее, что нам необходимо рассмотреть, – применение регрессионных моделей для определения важности признаков. Здесь в качестве примера возьмем регрессионную проблему Фридмана, когда на вход моделей подается 14 факторов, выход рассчитывается по формуле, использующей только пять факторов, но факторы 1-5, а также 10-14 взаимозависимы. Наша задача – посмотреть, как разные виды регрессий оценят важности факторов и какой из них будет иметь наибольшую среднюю значимость по всем моделям. Для работы мы возьмем три модели: линейную регрессию, гребневую и лассо (линейную регрессию с L1-регуляризатором). О реализации в `scikit-learn` последней модели можно подробнее прочитать здесь: [46].

Первое, что нам необходимо сделать для решения поставленной задачи, – подключить необходимые модули, как в листинге 38. Вторая строка кода из листинга 38 – импорт класса, выполняющего масштабирование данных до заданного диапазона (например, так, чтобы все значения находились в диапазоне от 0 до 1). Подробнее о реализации класса можно прочитать здесь: [47]. Необходимость его использования объясняется следующим: каждая модель регрессии дает оценки важности признаков в своем диапазоне. Для того чтобы найти признак с максимальной средней важностью по трем моделям, нам необходимо привести выданные ими оценки к одному виду, в чем и поможет нам указанный класс.

Листинг 38

```
from sklearn.linear_model import LinearRegression, Ridge, Lasso
from sklearn.preprocessing import MinMaxScaler
import numpy as np
```

После импорта необходимых классов мы должны сгенерировать исходные данные, как показано в листинге 39.

Листинг 39

```
#генерируем исходные данные: 750 строк-наблюдений и 14 столбцов-признаков
np.random.seed(0)
size = 750
X = np.random.uniform(0, 1, (size, 14))
#Задаем функцию-выход: регрессионную проблему Фридмана
Y = (10 * np.sin(np.pi*X[:,0]*X[:,1]) + 20*(X[:,2] - .5)**2 +
     10*X[:,3] + 5*X[:,4]**5 + np.random.normal(0,1))
#Добавляем зависимость признаков
X[:,10:] = X[:,4] + np.random.normal(0, .025, (size,4))
```

Теперь создаем и обучаем модели (листинг 40).

Листинг 40

```
#линейная модель
lr = LinearRegression()
```

Окончание листинга 40

```
lr.fit(X, Y)
#гребневая модель
ridge = Ridge(alpha=7)
ridge.fit(X, Y)
#Лассо
lasso = Lasso(alpha=.05)
lasso.fit(X, Y)
```

Наконец, нам нужно выгрузить в единый массив размера 3×14 (количество_моделей и количество_признаков) все оценки моделей по признакам. Найти средние оценки и вывести результат в формате списка пар {номер_признака – средняя_оценка}, отсортированном по убыванию. Получить оценки признаков можно через поле coef_ у каждой из наших моделей. Для удобства отображения данных поместим их в конструкцию вида: [имя_модели : [{имя_признака : оценка}, {имя_признака : оценка}]]. То есть верхним уровнем у нас будет словарь, где ключом будет имя модели. В нем будут располагаться три записи из четырнадцати пар каждая. Пример содержимого списка представлен в листинге 41.

Листинг 41

```
{'Lasso': {'x13': 0.0, 'x14': 0.0, 'x10': 0.0, 'x8': 0.0, 'x9': 0.0, 'x11': 0.0, 'x2': 0.72, 'x3': 0.0,
'x12': 0.0, 'x1': 0.69, 'x6': 0.0, 'x7': 0.0, 'x4': 1.0, 'x5': 0.29},
'Ridge': {'x13': 0.0, 'x14': 0.92, 'x10': 0.01, 'x8': 0.08, 'x9': 0.0, 'x11': 0.59, 'x2': 0.75, 'x3': 0.06,
'x12': 0.67, 'x1': 0.76, 'x6': 0.08, 'x7': 0.03, 'x4': 1.0, 'x5': 0.61},
'Linear reg': {'x13': 0.48, 'x14': 0.12, 'x10': 0.01, 'x8': 0.03, 'x9': 0.0, 'x11': 0.59, 'x2': 0.61, 'x3':
0.51, 'x12': 0.19, 'x1': 1.0, 'x6': 0.02, 'x7': 0.0, 'x4': 0.69, 'x5': 0.19}}
```

Первое, что нам необходимо сделать, чтобы реализовать описанное, – создать список вида ['x1', 'x2', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8', 'x9', 'x10', 'x11', 'x12', 'x13', 'x14'], содержащий имена признаков. Выполнить это можно с помощью кода из листинга 42.

Листинг 42

```
names = ["x%s" % i for i in range(1,15)]
```

Затем нам необходимо объявить функцию, которая на вход будет получать сформированный выше список и список оценок по признакам. Выходом же функции должен быть словарь, составленный из попарно соотнесенных элементов списков (то есть словарь из пар имя_признака: оценка_признака, где первое – ключ, а второе – значение). Причем оценки внутри функции должны быть приведены к единому диапазону от 0 до 1 и округлены до сотых.

В теле функции нам потребуются следующие объекты. Во-первых, класс `MinMaxScaler`, о котором было сказано выше. Во-вторых, функция `abs` пакета `Numpy` для получения абсолютных значений оценок(модуля). В-третьих, функция `map`, позволяющая применить операцию округления (`round`) к каждому элементу массива посредством оператора `lambda`. В-четвертых, нам понадобятся функции `rashape` и `ravel`. Первая – для переформирования входного массива признаков из одной строки и четырнадцати столбцов в четырнадцать строк и один столбец, вторая – для преобразования полученного двумерного массива в одномерный. В-пятых, нам потребуются функции `zip` и `dict`. Первая попарно соотнесет имена признаков и оценки, вторая – преобразует полученные пары в словарь. Примеры использования функций `zip`, `map` и `lambda` можно найти здесь: [48]. Код реализации описанных действий можно найти в листинге 43.

Листинг 43

```
def rank_to_dict(ranks, names):
    ranks = np.abs(ranks)

    minmax = MinMaxScaler()

    ranks = minmax.fit_transform(np.array(ranks).reshape(14,1)).ravel()
    ranks = map(lambda x: round(x, 2), ranks)
    return dict(zip(names, ranks))
```

Вызов функции для наших моделей представлен в листинге 44.

Листинг 44

```
ranks["Linear reg"] = rank_to_dict(lr.coef_, names)
ranks["Ridge"] = rank_to_dict(ridge.coef_, names)
ranks["Lasso"] = rank_to_dict(lasso.coef_, names)
```

Последним шагом мы должны сформировать среднее по каждому признаку, загрузив их в отдельный список. Затем отсортировать его по убыванию и вывести. Реализовать это можно, используя код из листинга 45.

Листинг 45

```
#Создаем пустой список для данных
mean = {}

#«Бежим» по списку ranks
for key, value in ranks.iteritems():
#«Пробегаемся» по списку значений ranks, которые являются парой имя:оценка
for item in value.iteritems():
#имя будет ключом для нашего mean
#если элемента с текущим ключем в mean нет - добавляем
if(item[0] not in mean):
mean[item[0]] = 0

#суммируем значения по каждому ключу-имени признака
    mean[item[0]] += item[1]

#находим среднее по каждому признаку
for key, value in mean.iteritems():
    res=value/len(ranks)
    mean[key] = round(res, 2)

#сортируем и распечатываем список
mean = sorted(mean.iteritems(), key=lambda (x, y): y, reverse=True)

print "MEAN"
print mean
```

Результат работы программы представлен в листинге 46.

Листинг 46

MEAN

```
[('x4', 0.9), ('x1', 0.82), ('x2', 0.69), ('x11', 0.39), ('x5', 0.36), ('x14', 0.35), ('x12', 0.29), ('x3', 0.19), ('x13', 0.16), ('x8', 0.04), ('x6', 0.03), ('x10', 0.01), ('x7', 0.01), ('x9', 0.0)]
```

Мы видим, что по средней оценке наиболее важным оказался четвертый признак, затем первый и потом второй. Сравним эти данные с показателями каждой отдельной модели. Для этого отсортируем массив `ranks` также по убыванию оценок и выведем его, используя код из листинга 47. Результат работы кода представлен в листинге 48.

Листинг 47

```
for key, value in ranks.iteritems():
    ranks[key] = sorted(value.iteritems(), key=lambda (x, y): y, reverse=True)
for key, value in ranks.iteritems():
    print key
    print value
```

Листинг 48

Lasso

```
[('x4', 1.0), ('x2', 0.72), ('x1', 0.69), ('x5', 0.29), ('x13', 0.0), ('x14', 0.0), ('x10', 0.0), ('x8', 0.0), ('x9', 0.0), ('x11', 0.0), ('x3', 0.0), ('x12', 0.0), ('x6', 0.0), ('x7', 0.0)]
```

Ridge

```
[('x4', 1.0), ('x14', 0.92), ('x1', 0.76), ('x2', 0.75), ('x12', 0.67), ('x5', 0.61), ('x11', 0.59), ('x8', 0.08), ('x6', 0.08), ('x3', 0.06), ('x7', 0.03), ('x10', 0.01), ('x13', 0.0), ('x9', 0.0)]
```

Linear reg

```
[('x1', 1.0), ('x4', 0.69), ('x2', 0.61), ('x11', 0.59), ('x3', 0.51), ('x13', 0.48), ('x12', 0.19), ('x5', 0.19), ('x14', 0.12), ('x8', 0.03), ('x6', 0.02), ('x10', 0.01), ('x9', 0.0), ('x7', 0.0)]
```

Как можно увидеть, метод Lasso отобрал признаки `x1`, `x2`, `x4`, `x5` как значимые параметры, а все остальные совсем незначимые (так как они равны нулю). Что довольно близко к истине, так как параметры `x1-x5` заданы как независимые случайные величины. Но взвешивание по такому методу довольно грубое, резкое.

Метод Ridge сделал более мягкое взвешивание. Но несмотря на то, что `x3` также упущен, его влияние может быть учтено через скорре-

лированные параметры x_{14} , x_{12} , x_{11} (которые были также отобраны как важные). Поэтому данный метод может быть очень полезным.

Метод линейной регрессии также отдал предпочтение многим действительно важным параметрам (причем x_3 здесь имеет гораздо больший вес).

В заключение можно сказать, что учет оценок от разных методов позволит более оптимально выделить значимые признаки и отделить от менее значимых. Но, конечно, без реальных экспериментов, проверяющих выбранное множество параметров, невозможно гарантировать идеального разбиения для всех задач.

Кроме рассмотренных выше моделей для отбора признаков еще можно использовать:

1. `RandomizedLasso` из `sklearn.linear_model` (оценки признаков будут находиться в поле `scores_`). Это так называемое рандомизированное лассо работает, разделяя тренировочные данные на подвыборки и вычисляя Лассо-оценки признаков, где штраф случайного подмножества коэффициентов масштабирован. Применяя такую двойную рандомизацию несколько раз, метод присваивает наивысшие оценки признакам, выбор которых повторялся в процессе рандомизации. Это называется «выбором стабильности». Если же сказать короче, то признаки, которые выбираются наиболее часто, считаются значимыми. Подробнее об этом классе можно почитать здесь: [49].

2. `RFE` из `sklearn.feature_selection` (оценки признаков будут находиться в поле `ranking_`, однако в отличие от всех остальных моделей, класс выдает не веса при коэффициентах регрессии, а именно ранг для каждого признака. Так наиболее важные признаки будут иметь ранг – «1», а менее важные признаки ранг больше «1». Коэффициенты остальных моделей тем важнее, чем больше их абсолютное значение). Класс `RFE` выполняет ранжирование признаков через рекурсивный отбор путем отсева признаков. Целью рекурсивного устранения элемента (`RFE`) является выбор признаков путем рекурсивного рассмотрения меньших и меньших наборов признаков, учитывая внешнюю оценку, которая присваивает признакам веса

(например, коэффициенты линейной модели). Сначала оценщик обучается на начальном наборе признаков и ассоциированных с ними весами. Затем признаки, абсолютные веса которых являются наименьшими, удаляются из текущего набора. Эта процедура рекурсивно повторяется на оставшемся множестве до тех пор, пока в конечном итоге не будет достигнуто желаемое количество признаков. То есть работа с этим классом строится, как показано в листинге 49. Подробнее о нем можно почитать здесь: [50].

Листинг 49

```
lr = LinearRegression()
lr.fit(X, Y)

rfe = RFE(lr)
rfe.fit(X, Y)
```

3. RandomForestRegressor из sklearn.ensemble (оценки признаков будут находиться в поле feature_importances_). Регрессор на основе Случайного леса – это мета-оценщик, который обучает несколько классификационных деревьев решений на разных подвыборках данных и использует усреднение для улучшения точности и контроля переобучения. Размер подвыборки всегда соответствует размеру входной выборки, но образцы из выборки перемешиваются (если параметр bootstrap=True, что является значением по умолчанию). Подробнее об этом классе можно посмотреть здесь: [51].

4. f_regression из sklearn.feature_selection (для получения оценок см.код из листинга 50. Оценки будут находиться в объекте f). Это быстрая линейная модель, позволяющая оценить эффект одного признака и применяющаяся последовательно к множеству признаков. В основе ее работы лежит вычисление взаимной корреляции между каждым признаком и выходной переменной. Более подробно можно посмотреть здесь: [52].

Листинг 50

```
f, pval = f_regression(X, Y, center=True)
```


Работа с полиномиальной регрессией

Напоследок, заканчивая разговор о регрессионных моделях, мы познакомимся с полиномиальной регрессией, а также посмотрим на практике такое явление, как переобучение модели. Рассматривать мы это будем на следующем примере: сгенерируем точки функцией косинуса и попробуем аппроксимировать их линейной регрессией, полиномиальной регрессией со степенью 4 и 15, а также гребневой регрессией и созданными на основе нее полиномами 4 и 15 степеней. Затем рассчитаем точность моделей, выведем шесть графиков и сравним результаты.

Итак, первым шагом создадим обычную функцию косинуса от аргумента, записанную в виде lambda-выражения (листинг 51). Особенности Python позволяют использовать в качестве аргумента не только числа, но и массивы\векторы (так что использование циклов не требуется).

Листинг 51

```
import numpy as np
true_fun = lambda X: np.cos(1.5 * np.pi * X)
```

Представленная в листинге 51 функция будет использоваться как истинная закономерность в некоторых данных. То есть именно то, что аналитики и ученые пытаются выяснить, проводя измерения некоторого объекта природы или бизнеса. Собственно, поэтому название функции – true_fun. В дальнейшем эту функцию можно будет вызывать подобно функции, которая определена через def, как показано в листинге 52.

Листинг 52

```
result= true_fun(something_data)
```

Далее, для корректного проведения эксперимента необходимо создать некую выборку, которая будет служить аналогом измерений истинной закономерности. Для этого мы создадим распределение\выборку

на основе объявленной ранее функции с небольшим случайным смещением по X и Y (листинг 53).

Листинг 53

```
np.random.seed(0)
n_samples = 30
X = np.sort(np.random.rand(n_samples))
y = true_fun(X) + np.random.randn(n_samples) * 0.1
```

Полиномиальная регрессия задается в несколько шагов. Сперва необходимо определить коэффициенты полинома (которые автоматически генерируются с помощью функции `PolynomialFeatures`) и модель для основы (в нашем случае линейную или гребневую). Затем необходимо соединить вместе базовую регрессию и полиномиальные коэффициенты (вектор или матрицу преобразования) как последовательность трансформаций некоторых данных. Итоговый объект и будет являться полиномиальной моделью. Трансформацию данных можно выполнить с помощью объекта `Pipeline` из `sklearn.pipeline` (подробнее о нем можно почитать здесь: [53]).

Для удобства работы и сокращения кода сделаем следующее: объявим массив с нужными нам степенями полинома, как показано в листинге 54.

Листинг 54

```
degrees = [1, 4, 15]
```

Затем объявим функцию, которая будет принимать в качестве параметров номер степени (индекс элемента в массиве `degrees`), полотно, на котором нужно отобразить график, базовую модель и уровень графика на полотне (первый или второй ряд по вертикали). Внутри тела функции будет создаваться и обучаться нужная полиномиальная модель, результат работы которой будет выведен на возвращаемом функцией графике. Кроме точек на графике будет отображена точность модели, вычисленная методом кросс-валидации:

средние оценки качества по метрике «neg_mean_squared_error» (среднеквадратичная ошибка). Эти оценки будут получены с помощью объекта `cross_val_score` из `sklearn.model_selection`. Подробнее о кросс-валидации в `scikit` можно почитать здесь: [54, 55].

Объявление функции показано в листинге 56, с работой которого вам предлагается разобраться самостоятельно. Единственно, поясним назначение параметра `cv` в `cross_val_score`. Параметр `cv` определяет, на сколько групп будет разбита тренировочная выборка. Каждая группа будет содержать тренировочную часть А и тестовую Б. Причем эти множества (А и Б) не будут пересекаться. Так для следующих данных: входа `X = np.array([[11, 2], [3, 14], [1, 2], [3, 4]])`, выхода `y = np.array([1, 2, 3, 4])` `cross_val_score` с параметром `cv=2` даст следующее разбиение на 2 группы\итерации (листинг 55).

Листинг 55

```
TRAIN: [2 3] TEST: [0 1]
TRAIN: [0 1] TEST: [2 3]
```

Массив `TRAIN` содержит это индексы тех объектов, которые входят в тренировочную выборку. То есть `TRAIN: [2 3]` означает, что в тренировочную выборку войдут второй и третий объекты из `X`: `[1,2], [3, 4]`. Все остальные – в тестовую.

Листинг 56

```
import matplotlib.pyplot as plt
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression, Ridge
from sklearn.model_selection import cross_val_score
def MakeExample(index, plt, model,k):
    polynomial_features = PolynomialFeatures(degree=degrees[index],
                                             include_bias=False)
```

Окончание листинга 56

```
pipeline = Pipeline([("polynomial_features", polynomial_features),
                    ("linear_regression", model)])
pipeline.fit(X[:, np.newaxis], y)
scores = cross_val_score(pipeline, X[:, np.newaxis], y,
                        scoring="neg_mean_squared_error", cv=10)
X_test = np.linspace(0, 1, 100)
plt.plot(X_test, pipeline.predict(X_test[:, np.newaxis]), label="Model")
plt.plot(X_test, true_fun(X_test), label="True function")
plt.scatter(X, y, label="Samples")
if k == 1:
    plt.xlabel("x")
    plt.ylabel("y")
    plt.xlim((0, 1))
    plt.ylim((-2, 2))
    plt.legend(loc="best")
if k==0:
    plt.title("Degree {}\nMSE = {:.2e}{+/- {:.2e}"".format(degrees[i], -scores.mean(),
scores.std()))
else:
    plt.title("MSE = {:.2e}{+/- {:.2e}"".format(-scores.mean(), scores.std()))
return plt
```

Код вызова функции из листинга 56 (и, соответственно отрисовка графиков) представлен в листинге 57, а результат работы кода – на рисунке 61.

Листинг 57

```
plt.figure(figsize=(14, 6))
for i in range(len(degrees)):
    linear_regression = LinearRegression()
    plt.subplot(2, len(degrees), i+1)
    plt = MakeExample(i, plt, linear_regression,0)
for i in range(len(degrees)):
    ridge = Ridge(alpha=0.02)
    plt.subplot(2, len(degrees), len(degrees) + i + 1)
    plt = MakeExample(i, plt, ridge,1)
plt.show()
```

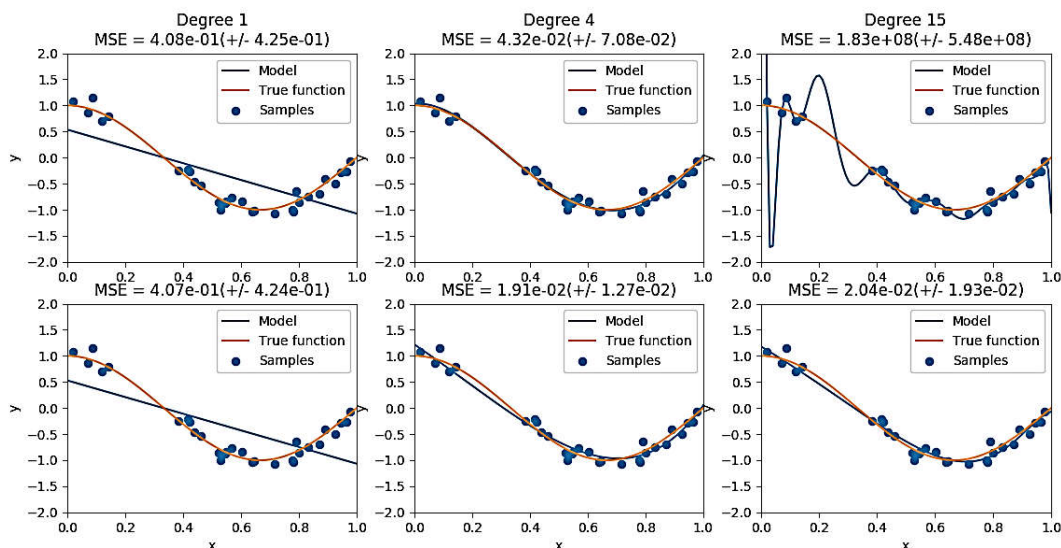


Рисунок 61. Результаты работы с разными моделями

Для лучшего понимания вам рекомендуется воспроизвести код и посмотреть на рисунок, выданный вашей программой. Если говорить о моделях, построенных на основе линейной регрессии, то, сравнивая первый и второй графики (полином 1 и 4 степени), можно сделать вывод, что повышение степени положительно влияет на качество модели. Но третий график (полином 15 степени) демонстрирует большую ошибку. Как мы видим, на тренировочной выборке модель показала хороший результат (кривая проходит через все точки выборки). Но на тестовой выборке модель «ведет себя плохо» (плохо приближает true_fun). Это и означает, что при повышении сложности модели есть риск перейти за критическую точку и переобучить модель. Однако, как мы видим на графиках гребневой регрессии, регуляризация позволяет улучшить качество решения.

Работа с простейшими моделями нейронных сетей

Следующее, с чем нам необходимо познакомиться, – работа с простейшими моделям нейронных сетей, которые содержатся в библиотеке Sklearn (в дальнейшем будет рассмотрена работа с другими библиотеками, которые содержат более сложные алгоритмы и позволяют создавать другие нейроклассификаторы и модели). Однако Sklearn не специализируется на нейросетевых моделях и не содержит более сложные алгоритмы.

Начнем мы с обычного персептрона, являющегося одним из вариантов линейных моделей. Это распространенный класс моделей, которые отличаются своей простотой и скоростью работы. Их можно обучать за разумное время на очень больших объемах данных, и при этом они могут работать с любыми типами признаков: вещественными, категориальными, разреженными. Рассмотрим решение следующей задачи: пусть у нас есть данные вида, показанного на рисунке 62, и сохраненные в файлах `perceptron-train.csv` и `perceptron-test.csv` (скачать файлы можно отсюда: [41]). Целевая переменная записана в первом столбце, признаки – во втором и третьем. Выполним следующее: применим к этим данным однослойный персептрон (о реализации модели `Perceptron` более подробно здесь: [56]), многослойный персептрон (о реализации модели `MLPClassifier` более подробно здесь: [57]), сравним точности предсказания моделей, затем нормализуем данные и повторим сравнение. Причем данные мы будем прогонять 100 раз, и на каждом прогоне выполнять 2000 эпох обучения. В качестве показателей выведем минимальные, максимальные, средние значения ассигасы и стандартное отклонение, полученные на 100 прогонах. А также выведем график, визуализирующий результаты.

1	-1.0,1.6514365371,1337.45382564
2	1.0,-0.86649520376,1191.23245707
3	-1.0,0.789828034734,-475.647767906
4	-1.0,0.179549484639,1959.09535251
5	1.0,-0.434351275619,568.504206668
6	1.0,-1.50629471392,929.584469943
7	1.0,-1.25388066775,-448.391638461
8	1.0,-0.393310851781,-1109.23094325
9	1.0,0.474347297872,-149.290941747
10	1.0,-0.882828987722,-1793.88892093
11	1.0,0.680071532867,380.219144866
12	-1.0,0.0792270140906,1648.81786149
13	-1.0,0.553856166003,371.5246818
14	-1.0,0.753868778838,3571.57921803

Рисунок 62. Данные для тестовой задачи

Поясним необходимость нормализации. Как и в случае с метрическими методами, качество линейных алгоритмов зависит от некоторых свойств данных. В частности, признаки должны иметь одина-

ковый масштаб. Если это не так, и масштаб одного признака сильно превосходит масштаб других, то качество может резко упасть.

Один из способов достижения нужного свойства признаков заключается в их стандартизации. Для этого берется набор значений признака на всех объектах, вычисляется их среднее значение и стандартное отклонение. После этого из всех значений признака вычитается среднее, и затем полученная разность делится на стандартное отклонение. В ходе нашего эксперимента мы выясним, улучшает ли нормализация качество работы модели. Код решения задачи (с комментариями) представлен в листинге 58. Результат программы, выдаваемый в консоль, представлен на рисунке 63, а построенный график – на рисунке 64.

Листинг 58

```
# -*- coding: utf-8 -*-
#Необходимые импорты
import numpy as np
import pandas
from sklearn.linear_model import Perceptron
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import accuracy_score
from sklearn.neural_network import MLPClassifier
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
#функция расчета медианы
def median(lst):
    return np.median(np.array(lst))
#Загрузка тренировочных данных
data_train = pandas.read_csv('perceptron-train.csv')
X_train = data_train.ix[:, 1:3].values
y_train = data_train.ix[:, 0].values
#Загрузка тестовых данных
data_test = pandas.read_csv('perceptron-test.csv')
X_test = data_test.ix[:, 1:3].values
y_test = data_test.ix[:, 0].values
```

Продолжение листинга 58

```
#Инициализация массива для счетчика итераций
rs = np.linspace(0,100,num=100)
#Инициализация списков для сохранения accuracy моделей
acc_p = []
acc_pn = []
acc_mlp = []
acc_mlpn = []
#Цикл прогона моделей
for i in rs:
    i = int(i)
#Распечатка номера итерации
    print "Random: ", i
#Создание модели перцептрона
    clf = Perceptron(random_state=i, alpha=0.01, n_iter=2000)
#Обучение модели
    clf.fit(X_train, y_train)
#Получение прогноза
    predictions = clf.predict(X_test)
# Расчет показателя accuracy
    acc = accuracy_score(y_test, predictions)
#Распечатка результата
    print "Perceptron: ", acc
#Добавление оценки в список оценок для модели перцептрона
    acc_p.append(acc)
#Нормализация данных
    scaler = StandardScaler()
    X_train_scaled = scaler.fit_transform(X_train)
    X_test_scaled = scaler.transform(X_test)
#Работа перцептрона с нормализованными данными
    clf = Perceptron(random_state=i, alpha=0.01, n_iter=2000)
    clf.fit(X_train_scaled, y_train)
    predictions = clf.predict(X_test_scaled)
    acc = accuracy_score(y_test, predictions)
    print "Perceptron with normalization: ", acc
    acc_pn.append(acc)
#Создание многослойного классификатора
    mlp = MLPClassifier(random_state=i, solver="sgd", activation="tanh", alpha=0.01,
hidden_layer_sizes=(2, ), max_iter=2000, tol=0.00000001)
```


Окончание листинга 58

```
mlp.fit(X_train, y_train)
#Работа с ненормализованными данными
predictions = mlp.predict(X_test)
acc = accuracy_score(y_test, predictions)
print "MLP: ", acc
acc_mlp.append(acc)
#Работа с нормализованными данными
mlp = MLPClassifier(random_state=i, solver="sgd", activation="tanh", alpha=0.01,
hidden_layer_sizes=(2, ), max_iter=2000, tol=0.00000001)
mlp.fit(X_train_scaled, y_train)
predictions = mlp.predict(X_test_scaled)
acc = accuracy_score(y_test, predictions)
print "MLPwith Norm: ", acc
acc_mlpn.append(acc)
#Распечатка итоговых результатов
print "Perceptron: ", min(acc_p), median(acc_p), max(acc_p), np.std(acc_p)
print "Perceptron with Norm: ", min(acc_pn), median(acc_pn), max(acc_pn),
np.std(acc_pn)
print "MLP: ", min(acc_mlp), median(acc_mlp), max(acc_mlp), np.std(acc_mlp)
print "MLP with Norm: ", min(acc_mlpn), median(acc_mlpn), max(acc_mlpn),
np.std(acc_mlpn)
#Расчет минимума и максимума для графика
X = np.concatenate((X_train, X_test), axis=0)
x_min, x_max = X[:, 0].min() - .5, X[:, 0].max() + .5
y_min, y_max = X[:, 1].min() - .5, X[:, 1].max() + .5
#Построение графика
figure = plt.figure(figsize=(17, 9))
cm = plt.cm.RdBu
cm_bright = ListedColormap(['#FF0000', '#0000FF'])
ax = plt.subplot(1, 1, 1)
# Точки из обучающей выборки
ax.scatter(X_train[:, 0], X_train[:, 1], c=y_train, cmap=cm_bright)
# Тестовые точки
ax.scatter(X_test[:, 0], X_test[:, 1], c=y_test, cmap=cm_bright, alpha=0.6)
ax.set_xlim(x_min, x_max)
ax.set_ylim(y_min, y_max)
ax.set_xticks(())
ax.set_yticks(())
plt.show()
```

```
Random: 100
Perceptron: 0.743718592965
Perceptron with normalization: 0.894472361809
MLP: 0.653266331658
MLP with norm: 0.889447236181

Perceptron: 0.35175879397 0.683417085427 0.894472361809 0.130098348038
Perceptron with Norm: 0.251256281407 0.793969849246 0.899497487437 0.12811275683
MLP: 0.346733668342 0.653266331658 0.668341708543 0.0370981920263
MLP with Norm: 0.874371859296 0.889447236181 0.899497487437 0.00389763204616
```

Рисунок 63. Результат работы программы

Воспроизведите код, посмотрите на полученный график. Он должен быть похож на изображенный на рисунке 64. Обратим внимание на один важный технический момент. На графике можно увидеть, что принадлежность к классу (то есть по сути Y) отображается не отдельным измерением, а цветом. Такой механизм позволяет рисовать двумерные графики вместо трехмерных, что проще и нагляднее.

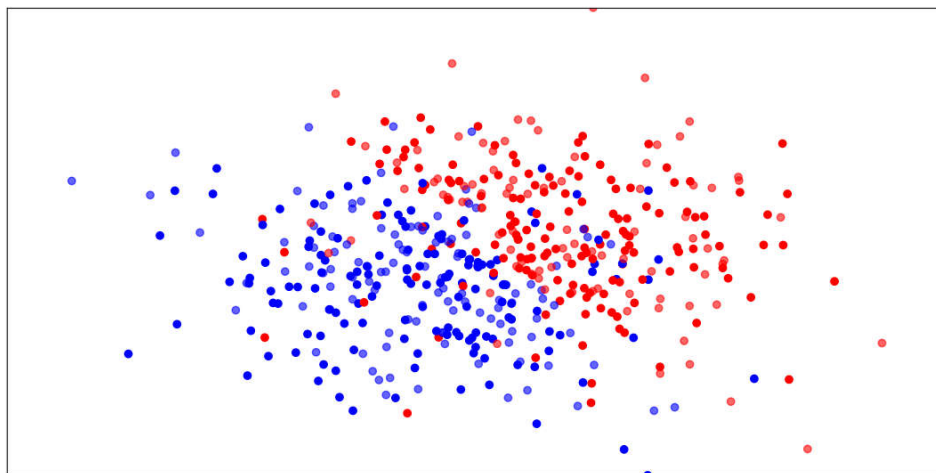


Рисунок 64. Итоговый график

Содержательный анализ результатов работы программы и доработку кода вам предлагается выполнить самостоятельно (см. практическое задание «Многослойный персептрон»).

Реализация алгоритма обучения нейронной сети

Данный пункт посвящен алгоритму обратного распространения ошибки как классическому подходу к обучению нейронных сетей. Поскольку нейронные сети сегодня – это важный класс моделей для построения сложных иерархических признаков, то возникает необходимость детального понимания протекающих в них процессов. Важность именно понимания нейронных сетей обуславливается еще тем, что сложнейшие на сегодняшний день сети еще плохо описаны математически, и требуется проделать много исследовательской работы в этом направлении.

Задача состоит в том, чтобы реализовать обучение полносвязной трехслойной сети прямого распространения на каких-то простых синтетических данных.

Если вы собираетесь заниматься именно исследованием в ML, то подобного рода задачи – это ваша обычная работа. Если вы собираетесь заниматься решением конкретных проблем бизнеса, то с такими задачами вам сталкиваться вряд ли придется, и основная работа будет посвящена подготовке и интерпретации данных. Подготовка данных было уже уделено много внимания, поэтому сейчас рассмотрим исследовательскую задачу. Однако прежде чем перейти непосредственно к ее реализации, рассмотрим применение матричной записи данных в ней.

Так, если вы хотите взвесить некоторые значения X по некоторым весам W и получить сумму, то не обязательно писать выражение $w_1 \times x_1 + w_2 \times x_2$. Ту же самую операцию можно реализовать в виде матричного умножения. Первая матрица A будет представлять из себя X , где столбцы – признаки, а строки – традиционно объекты (если вы обрабатываете один объект за раз, то это может быть матрица с одной строкой). Вторая матрица B будет представлять из себя матрицу весов, где каждая колонка будет отождествляться с нейроном следующего слоя, а каждая строка будет отвечать за конкретный вес (значение веса). Однако даже матрица с одной строкой

в Python будет представлена как двумерный массив. Вы можете поэкспериментировать самостоятельно, выводя различные массивы\матрицы в консоль. После перемножения таких двух матриц получим новую матрицу C, где каждая строка будет представлять собой взвешенный набор признаков $X(x_1, \dots, x_n)$ по $W(w_1, \dots, w_n)$. Схематично это выглядит так, как показано на рисунке 65 [1].

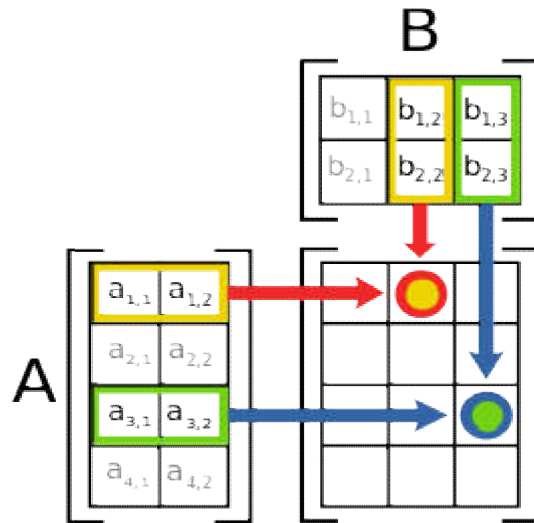


Рисунок 65. Умножение матриц

Необходимо отметить, что взвешивание значений с точки зрения матрицы происходит так, что из матрицы A берутся строки, а из матрицы B берутся колонки. Поэтому иногда массивы\матрицы нужно будет транспонировать. Напомним суть операции транспонирования, приведя в качестве примера рисунок 66.

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \text{и} \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Рисунок 66. Транспонирование матриц

Объявления функций активации и ее производной представлены в листинге 59.

Листинг 59

```
# -*- coding: utf-8 -*-
#Необходимый импорт
import numpy as np
#Определяем функцию активации
def activation(x):
    return 1 / (1 + np.exp(-x))
#Определяем производную от функции активации
def sigma_derivative(x):
    return x * (1 - x)
```

Реализация алгоритма обраного распространения ошибки с подробными комментариями приведена в листинге 60. Забегая вперед, скажем, что в практической части вам предлагается задача модификации данного кода и интерпретации его результатов.

Листинг 60

```
# X - это матрица, где столбцы это признаки,
# а строки это объекты выборки.
X = np.array([[0, 0, 1],
              [0, 1, 0],
              [1, 0, 0],
              [1, 1, 1]])
# Y - это матрица, где столбцы это признаки
# (в данном случае один целевой признак),
# а строки это объекты выборки.
y = np.array([[0],
              [1],
              [1],
              [0]])
# Зафиксируем генератор случайных чисел, чтобы получать каждый раз
#предсказуемый (одинаковый) результат
np.random.seed(4)
# Нотация слоев - L1, L2, L3.
# В каждом слое есть n1, n2, n3 нейронов.
# Строки в матрицах индексируются как i, столбцы как j.
# W_1_2 - это матрица весов между первым и вторым слоем.
# Строки определяют левый нейрон (т.е. нейрон первого слоя) и соответственно
количество строк равно n1.
```

Продолжение листинга 60

```
# Столбцы определяют правый нейрон (т.е. нейрон второго слоя) и соответственно
# количество столбцов равно n2.
# На пересечении i-й строки и j-го столбца получаем ячейку с конкретным весом,
# который
# связывает i-ый левый нейрон (слоя L1) и j-й правый нейрон (слоя L2)
W_1_2 = 2 * np.random.random((3, 2)) - 1
# W_2_3 - это матрица весов между вторым и третьим слоем. Все остальное как в
W_1_2
W_2_3 = 2 * np.random.random((2, 1)) - 1
# скорость движения по антиградиенту
speed = 1.1
# цикл прогона модели
for j in range(100000):
    # I0, I1, I2 - матрицы определенного слоя сети. Каждая строка матрицы это реакция
# на i-й объект входа. Каждая колонка матрицы это реакция j-го нейрона
# соответствующего слоя на разные входные образы.
    # На пересечении i-й строки и j-го столбца получаем ячейку с конкретной
# реакцией j-го нейрона на конкретный вход.
    # Первый слой нейронов полностью принимает значения входа X (т.е. все реакции
# единичные).
    I1 = X
    # Второй слой нейронов рассчитывается как функция активации по каждому
# элементу матрицы U
    # По сути I2 это матрица 4 на 4, где в каждой ячейке результат активации нейрона
# второго слоя для всех 4-х входных образов
    # Детально:
    # матрица U = np.dot(I0, W_0_1) является результатом
    # матричного произведения выходов нейронов предыдущего слоя на веса между 1
# и 2 слоем.
    # Строки матрицы U отвечают за конкретный входной образ (объект).
    # Столбцы матрицы U отвечают за нейроны правого слоя (L2).
    # На пересечении i-й строки и j-го столбца получаем ячейку с конкретной
# взвешенной суммой
    # для i-го входного образа и j-го нейрона слоя (I2). Иными словами, для каждого
# нейрона слоя L2 и для каждого входа
    # считается  $U = W1X1 + W2X2 + W3X3$  (поскольку входной нейрон содержит 3
# нейрона).
    I2 = activation(np.dot(I1, W_1_2))
    # тоже самое проделываем для Третьего слоя
```

Продолжение листинга 60

```
# По сути I3 это матрица 4 на 1, где в каждой ячейке результат активации нейрона
#третьего слоя (а он один)
# для всех 4-х входных образов
I3 = activation(np.dot(I2, W_2_3))
# рассчитывает ошибку на выходе
I3_error = y - I3
# рассчитывает модуль средней ошибки
if (j % 10000) == 0:
    print "Error:" + str(np.mean(np.abs(I3_error)))
# sigma - есть локальный градиент ошибки
# I3_sigma рассчитывается как ошибка выхода всей сети на производную функции
#активации всех нейронов L3.
# На пересечении i-й строки и j-го столбца получаем ячейку с конкретной сигмой
# для i-го входного образа и j-го нейрона слоя (I3). Иными словами, для каждого
#нейрона слоя L3
# и для каждого входа рассчитывается производная по функции активации
#умноженная на ошибку.
# То есть I3_sigma будет матрица 4 на 1 (поскольку в L3 только один нейрон).
I3_sigma = I3_error * sigma_derivative(I3)
print I3_sigma
# Ошибка L2 слоя оценивается через взвешенную сигму слоя L3 по весам между
#L2 и L3
# Поскольку I3_sigma это матрица 4 на 1 и матрица W_2_3 4 на 1, то последнюю
#матрицу
# надо Транспонировать, чтобы выполнить правило умножения матриц и взвесить
#элементы I3_sigma
# по элементам W_2_3.
# Тогда итоговая матрица I1_error будет 4 на 4, где на пересечении i-й строки и j-го
#столбца
# будет ячейка с конкретной ошибкой j-го нейрона слоя L2 для i-го входного
#образа.
I2_error = I3_sigma.dot(W_2_3.T)
print I2_error
# I2_sigma рассчитывается как ошибка слоя L2 на производную функции активации
#всех нейронов L2. На пересечении i-й строки и j-го столбца получаем ячейку с
#конкретной сигмой для i-го входного образа и j-го нейрона слоя (L2). Иными
#словами, для каждого нейрона слоя L2 и для каждого входа рассчитывается
#производная по функции активации, умноженная на ошибку.
```

Окончание листинга 60

```
# То есть l2_sigma будет матрица 4 на 4 (поскольку в L2 четыре нейрона).
l2_sigma = l2_error * sigma_derivative(l2)
print l2_sigma
# обновляем веса
# l2 это матрица 4 на 4, где в каждой ячейке результат активации нейрона второго
#слоя для всех 4-х входных образов (по строкам).
# А l3_sigma это матрица 4 на 1, где в каждой строке локальный градиент ошибки
#для 4-х входных образов.
# Чтобы взвесить столбец матрицы l2 по локальному градиенту (l3_sigma), нужно
#транспонировать матрицу l2, чтобы
# результат активации каждого нейрона в слое L2 по каждому входному образу
#вытянулся в строку.
# Тогда соответствующая строка умножится на столбик l3_sigma.
# Заметьте, что транспонирование l3_sigma не даст результата, так как тогда в
#каждом столбце
# l3_sigma будет по одному значению, а в каждой строке l2 по 4 значения, а значит
#такое перемножение матриц не пройдет. Применяем транспонирование к l2.
W_2_3 += speed * l2.T.dot(l3_sigma)
# аналогично W_2_3
W_1_2 += speed * l1.T.dot(l2_sigma)
# Прямое распространение для тестовых данных
X_test = np.array([[0, 0, 0],
                   [0, 1, 1],
                   [1, 0, 1],
                   [1, 1, 0],
                   [0.5, 0.5, 0],
                   [0.5, 0.5, 1]])
# Y_test должен получиться [1, 0, 0, 1, 1, 0]
l1 = X_test
l2 = activation(np.dot(l1, W_1_2))
l3 = activation(np.dot(l2, W_2_3))
print l3
```

Регуляризация и сеть прямого распространения

В отличие от предыдущих пунктов здесь не будет кода целиком, а только общие рекомендации по выполнению экспериментов, позволяющих выявить влияние регуляризации на многослойную сеть прямого распространения. В ходе работы над пунктом вы научитесь

генерировать синтетические данные, проводить серии экспериментов, анализировать возможности моделей на разнотипных данных и сравнивать результаты экспериментов.

Необходимые для выполнения работы операторы `import` представлены в листинге 61.

Листинг 61

```
import numpy as np
from matplotlib import pyplot as plt
from matplotlib.colors import ListedColormap
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.datasets import make_moons, make_circles, make_classification
from sklearn.neural_network import MLPClassifier
```

Первый шаг, который вам необходимо выполнить, – генерация данных. Библиотека Scikit-learn содержит множество наборов данных (`sklearn.datasets`). Среди этого набора есть реальные данные (выборки, которые можно загрузить при помощи функций – Loaders), так и специальные генераторы синтетических данных – Samplesgenerator. Подобные генераторы позволяют создать данные на лету по заданными распределениям, функциям, с возможностью добавления шума, случайных отклонений, смещений и т. п.

Есть несколько классических сценариев для проверки возможностей моделей в машинном обучении. Мы воспользуемся одним генератором линейных задач (`datasets.make_classification`) и двумя нелинейными генераторами (`make_moons`, `make_circles`). Отметим, что одной из важнейших характеристик данных является размерность. Важно понимать, что величина размерности данных соответствует количеству признаков, которые описывают исходные объекты. То есть каждая ось декартового пространства закрепляется за одним количественным признаком. Не все генераторы данных позволяют задавать размерность данных. Тем ни менее ниже приведены крайне важные распределения данных, которые иллюстрируют разные типы задач в машинном обучении. Соответственно тестирование моделей

на подобных данных позволяет без наличия реальных данных проверять какие-либо гипотезы.

Пример генераторов:

- `make_classification` – позволяет сгенерировать такие классы-признаки, которые будут линейно разделяться в n -мерном пространстве (см. рисунок 67. а).
- `make_circles` – позволяет сгенерировать такие классы-признаки, что признаки одного класса геометрически окружают признаки другого класса (см. рисунок 67.б).
- `make_moons` – позволяет сгенерировать такие классы-признаки, что признаки одного класса частично пересекаются с признаками другого класса (см. рисунок 67. в).

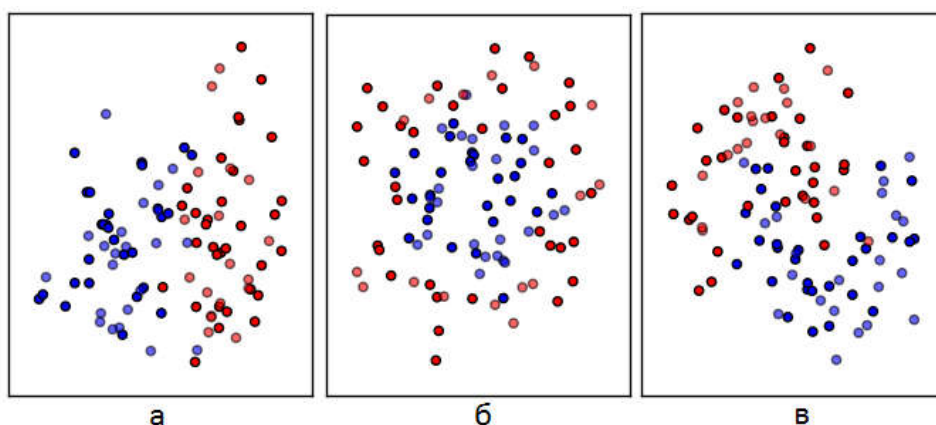


Рисунок 67. Виды распределений признаков, сгенерированных инструментами Scikit-learn

Самостоятельно найдите в Scikit информацию по различным генераторам и сгенерируйте синтетические данные для проведения экспериментов. Пример кода для создания данных трех разнотипных задач классификации приведен в листинге 62.

Для удобства дальнейшей работы рекомендуется объединить все наборы данных в список, как показано в листинге 63, тогда вы сможете организовать цикл по каждому элементу этого списка (что приведено в коде). В этом же листинге есть пример масштабирования данных и разбиения их на тренировочную тестовую выборку.

Листинг 62

```
X, y = make_classification(n_features=2, n_redundant=0, n_informative=2,
random_state=0, n_clusters_per_class=1)
rng = np.random.RandomState(2)
X += 2 * rng.uniform(size=X.shape)
linearly_dataset = (X, y)
moon_dataset = make_moons(noise=0.3, random_state=0)
circles_dataset = make_circles(noise=0.2, factor=0.5, random_state=1)
```

Листинг 63

```
datasets = [moon_dataset, circles_dataset, linearly_dataset]
for ds_cnt, ds in enumerate(datasets):
    X, y = ds
    X = StandardScaler().fit_transform(X)
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.4, random_state=42)
```

Функция `train_test_split` делит данные так, что тестовая выборка составляет 40% от исходного набора данных. Разделение происходит случайным образом (т.е. элементы берутся из исходной выборки не последовательно). В этом же цикле можно организовать обучение моделей, их тестирование и формирования графиков.

В теле цикла создайте коллекцию многослойных сетей (`MLPClassifier`) с различными коэффициентами регуляризации (`alpha`) и обучите каждую модель на каждом типе данных. Всего должно получиться 15 моделей (3 типа данных\задачи и 5 видов коэффициентов регуляризации). Для реализации используйте код из листинга 64.

Листинг 64

```
alphas = np.logspace(-5, 3, 5)
```

Следующим шагом отобразите исходные данные и результаты на графиках. Пример отображения исходных данных уже был приведен в предыдущих пунктах. Однако на этот раз необходимо построить не один график, а серию графиков в одном окне. В итоге должна получиться таблица из графиков (3 строки, 6 столбцов). Для этого изучите функцию `subplot` из библиотеки `matplotlib`. Пример кода для инициализации нового графика (если точнее, то подграфика в рамках одного полотна\окна) показан в листинге 65.

Листинг 65

```
current_subplot= plt.subplot(3, 5 + 1, i)
```

Показанным в листинге 65 способом мы можем получить один из графиков в полотне и заполнять его данными, а потом таким же образом получить следующий график (то есть объект `current_subplot`). В первую очередь на всех графиках необходимо отобразить исходные данные (точки\признаки). Можете отобразить тренировочную и тестовую выборку или только одну. Пример кода, в котором точки тестовой выборки рисуются полупрозрачными (`alpha=0.6`), приведен в листинге 66.

Листинг 66

```
current_subplot.scatter(X_train[:, 0], X_train[:, 1], c=y_train, cmap=cm_bright)
current_subplot.scatter(X_test[:, 0], X_test[:, 1], c=y_test, cmap=cm_bright,alpha=0.6)
```

В листинге 67 `cm` и `cm_bright` это цветовые схемы для отрисовки графиков. Их можно взять как из готового набора, так и задать вручную. Это делается следующим образом (листинг 65):

Листинг 67

```
cm = plt.cm.RdBu
cm_bright = ListedColormap(['#FF0000', '#0000FF'])
```

При отрисовке графиков рекомендуется придерживаться следующего. Пусть левый столбец графиков отображает только исходные данные и ничего больше. Остальные же столбцы должны отображать функции обученных моделей в пространстве исходных данных. Это необходимо, чтобы визуально оценить, как каждая модель разделила исходное пространство признаков. Вместо построения значений функции Y на отдельной оси отобразите значения Y в виде цветовых градиентов (чтобы не строить трехмерные графики). Пример визуализации показан на рисунке 68.

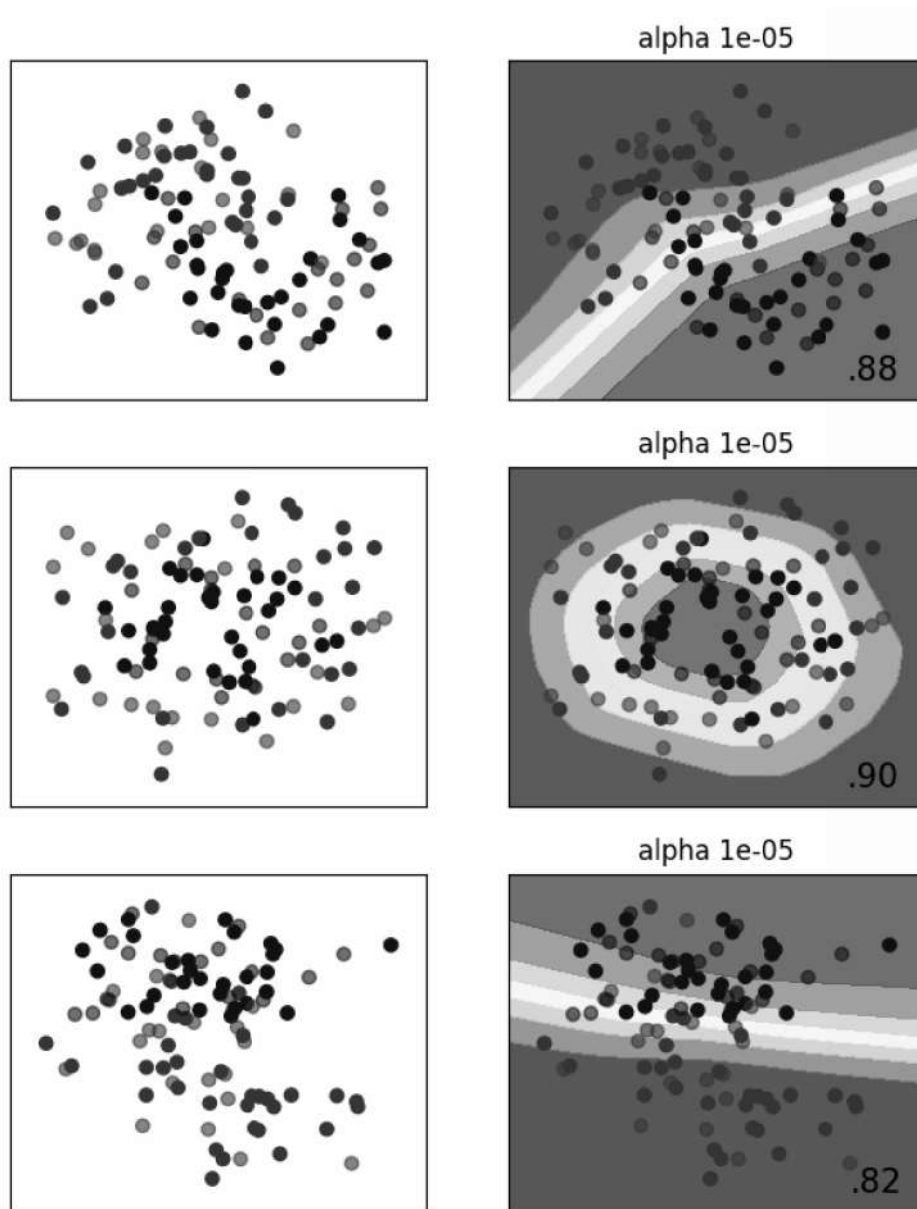


Рисунок 68. Пример визуализации качества моделей

Здесь слева представлена колонка с тремя графиками исходных распределений признаков (на каждом графике две группы точек исходных данных). Каждая строка отвечает за свою задачу классификации (moondataset, circledataset, linearseparabledataset). Правая колонка демонстрирует как конкретная модель многослойной сети MLP (с параметром $\alpha=0.00001$) классифицирует исходное распределение. Значения модели в каждой точке признаков представлены цветовым градиентом. Это гораздо более удобный способ визуализации, чем построение трехмерных поверхностей.

Как мы видим, правый столбец отображает градиент. Чтобы получить градиент, необходимо определить множество точек, по которым будет строиться график. Это можно сделать при помощи регулярной сетки, как показано в листинге 68.

Листинг 68

```
h = .02 #шаг регулярной сетки
x0_min, x0_max = X[:, 0].min() - .5, X[:, 0].max() + .5
x1_min, x1_max = X[:, 1].min() - .5, X[:, 1].max() + .5
xx0, xx1 = np.meshgrid(np.arange(x0_min, x0_max, h), np.arange(x1_min, x1_max, h))
```

В листинге 69 представлено несколько вариантов кода для получения значений функции модели в указанных точках сетки.

Листинг 69

```
Z = clf.decision_function(np.c_[xx0.ravel(), xx1.ravel()])#1
Z = clf.predict_proba(np.c_[xx0.ravel(), xx1.ravel()])[:, 1]#2
Z = clf.predict(np.c_[xx0.ravel(), xx1.ravel()])#3
```

Сделаем некоторые пояснения к коду:

1. `decision_function` возвращает расстояние до разделяющей гиперплоскости в соответствующем измерении (проекционной оси).
2. `predict_proba` возвращает вероятность принадлежности к каждому классу (так для двух классов можно брать одну из вероятностей (один из столбцов двумерного массива) и исходя из значения раскрашивать область в соответствующий цвет).
3. `Predict` возвращается целочисленную оценку принадлежности к классу (как бинарная маска).

Третий вариант в данном случае выдаст округленные значения классов (0, 1 и т. д.), а функция `predict_proba` выдает сглаженную функцию вероятности класса. Чтобы понять разницу, попробуйте различные варианты.

Однако не все классификаторы имеют `decision_function`. Чтобы узнать, содержит ли текущий объект интересующую нас функцию, можно воспользоваться методом `hasattr`. Этот метод возвращает флаг,

указывающий на то, содержит ли объект указанный атрибут и нужен для обработки разных вариантов классификаторов (можно выстроить на этом if-else логику). Пример ее вызова показан в листинге 70.

Листинг 70

```
hasattr(clf, "decision_function")
```

Далее вам необходимо нарисовать результат, оси, заголовок и текст со значением точности модели. Выполнить это можно с помощью кода из листинга 71.

Листинг 71

```
Z = Z.reshape(xx0.shape)
current_subplot.contourf(xx0, xx1, Z, cmap=cm, alpha=.8)
current_subplot.set_xlim(xx0.min(),xx0.max())
current_subplot.set_ylim(xx0.min(),xx1.max())
current_subplot.set_xticks(())
current_subplot.set_yticks(())
current_subplot.set_title(name)
current_subplot.text(xx0.max() - .3, xx1.min() + .3, ('%.2f' % score).lstrip('0'),
size=15, horizontalalignment='right')
```

После отрисовки градиента добавьте на каждый график тестовую выборку (точки из каждой задачи для графика каждой модели). Это можно сделать таким же образом, как показано в листинге 72.

Листинг 72

```
current_subplot.scatter(X_test[:, 0], X_test[:, 1], c=y_test, cmap=cm_bright,alpha=0.6)
```

Не забудьте, что после построения каждого графика необходимо инкрементировать переменную *i* для функции subplot(), а также помните о необходимости применить функцию plt.show() в конце всех операций формирования графиков для вывода результата в окне ОС.

При желании вы можете настроить отступы графиков в окне просмотра при помощи следующего кода (листинг 73).

Листинг 73

```
figure.subplots_adjust(left=.02, right=.98)
```

Для задания заголовков и вывода текста на графиков можно использовать следующие функции из листинга 74.

Листинг 74

```
current_subplot.set_title("Input data")  
current_subplot.text(xx0.max() - .3, xx1.min() + .3, ('%.2f' % score).rstrip('0'), size=15,  
horizontalalignment='right')
```

Работа с библиотеками Keras и Theano. Настройка под Windows

Мы закончили рассмотрение простых моделей нейронных сетей и переходим к более сложным. В частности, к сверточным сетям. Однако поскольку в библиотеке scikit нет полноценной поддержки таких сетей, то в данном пункте мы рассмотрим установку и настройку библиотек Keras и Theano, потому что для пользователей Windows установка этих библиотек нетривиальна.

Для разворачивания Keras и Theano для CPU на Windows выполните следующее:

1. Откройте Anaconda prompt (командную строку Анаконды через меню «Пуск» или просто выполните стандартную команду cmd. Если все установлено верно, то Anaconda «перехватит» адресованные ей команды)
2. Выполните `conda update conda`
3. Выполните `conda update --all`
4. Выполните `conda install mingw libpython`
5. Установите Theano, `pip install Theano`
6. Установите Keras `pip install keras`
7. Выполните код для проверки корректности всех действий (он показан в листинге 75).

Листинг 75

```
import numpy as np
import sklearn.linear_model as lin
from keras.layers.core import Dense, Activation
from keras.models import Sequential
from keras.optimizers import RMSprop

x_train = np.array([[0, 0], [0, 1], [1, 0], [1, 1]])
y_train = np.array([0, 1, 1, 0]).reshape(4, 1)
x_test = np.array([[0, 0], [0, 1], [1, 0], [1, 1]])

model = Sequential()
model.add(Dense(2, init='lecun_uniform', input_shape=(2,)))
model.add(Activation('relu'))
model.add(Dense(1, init='lecun_uniform'))
model.add(Activation('linear'))
rms = RMSprop(lr=0.01)
model.compile(loss='mse', optimizer=rms)

epochs = 200
model.fit(x_train, y_train, batch_size=1, nb_epoch=epochs, verbose=0)
y_predict = model.predict(x_test, batch_size=1)
print y_predict
```

Теперь опишем разворачивание Theano для GPU на Windows. Ведь обучение даже относительно простых сверточных сетей – крайне затратный вычислительный процесс, который может занять несколько часов на довольно мощном процессоре (здесь также стоит учитывать, что не все библиотеки и платформы поддерживают параллельное исполнение кода между ядрами процессоров, а значит, попросту не используют всю вычислительную мощь CPU).

Современные графические карты позволяют запускать очень много параллельных вычислений на аппаратном (а не только на программном) уровне. Лидером таких вычислений на сегодняшний день является компания NVidia, которая предлагает специальную платформу под названием CUDA. Это программно-аппаратная платформа, которая встроена в видеокарты и драйвера NVidia. Подобная

технология может ускорить вычисления во много раз (от 4 до 100). Конечно же, GPU вариант предпочтительнее не только в реальных проектах и промышленных системах, но и в процессе обучения, так как может существенно сэкономить время на эксперименты и выполнение учебных заданий.

Библиотека Theano может исполнять свои модели на графическом процессоре (GPU), который установлен в системе. Однако поддерживаются только графические процессоры от NVidia и только те, которые поддерживают режим `CUDAComputeCapability 3.0` или выше. Узнать о том, какой режим поддерживает ваша видеокарта можно на официальном сайте NVidia.

Само разворачивание Theano состоит из следующих шагов:

1. Проверьте, что у вас есть свежий драйвер на видеокарту Nvidia. Необходима именно карта `NvidiacCompute Capablity = 3.0` или выше (проверить `Compute Capablity` своей видеокарты можно тут [58]). Если карты от Nvidia нет, то этот сценарий реализовать не получится и просто используйте CPU.

2. Скачайте и установите `CUDA Toolkit 8.0` отсюда: [59].

3. Скачайте `Cudnn 5.0` для `CUDA Toolkit 8.0` отсюда [60]. Это библиотека для низкоуровневой поддержки нейронных сетей. Распакуйте архив `Cudnn`. Там будет несколько папок. В каждой папке по одному файлу. Эти файлы нужно переместить в место, куда был установлен `CUDA Toolkit 8.0`. Например, файл `cudnn64_5.dll` из папки `cuda\bin` положить в папку `bin` в место, куда установился `CUDA Toolkit 8.0` и так далее.

4. Теперь код Theano может запускаться на GPU. Но его еще нужно скомпилировать. Для компиляции нужен `VC++` компилятор от Майкрософта. Нужную версию компилятора можно получить, установив `Visual Studio 2012`.

5. Создайте текстовый файл с именем `.theanorc`, заполненный как показано в листинге 76. Описание приведено в листинге 77.

6. Созданный файл нужно положить в домашнюю или пользовательскую директорию ОС. Для Win это C:\Users\\$UserName\$. Чтобы четко определить директорию, можно выполнить следующий код (листинг 78).

Листинг 76

```
[global]
floatX=float32
device=gpu
mode=FAST_RUN
cnmem=0.7
[cuda]
root=C:\Program Files\NVIDIA GPU Computing Toolkit\CUDA\v7.5
[nvcc]
flags=-LD:\Program_Files\Anaconda\libs
-compiler_bindir=C:\Program_Files_(x86)\Microsoft_Visual_Studio_11.0\VC\bin\
-fastmath=True
```

Листинг 77

```
[global]
Cnmem - объем памяти, который используется на видео карте
[cuda]
Root - путь на корневую папку CUDAtoolkit. Например:
«root=C:\Program Files\NVIDIA GPU Computing Toolkit\CUDA\v8.0»
[nvcc]
Flags - указывает путь до библиотек Anaconda
compiler_bindir - путь до компилятора cl.exe из VisualStudio
Во всех параметрах [nvcc] не должно быть пробелов. Так «Program Files (x86)» нужно
заменить на «Program_Files_(x86)». Пример пути:
«D:\Program_Files\Visual_Studio_1\VC\bin\»
```

Листинг 78

```
os.environ['THEANO_FLAGS']="device=gpu0,lib.cnmem=0.5"
path=os.path.expanduser('~/.theanorc.txt')
```

7. Выполните проверочный код из листинга 79. Он должен вывести 'Used the gpu' и нулевой код ошибки.

Листинг 79

```
from theano import function, config, shared, sandbox
import theano.tensor as T
import numpy
import time
vlen = 10 * 30 * 768 # 10 x #cores x # threads per core
iters = 10000
rng = numpy.random.RandomState(22)
x = shared(numpy.asarray(rng.rand(vlen), config.floatX))
f = function([], T.exp(x))
print(f.make_fgraph().toposort())
t0 = time.time()
for i in range(iters):
    r = f()
t1 = time.time()
print("Looping %d times took %f seconds" % (iters, t1 - t0))
print("Result is %s" % (r,))
if numpy.any([isinstance(x.op, T.Elemwise) for x in f.make_fgraph().toposort()]):
    print('Used the cpu')
else:
    print('Used the gpu')
```

Получение данных средствами Keras

Подобно библиотеке Scikit-learn библиотека Keras также имеет встроенные функции для получения готового набора данных. Пример показан в листинге 80. На выходе у нас получаются стандартные массивы Numpy.

Листинг 80

```
from keras.datasets import mnist
(X_train, y_train), (X_test, y_test) = mnist.load_data()
```

Забежав немного вперед, скажем, что сверточные модели, представленные библиотекой Keras, принимают на вход немного непривычный формат данных для обучения. Так параметр `Input_shape` функции `model.fit()` имеет следующую форму: `(samples, channels, rows, cols)`. Здесь `samples` – количество элементов в выборке (может

не указываться), `channels` – количество каналов (если речь идет об RGB-изображениях, то на каждую компоненту будет по матрице и этот параметр будет равен трем), `rows` и `cols` соответственно ширина и высота матрицы (изображения).

Исходя из этого, нам необходимо модифицировать полученные данные. Чтобы сменить форму массива Numpy, выданного `load_data`, воспользуйтесь функцией `reshape`, которая в качестве параметров принимает размер каждого измерения. Пример функции показан в листинге 81.

Листинг 81

```
X_train = X_train.Reshape(N, 1, r, c)
```

Кроме смены формы массива его элементы необходимо привести в диапазон от 0 до 1 для корректной работы с ним сверточной нейронной сетью. Сделайте это самостоятельно.

Над выборкой `Y` также необходимо провести преобразования. Целое число нужно перевести в бинарную маску (вектор) так, чтобы класс 2 из трех возможных классов представлял собой вектор `[0,0,1]` при отсчете от нуля. Подобное преобразование можно сделать при помощи функции `np_utils.to_categorical`.

Создание и обучение модели сверточной сети

Общая логика обучения модели такая же, как и в моделях библиотеки `Scikit-Learn`. Но поскольку `Keras` больше ориентирована на нейросетевые модели, то она позволяет более детально настраивать структуру и работу сети. В этом пункте мы рассмотрим лишь некоторые из возможностей.

Модели в `Keras` описываются как вычислительные графы. Поэтому, чтобы создать и обучить сверточную сеть, сначала необходимо создать последовательную модель (граф последовательных вычислений\обработчиков, «контейнер моделей»). Сделать это можно так, как показано в листинге 82.

Листинг 82

```
model = Sequential()
```

После создания граф необходимо заполнить рядом обработчиков. Описание всех возможных обработчиков есть в документации Keras, например, здесь: [61]. Ниже представлен список основных обработчиков, с которыми мы столкнемся в этом пункте.

Dense – одномерный слой обычных нейронов. По сути, этот обработчик вычисляет взвешенные суммы для каждого нейрона слоя, но не вычисляет для них функцию активации (так как она задается отдельным обработчиком). Функция, создающая обработчик, принимает на вход число нейронов. Использовать ее можно так, как показано в листинге 83.

Листинг 83

```
Dense(128)
```

ConvolutionND – сверточный слой нейронов. Данный обработчик подсчитывает взвешенные суммы для каждой карты и каждого нейрона в слое, но также не вычисляет для них функцию активации. Функция, создающая обработчик принимает на вход число карт в слое и размер ядра\фильтра, также определяется форма входа и режим обработки границ (valid означает, что все ядра поместятся в исходную матрицу и не будет выступающих границ, при этом карта будет меньше исходной матрицы; в случае same карта будет такого же размера, как и входная матрица). Можно задать и ряд других параметров, но с ними вам предлагается ознакомиться самостоятельно. Пример использования приведен в листинге 84.

Activation – обработчик функции активации. Задается после каждого слоя нейронов (Convolution или Dense), такие обработчики, как MaxPooling2D или Dropout, не являются слоями нейронов. Данный обработчик рассчитывает значение активации для взвешенных сумм предыдущего слоя. В качестве аргумента функция прини-

мает строку с названием собственно функции активации (relu, tanh, linear, sig, prelu). Пример использования – в листинге 85.

Листинг 84

```
Convolution2D(nb_filters, nb_conv, nb_conv,  
border_mode='valid',  
input_shape=(1, img_rows, img_cols)))
```

Листинг 85

```
Activation('relu')
```

MaxPoolingND – обработчик объединения по максимуму. Обработчик проходит по матрице, поступающей на вход, окном размера pool_size (например, 2X2) и подает на выходную матрицу одно максимальное значение из данной области входной матрицы. Иллюстрация работы функции – рисунок 69 [1].

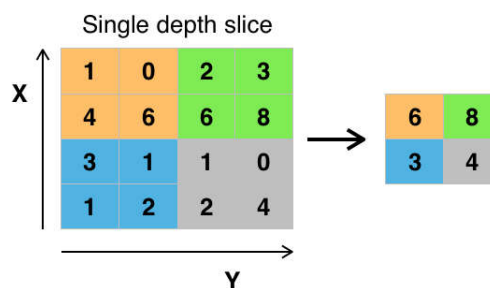


Рисунок 69. Иллюстрация действия обработчика MaxPooling2D

Как мы видим, из входной матрицы 4X4 формируется матрица 2X2 максимальных значений. Пример вызова функции приведен в листинге 86.

Листинг 86

```
MaxPooling2D(pool_size=(nb_pool, nb_pool))
```

Dropout – данный обработчик задает степень разрыва нейронных связей между соседними слоями. В качестве параметра функция принимает значение от 0 до 1, выражающее процент разрыва. В данном случае 25% связей между слоями выродятся в ноль и не будут меняться в процессе обучения.

Обратите внимание!

- *Такой возможности не было в библиотеке scikit-learn. Это важная настройка, так как она позволяет контролировать переобучение сети не только регуляризацией. Кроме того, таким способом можно добиться эффекта локальности.*

Вызов обработчика – в листинге 87.

Листинг 87

```
Dropout(0.25)
```

Flatten – выравнивающий обработчик. Преобразует связи от нейронов слоя свертки к одномерному слою обычных нейронов. Такой обработчик используется ближе к завершению сети, чтобы формировать выход модели. Пример вызова – в листинге 88.

Листинг 88

```
Flatten()
```

Если создать модель из двух сверточных слоев с функцией активации «relu», затем перейти к обычным слоям, добавить разрывы и завершите модель слоем из 10 нейронов с функцией активации «softmax», то данная функция сформирует выходной вектор таким образом, чтобы сумма всех элементов была равна единице, а отдельный элемент вектора выражал вероятность принадлежности входного образа к классу, который отображается этим элементом вектора.

Чтобы добавить тот или иной обработчик в вычислительный граф, используется функция add, как показано в листинге 89.

После формирования графа необходимо скомпилировать модель. Причем то же самое придется выполнить и после загрузки модели из файла, поскольку сохраняется лишь граф вычислений, а не готовый к исполнению объект. Пример компиляции модели приведен в листинге 90.

Листинг 89

```
model.add(<обработчик>)
```

Листинг 90

```
model.compile(loss='categorical_crossentropy',  
optimizer='adadelta',  
metrics=['accuracy'])
```

Обучается модель уже знакомым способом (с помощью функции `fit`, куда передаются тренировочные выборки X , Y). Однако в данных моделях имеет смысл рассмотреть еще несколько параметров:

- `batch_size` – определяет количество предъявляемых образов в рамках одного обновления весов (так если `batch_size > 1`, то производится обучение по группе, а не по одиночным образам). Данный параметр может повлиять на скорость обучения (особенно на CPU) и на объем требуемой памяти. Так небольшие группы по 50 образов могут потребовать меньше памяти и вычислительных ресурсов (слово «могут» используется потому, что есть и другие факторы, в том числе аппаратная реализация некоторых функций и поведение кэша).

- `nb_epoch` – определяет количество эпох обучения.

- `verbose` – определяет, выводить ли в консоль детали обучения или нет. Под деталями понимается информация о ходе обучения, времени и оценке качества за каждую итерацию.

- `validation_data` – определяет набор данных, по которому будет проводиться оценка качества модели.

Пример обучения модели приведен в листинге 91.

Листинг 91

```
model.fit(X_train, Y_train, batch_size=batch_size, nb_epoch=nb_epoch,  
verbose=1, validation_data=(X_test, Y_test))
```

Качество обученной модели проверяется кодом из листинга 92.

Листинг 92

```
score = model.evaluate(X_test, Y_test, verbose=0)
print('Test loss:', score[0])
print('Test accuracy:', score[1])
```

Напомним, что:

- Величина `loss` характеризует интегральную оценку всех средне-квадратических ошибок. Единицы этой величины совпадают с единицами Y . Чем меньше это значение, тем лучше. Однако для каждой задачи (каждого типа данных) конкретные величины будут разными, и оценивать эту величину нужно\можно, только зная характер данных и допустимые величины. Так, для оценки стоимости квартиры общая ошибка 300 000 на тренировочной выборке в несколько тысяч квартир может быть вполне приемлемой. А вот такая же величина `loss` для оценки уровня лейкоцитов в крови на тысячу пациентов просто неприемлема.

- Величина `accuracy` характеризует процент верных ответов из тестовой выборки. Чем ближе это значение к 100% (или 1.00), тем лучше.

Загрузка и сохранение сложных моделей

В одном из прошлых пунктов был рассмотрен вариант сохранения объектов посредством функций `pickle.dump(obj, output)` и `pickle.load(f)`. Однако многие модели Keras имеют такую сложную структуру, что стандартный сериализатор объектов вызывает исключение из-за глубины ссылок. Поэтому для сохранения и загрузки сложных моделей нейронных сетей необходимо использовать следующий код (листинг 93).

Обратите внимание!

- *Настройки модели и веса сохраняются отдельно.*

Листинг 93

```
def SaveToJson(model, fileName):
    model_json = model.to_json()
    with open(fileName + ".net", "w") as json_file:
        json_file.write(model_json)
    model.save_weights(fileName + ".wgt")
    print("model saved to disk")
def LoadFromJson(fileName):
    json_file = open(fileName + ".net", 'r')
    loaded_model_json = json_file.read()
    json_file.close()
    loaded_model = model_from_json(loaded_model_json)
    loaded_model.load_weights(fileName + ".wgt")
    print("Loaded model from disk")
    return loaded_model
```

Рекуррентные сети для прогнозирования временных рядов

Теперь рассмотрим пример работы с рекуррентными сетями средствами библиотеки Keras, а именно, их применение для решения задачи прогнозирования временных рядов. Пусть в файле «international-airline-passengers.csv» (скачать его можно отсюда: [41]) находятся данные о ежемесячном количестве пассажиров международного аэропорта в виде, представленном на рисунке 70. Применим рекуррентную сеть для прогнозирования количества пассажира в будущем. Глубину прогноза возьмем равной одному месяцу, как и количество точек, по которому будем строить прогноз. Качество работы модели оценим по величине среднеквадратичной ошибки. Данная задача и код ее решения, показанный в листинге 94, взяты из [62].

Листинг 94

```
# -*- coding: utf-8 -*-
#Необходимые импорты
import numpy
import matplotlib.pyplot as plt
```

Продолжение листинга 94

```
from pandas import read_csv
import math
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
# функция для создания выборки из исходного массива данных
def create_dataset(dataset, look_back=1):
    dataX, dataY = [], []
    for i in range(len(dataset)-look_back-1):
        a = dataset[(i+look_back), 0]
        dataX.append(a)
        dataY.append(dataset[i + look_back, 0])
    return numpy.array(dataX), numpy.array(dataY)
# фиксация состояния рандома для получения предсказуемого результата
numpy.random.seed(7)
# загрузка данных
dataframe = read_csv('international-airline-passengers.csv', usecols=[1], engine='python',
skipfooter=3)
dataset = dataframe.values
dataset = dataset.astype('float32')
# нормализация данных
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(dataset)
# ручное разделение данных на обучающую и тестовую выборку
train_size = int(len(dataset) * 0.67)
test_size = len(dataset) - train_size
train, test = dataset[0:train_size,:], dataset[train_size:len(dataset),:]
# Изменение исходного набора данных для того, чтобы в каждой строке X оказался
# временной ряд размером look_back, начиная от момента времени t0. А в Y
# временной ряд единичного размера от момента времени t0 + look_back.
# В данном примере, для look_back=4 X1 будет состоять из объектов: Data[t0],
# Data[t1], Data[t2], Data[t4], а Y1 из объектов: Data[t5]. X2 начнется с элемента
# Data[t1] и т.д.
trainX, trainY = create_dataset(train, look_back)
testX, testY = create_dataset(test, look_back)
# изменение формы входных данных и приведение их к виду
# [количество объектов, размер временного ряда из объектов, количество признаков
каждого объекта]
```

Окончание листинга 94

```
trainX = numpy.reshape(trainX, (trainX.shape[0], trainX.shape[1], 1))
testX = numpy.reshape(testX, (testX.shape[0], testX.shape[1], 1))
# Создание графа вычислений (некоторой абстрактной модели)
model = Sequential()
#4 – это количество блоков LSTM, input_shape - это форма вектора входа
model.add(LSTM(4, input_shape=(look_back, 1)))
model.add(Dense(1))
#Компиляция модели
model.compile(loss='mean_squared_error', optimizer='adam')
#Обучение модели
model.fit(trainX, trainY, epochs=100, batch_size=1, verbose=2)
# выполнение предсказания
trainPredict = model.predict(trainX)
testPredict = model.predict(testX)
# так как прогнозные значения нормализованы, то необходимо его восстановление
# под восстановлением понимается приведение к исходному масштабу данных
trainPredict = scaler.inverse_transform(trainPredict)
trainY = scaler.inverse_transform([trainY])
testPredict = scaler.inverse_transform(testPredict)
testY = scaler.inverse_transform([testY])
# вычисление ошибки прогноза
trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))
print('Train Score: %.2f RMSE' % (trainScore))
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))
print('Test Score: %.2f RMSE' % (testScore))
# подготовка данных для отрисовки
trainPredictPlot = numpy.empty_like(dataset)
trainPredictPlot[:, :] = numpy.nan
trainPredictPlot[look_back:len(trainPredict)+look_back, :] = trainPredict
testPredictPlot = numpy.empty_like(dataset)
testPredictPlot[:, :] = numpy.nan
testPredictPlot[len(trainPredict)+(look_back*2)+1:len(dataset)-1, :] = testPredict
# построение графика
plt.plot(scaler.inverse_transform(dataset))
plt.plot(trainPredictPlot)
plt.plot(testPredictPlot)
plt.show()
```

1	"Month", "International airline passengers: monthly totals in thousands. Jan 49 ? Dec 60"
2	"1949-01", 112
3	"1949-02", 118
4	"1949-03", 132
5	"1949-04", 129
6	"1949-05", 121
7	"1949-06", 135
8	"1949-07", 148
9	"1949-08", 148
10	"1949-09", 136
11	"1949-10", 119
12	"1949-11", 104
13	"1949-12", 118
14	"1950-01", 115
15	"1950-02", 126
16	"1950-03", 141
17	"1950-04", 135
18	"1950-05", 125
19	"1950-06", 149
20	"1950-07", 170

Рисунок 70. Данные о пассажирах

На этом мы заканчиваем рассмотрение инструментальных средств для разработки приложений сферы машинного обучения на языке Python. Как вы могли заметить, в пособии не были рассмотрены реализации генетических алгоритмов и методов нечеткой логики. Это связано с тем, что их реализация на языке Python не требует каких-либо специализированных библиотек, поэтому эту часть практического материала вам предлагается освоить самостоятельно в рамках выполнения задач по этим темам (см. раздел «Практические задания»). В случае затруднения вы можете обратиться к следующим ресурсам: реализация генетического алгоритма на Python – [63] и материалы учебного пособия [64], где рассматривается реализация генетических алгоритмов и некоторых операций нечеткой логики на языке Java.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ТЕСТОВЫЕ ЗАДАНИЯ

Тест «Общие сведения о машинном обучении»

1. Выберите верные утверждения
 - a) Одна из задач машинного обучения – научиться делать прогнозы для признаков
 - b) Объекты описываются с помощью признаков
 - c) Одна из задач машинного обучения – научиться делать прогнозы для объектов
 - d) Признаки описываются с помощью объектов
2. Какие из этих задач являются задачами классификации?
 - a) Прогноз температуры на следующий день
 - b) Разделение книг, хранящихся в электронной библиотеке, на научные и художественные
 - c) Поиск групп похожих пользователей интернет-магазина
 - d) Прогноз оценки студента по пятибалльной шкале на экзамене по машинному обучению в следующей сессии
3. Какие свойства данных препятствуют однозначному построению разделяющей поверхности?
 - a) Ортогональность
 - b) Мультиколлинеарность
 - c) Противоречивость
 - d) Категориальность
4. Какая способность людей и систем позволяет получать им новые знания по наблюдению отдельных прецедентов (примеров)?
 - a) Корректировать ошибку
 - b) Обобщать
 - c) Запоминать
 - d) Распознавать образы
5. Какая задача лучше всего подходит под следующее описание. Нахождение такой функции F , которая бы наилучшим образом отображала неизвестные ранее объекты X в конечное множество

целочисленных номеров (имен, меток), на основании обучающих пар (X, Y) ?

- a) Прогнозирование денежных затрат
 - b) Кластеризация клиентов
 - c) Классификация образов
 - d) Выявление особенностей в данных
6. Почему для обучения моделей используются такие методы, как Градиентный спуск?
- a) Потому что метод позволяет корректировать параметры модели постепенно
 - b) Потому что аналитические решения не всегда дают корректное решение
 - c) Потому что такой подход позволяет получать более точные решения (Глобальный экстремум в отличие от локального)
 - d) Потому что при большой размерности входных данных подобные методы работают быстрее
7. Выберите верные утверждения
- a) Метод Байеса – это во многом классический подход к классификации, основанный на оценке частоты встреч объектов со схожими признаками
 - b) Благодаря универсальности статистического подхода метод Байеса позволяет решать любые задачи без априорной информации
 - c) Данный метод позволяет очень хорошо обобщать высокоуровневые признаки
 - d) Закон, задающий распределение вероятностей, который используется в предсказательной модели, сильно влияет на способ обобщения

Проблема переобучения

1. Какие факторы влияют на переобучение?
2. Какие есть способы оценки переобучения?
3. Какие есть способы борьбы с переобучением?

Регрессия

1. Почему такая простая формула, как $y=kx+b$, позволяет делать прогнозы или классификацию?
2. В чем отличие линейной и логистической регрессий?
3. В чем отличие линейной от нелинейной регрессии?
4. В чем отличие линейной регрессии от полиномиальной?
5. Что позволяет делать LASSO?
6. В чем заключаются особенности Ridge регрессии?

Модели и методы нечеткой логики

1. Охарактеризуйте следующие понятия: нечеткие множества, операции нечеткой логики, нечеткие модели или нечеткие системы.
2. Сформулируйте понятие лингвистической неопределенности.
3. Напишите функцию лингвистической неопределенности некоторого объекта.
4. Дайте определение функции принадлежности.
5. К какому типу относятся нечеткие множества, если значения функции принадлежности нечеткого множества представлены точными числовыми значениями?
6. К какому типу относятся нечеткие множества, если значения функции принадлежности нечеткого множества моделируются другими нечеткими множествами?
7. Какие существуют способы задания функции принадлежности?
8. Приведите три примера функции принадлежности, задаваемых функциональным способом.
9. Напишите формулу компактной записи функции принадлежности.
10. Сформулируйте определение лингвистической переменной.
11. Опишите набор переменных, с помощью которого описывается лингвистическая переменная.
12. Дайте определения следующих понятий: X – универсальное множество объектов, \tilde{X} – базовое терм-множество, G – синтакси-

ческие правила вывода (порождения) новых термов, Р – семантические правила.

13. Сформулируйте определение нечеткой логики.
14. Какие необходимые характеристики нечеткой логики ввел Л. Заде?
15. Что лежит в основе операций нечеткой логики?
16. Какие операции используются для моделирования основных логических связок И, ИЛИ над нечеткими множествами в нечеткой логике?
17. Сформулируйте определения триангулярной нормы и триангулярной конормы (t-норма и s-конорма) и докажите их двойственность. Приведите примеры.
18. Запишите композиционное правило Л. Заде.
19. Сформулируйте содержательное определение операции импликации.
20. Сформулируйте определение системы нечеткого логического вывода.
21. Какие объекты входят в систему нечеткого логического вывода?
22. Какие условия должны соблюдаться при построении правил на основе системы нечеткого логического вывода?
23. Перечислите пять способов реализации нечеткого логического вывода.
24. Подробно опишите реализацию нечеткого логического вывода с помощью алгоритма Мамдани. Нарисуйте схему процесса нечеткого вывода этого алгоритма.

Нечеткие временные ряды

1. Сформулируйте определение уровня временного ряда.
2. Сформулируйте определение нечеткого временного ряда.
3. Сформулируйте определение носителя нечеткой метки \tilde{x}_i .
4. Приведите примеры прикладных задач обработки нечетких ВР.
5. Дайте определение нечеткому разбиению.

6. Опишите основные этапы алгоритма моделирования нечетких ВР в соответствии с нечеткой моделью Сонга.
7. В чем преимущество использования моделей нечетких ВР?
8. Что понимается под интеллектуальным анализом данных или Data Mining?
9. Какие задачи решаются на основе Data Mining?
10. Приведите виды темпорально-логических концептов ВР.
11. В чем заключается метод аппроксимации временного ряда на основе F-преобразования?

Нечеткая регрессия

1. Какие существуют подходы к построению моделей нечеткой линейной регрессии?
2. Какие существуют критерии для определения нечетких коэффициентов модели?
3. Какие вы знаете варианты методов на основе классификации «вход – выход»?

Генетические алгоритмы

1. Поясните происхождение термина «генетические алгоритмы».
2. Опишите сферу применения генетических алгоритмов.
3. Дайте определение гену в контексте генетических алгоритмов.
4. Дайте определение хромосоме в контексте генетических алгоритмов.
5. Дайте определение популяции в контексте генетических алгоритмов.
6. Дайте определение степени приспособленности в контексте генетических алгоритмов.
7. Дайте определение кроссовера в контексте генетических алгоритмов.
8. Дайте определение мутации в контексте генетических алгоритмов.
9. Перечислите методы отбора хромосом для кроссовера.
10. Расскажите о типовой схеме генетического алгоритма.

Нечеткая кластеризация

1. Дайте определение классификации.
2. Дайте определение кластеризации.
3. Поясните разницу между классификацией и кластеризацией.
4. Определите область применения кластеризации.
5. Определите цель алгоритма FCM-кластеризации.
6. Определите перечень входных данных для алгоритма FCM-кластеризации.
7. Перечислите шаги алгоритма FCM-кластеризации.
8. Почему кластеризация, проводимая с помощью алгоритма FCM, является нечеткой?
9. Поясните назначение параметра «степень нечеткости» в алгоритме FCM.
10. Что является выходными данными алгоритма FCM-кластеризации?

Искусственные нейронные сети и глубинное обучение

1. Какие достоинства и недостатки есть у ИНС по сравнению с Регрессией и Решающими Деревьями?
2. Сеть какого типа лучше использовать для прогнозирования?
3. Сеть какого типа можно использовать в условиях постоянного изменения данных, когда точной выборки еще не существует?

Тест «Искусственные нейронные сети»

1. Выберите верные утверждения:
 - а) ИНС проще подобрать под любую нелинейную задачу. Все, что нужно сделать, это увеличивать число слоев пропорционально числу признаков
 - б) ИНС позволяют обрабатывать более высокоуровневые признаки за счет нелинейной функции активации и последовательным слоям

- c) По сравнению с Регрессией ИНС практически не подвержены Переобучению при любом количестве нейронов
- d) С точки зрения математического аппарата ИНС – это комбинация полиномиальной регрессии высокого порядка и формулы Байеса
- e) ИНС может аппроксимировать любую нелинейную непрерывную функцию, но это еще не гарантирует 100% сходимости на произвольных данных
- f) ИНС в отличие от регрессии может хорошо обрабатывать высокую степень мультиколлинеарности и противоречивости в данных

2. Сеть какого типа лучше использовать для прогнозирования временных рядов?

- a) Сверточную
- b) ART MAP
- c) Импульсную
- d) MLP
- e) Рекуррентную
- f) Когнитрон

3. Сеть какого типа лучше использовать для обработки трехмерных сцен?

- a) MLP
- b) Рекуррентную
- c) ART MAP
- d) Сверточную
- e) Когнитрон
- f) Импульсную

4. Сеть какого типа лучше использовать для решения задачи классификации клиентов по одиночному вектору клиентских характеристик (с учетом того, что этот вектор содержит большое количество категориальных признаков)?

- a) Автокодировщик
- b) MLP

- c) Когнитрон
- d) ART MAP
- e) Сверточную
- f) Рекуррентную
- g) Импульсную

5. Сеть какого типа можно использовать в условиях постоянного изменения данных, когда точной выборки еще не существует и сеть приходится постоянно дообучивать на новых классах?

- a) MLP
- b) Сверточную
- c) Когнитрон
- d) Рекуррентную
- e) ART MAP
- f) Автокодировщик
- g) Импульсную

ПРАКТИЧЕСКИЕ ЗАДАНИЯ

Работа с файлом данных «Титаника»

Приведенные ниже задания основаны на данных 'titanic.csv', где содержатся сведения о пассажирах «Титаника».

Задачи:

1. Какое количество мужчин и женщин плыло на корабле?
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров. Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров? Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров. В качестве ответа приведите два числа через пробел.
5. Коррелируют ли число братьев/сестер/супругов с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch.
6. Какое самое популярное женское имя на корабле? Извлеките из полного имени пассажира (колонка Name) его личное имя (First Name). Это задание – типичный пример того, с чем сталкивается специалист по анализу данных. Данные очень разнородные и шумные, но из них требуется извлечь необходимую информацию. Попробуйте вручную разобрать несколько значений столбца Name и выработать правило для извлечения имен, а также разделения их на женские и мужские.

Работа по отбору признаков

Доработайте код отбора признаков, использовав все модели из списка:

1. Линейная регрессия (LinearRegression)
2. Гребневая регрессия (Ridge)
3. Лассо (Lasso)
4. Случайное Лассо (RandomizedLasso)
5. Рекурсивное сокращение признаков (Recursive Feature Elimination – RFE)
6. Сокращение признаков Случайными деревьями (Random Forest Regressor)
7. Линейная корреляция (f_regression)

Отобразите получившиеся значения\оценки каждого признака каждым методом\моделью и среднюю оценку. Проведите анализ получившихся результатов. Какие 4 признака оказались самыми важными по среднему значению? (Названия\индексы признаков и будут ответом на задание). Какие выводы можно сделать по результатам отдельных методов?

Многослойный персептрон

1. Доработайте код задачи классификации (листинг 58). Попробуйте настроить различные параметры сети, такие как количество нейронов в первом или втором слое, вид функции активации (activation), алгоритм градиентного спуска (solver), количество максимальных итераций при обучении (max_iter), порог точности при обучении (tol). Варьируя эти параметры, попробуйте найти наилучший вариант (возможные значения переменных можно найти в документации Scikit-learn).

2. Насколько лучше или хуже работает MLPClassifier по сравнению с персептроном? Без нормализации данных и с нормализацией данных? Предположите, почему именно так ведут себя модели.

3. Проведите серию экспериментов с различными значениями `random_state`. Как случайная инициализация весов влияет на Персептрон и многослойную сеть?

4. Дополнительно постройте график распределения исходных данных.

Реализация алгоритма обратного распространения ошибки

1. Модифицируйте сеть из листинга 59. Добавьте еще один слой и восстановите работоспособность алгоритма. Критерием его работоспособности может быть стремительно уменьшающаяся ошибка.

2. Дополнительно обучите старую и новую модели на следующих данных:

$$X = \begin{bmatrix} [0, 0, 1], [0.3, 1, 0], \\ [1, 0.3, 0], [0.6, 0.2, 1], \\ [0.6, 0.2, 1] \end{bmatrix}$$

Проанализируйте исходные данные. Какая зависимость скрыта в них? Объясните, что произошло с качеством классификации. Почему? С чем может быть связано подобное в реальных задачах и как на это можно повлиять (если это вообще возможно)?

Регуляризация и сеть прямого распространения

1. Вернитесь к задаче из пункта «Влияние регуляризации на многослойную сеть прямого распространения». Проанализируйте результат кода при количестве нейронов, равном 100, в скрытом слое MLP. Напомним, что регуляризация – это механизм снижения больших весов модели при повышении ее сложности. Этот механизм помогает избежать переобучения. Параметр регуляризации `alpha` регулирует размер штрафов, которые накладываются на веса. Какой коэффициент регуляризации оптимален для данной задачи?

2. Задача сравнения персептрона, многослойной сети и регрессий. В пункте «Влияние регуляризации на многослойную сеть прямого распространения» была описана общая методика по проведению

серии экспериментов и выводу результатов (понятно, что это можно сделать не только в виде графиков, но и выводом информации в консоль или записи более детальной информации в файл\таблицу для дальнейшей обработки). Соответственно каркас этой программы можно повторно использовать во многих других экспериментах не только по анализу влияния регуляризации.

По аналогии с заданием выше сгенерируйте 3 задачи классификации со следующими параметрами: `make_moons (noise=0.3, random_state=rs)`, `make_circles (noise=0.2, factor=0.5, random_state=rs)`, `X, y = make_classification (n_samples=500, n_features=2, n_redundant=0, n_informative=2, random_state=rs, n_clusters_per_class=1)` и сравните следующие модели:

- Линейную регрессию
- Полиномиальную регрессию (например, со степенью 3)
- Гребневую полиномиальную регрессию (например, со степенью 4, $\alpha = 1.0$)
- Перцептрон
- Многослойный перцептрон с десятью нейронами в скрытом слое ($\alpha = 0.01$)
- Многослойный перцептрон со ста нейронами в скрытом слое ($\alpha = 0.01$)

Создайте отдельный файл и замените только код моделей. Постройте графики и отобразите качество моделей. В результате у вас должно получиться полотно графиков, похожее на рисунок 71 (причем в данном случае представлена только первая задача классификации на Moondataset).

Проанализируйте полученные результаты, рассчитав их при параметре `random_state=25` для следующих функций\объектов:

- `Perceptron`
- `MLPClassifier`
- `Ridge`
- `make_classification`

- `rng = np.random.RandomState(25)`
- `make_moons`
- `make_circles`
- `train_test_split`

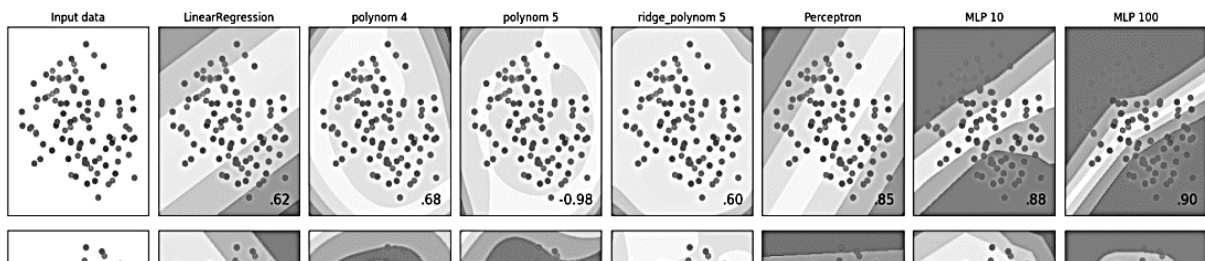


Рисунок 71. Пример работы программы по сравнению нескольких моделей на задачах moon, circle, linear

Сравнение эффективности моделей из библиотеки Keras

Вы научитесь:

- Проводить самостоятельный анализ результатов экспериментов;
- Проводить серию экспериментов;
- Интерпретировать результаты моделей;
- Подбирать соответствующие данным инструменты и параметры.

Необходимые `import` для выполнения работы представлены в листинге 95.

Листинг 95

```
import numpy as np
from keras.datasets import mnist
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation, Flatten
from keras.layers import Convolution2D, MaxPooling2D
from keras.utils import np_utils
import matplotlib.pyplot as plt
from keras.models import model_from_json
from sklearn.datasets import fetch_lfw_people
from sklearn.model_selection import train_test_split
```

Инструкции по выполнению представлены в листинге 96 (первые 3 шага), а также ниже. Прежде чем перейти к шагу 3, сравните качество полученных трех моделей на данных MNIST. Объясните полученные результаты. Есть ли превосходство сверточной сети перед обычной или перед полиномиальной регрессией?

Листинг 96

```
#Шаг 1. Загрузите данные MNIST
(X_train, y_train), (X_test, y_test) = mnist.load_data()
#Шаг 2. Создайте модель многослойного перцептрона (на основе Keras)и сверточной
#сети (на основе Keras) и полиномиальной регрессии из Scikit-Learn.
#Для задания обычного линейного слоя в качестве первого слоя сети потребуется
#указать форму входного вектора (размер). Это можно сделать следующим образом:
Dense(128, input_dim=input_size)
#Шаг 3. Далее попробуйте эти три типа модели уже на других данных.
#Для этого загрузите и подготовьте данные с фотографиями людей с помощью
#библиотеки scikit-learn (изучите код самостоятельно):
lfw_people = fetch_lfw_people(min_faces_per_person=70, resize=0.4)
# introspect the images arrays to find the shapes (for plotting)
n_samples, h, w = lfw_people.images.shape
X = lfw_people.images
# the label to predict is the id of the person
y = lfw_people.target
target_names = lfw_people.target_names
n_classes = target_names.shape[0]
print("n_classes: %d" % n_classes)
# split into a training and testing set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
X_train = X_train.reshape(X_train.shape[0], 1, img_rows, img_cols)
X_test = X_test.reshape(X_test.shape[0], 1, img_rows, img_cols)
X_train = X_train.astype('float32')
X_test = X_test.astype('float32')
X_train /= 255
X_test /= 255
```

Шаг 4. Отобразите данные, чтобы удостовериться в их корректной подготовке для дальнейшего обучения.

Шаг 5. Обучите модель многослойного персептрона, полиномиальной регрессии и сверточной сети. Попробуйте разное число нейронов и слоев в сетях (10, 20, 50, 100, 200, 500 нейронов) и (1-3 слоя). Больше число нейронов и слоев не нужно.

Замечание 1: Конечно же, не следует обучать модели сперва на MNIST, а потом эти же модели (конкретные экземпляры) на фотографиях. Это совсем разные задачи, и при малых количествах слоев и таких выборках вряд ли произойдет хорошее обобщение. Имеется в виду, что те же три типа моделей теперь надо сравнить на фотографиях.

Замечание 2: Если на CPU обучение идет слишком долго, то возьмите не всю обучающую выборку, а лишь половину или одну треть.

В заключение сравните качество обученных моделей на данных фотографий. Объясните полученные результаты и ответьте на вопросы:

- Почему получилось именно так?
- На что способны и не способны используемые модели?
- Чем отличаются данные из первой части задания от данных из второй части задания?

Работа с библиотекой OpenCV

Докажите гипотезу, сформулированную по завершению предыдущего задания. Что нужно сделать с данными MNIST, чтобы работающая ранее модель стала давать на них плохие результаты? Для этого вам может понадобиться библиотека OpenCV для Python. Которая достаточно легко устанавливается. Смотрите раздел «**Installing OpenCV from prebuilt binaries**» по следующей ссылке: [65].

Согласно Википедии [1]: «OpenCV (англ. Open Source Computer Vision Library, библиотека компьютерного зрения с открытым исходным кодом) – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом. Реализована на C/C++, также разрабатывается для Python, Java, Ruby, Matlab, Lua и других языков. Может свободно

использоваться в академических и коммерческих целях – распространяется в условиях лицензии BSD.

Проект OpenCV возник в 1999 году в рамках исследования Intel по высокопроизводительным приложениям по обработке 3D сцен. Это крайне мощная библиотека и стандарт де-факто для обработки изображений. Сферы применения OpenCV:

- Инструмент выделения 2D и 3D признаков;

- Трёхмерная навигация;

- Системы распознавания лиц;

- Системы распознавания жестов;

- Человеко-компьютерные интерфейсы;

- Мобильные роботы;

- Распознавание эмоций;

- Идентификация объектов;

- Сегментация и распознавание;

- Стереозрение;

- Восстановление поверхности;

- Отслеживание движения;

- Дополненная реальность.

Кроме того, для поддержки некоторых вышеуказанных областей, OpenCV включает в себя библиотеки статистической обработки данных и машинного обучения:

- Решающие деревья;

- Алгоритм максимального правдоподобия;

- Алгоритм k-ближайших соседей;

- Наивный Байесовский классификатор;

- Искусственные нейронные сети;

- Случайные леса;

- Метод опорных векторов.

В библиотеке OpenCV есть методы для многих видов деформаций изображений. Для базовой проверки способностей ИНС на двумерных изображениях вполне хватит следующих:

Повороты;
Линейные сдвиги;
Масштабирование».

Ниже (в листинге 97) приведен код, который позволяет выполнить преобразование исходного изображения `images[i]`. Каждый метод преобразования `cv2` возвращает новое изображение. Для поворотных и линейных сдвигов используются специальные матрицы деформации. Код для их инициализации также приведен ниже. Недостающие переменные – это коэффициенты и константы. Определите их самостоятельно.

Листинг 97

```
import cv2
# setrotation
rot_Matrix = cv2.getRotationMatrix2D((cols / 2, rows / 2), ang, 1)
cv2.warpAffine(images[i], rot_Matrix, (cols, rows))
# set scaling
cv2.resize(images[i], dsize=(rows, cols), fx=scale, fy=scale,
interpolation=cv2.INTER_CUBIC)
# set translation
trans_Matrix = np.float32([[1, 0, tx], [0, 1, ty]])
cv2.warpAffine(images[i], trans_Matrix, (cols, rows))
```

После преобразования части изображений проведите соответствующие эксперименты с разными моделями. Измерьте качество получившихся моделей. Вероятно, придется проделать много экспериментов с различными параметрами, чтобы добиться такой ситуации, когда результаты будут объяснять теорию.

Нечеткая логика

1. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с треугольной функцией принадлежности и отобразить его название, параметры, а также степень принадлежности вводимого пользователем объекта.

2. На языке Python разработайте скрипт, позволяющий задать нечеткое множество с трапециевидной функцией принадлежности и отобразить его параметры, а также степень принадлежности вводимого пользователем объекта.

3. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

4. На языке Python разработайте скрипт, позволяющий выполнить операцию пересечения заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – пересечение данных нечетких множеств.

5. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

6. На языке Python разработайте скрипт, позволяющий выполнить операцию объединения заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – объединение данных нечетких множеств.

7. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с треугольной функцией принадлежности. Входными данными будут параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

8. На языке Python разработайте скрипт, позволяющий выполнить операцию дополнения, заданного пользователем нечеткого множества с трапециевидной функцией принадлежности. Входными данными будут параметры функции принадлежности и четкие объекты множества. Выходными – дополнение нечеткого множества.

9. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с треугольными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Причем результат вывести через лингвистические переменные. Импликацию моделировать минимумом.

10. На языке Python разработайте скрипт, позволяющий выполнить операцию импликации заданных пользователем нечетких множеств с трапециевидными функциями принадлежности. Входными данными будут параметры функций принадлежности и четкие объекты для каждого из множеств. Выходными – результат импликации данных нечетких множеств. Импликацию моделировать минимумом.

Генетические алгоритмы

1. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано N наименований продуктов, для каждого из которых известно m характеристик. Необходимо получить самый дешевый рацион из k наименований, удовлетворяющий заданным медицинским нормам для каждой из m характеристик.

2. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано n пунктов производства продуктов и k городов, которые в них нуждаются. Каждый город может потребить x продуктов, а каждый пункт произвести y продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов без сильного его превышения, а транспортные расходы были минимальными.

3. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано N наименований продуктов, для каждого из которых известно m характеристик. Необходимо получить самый лучший по характеристикам рацион из k наименований, удовлетворяющий заданным ценовым рамкам. Лучшим считается рацион с минимальным отклонением от нормы.

4. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано n пунктов производства продуктов и k городов, которые в них нуждаются. Каждый город может потребить x продуктов, а каждый пункт произвести y продуктов. Необходимо получить оптимальный маршрут, так, чтобы все города получили нужный им объем продуктов с минимальным его превышением, а транспортные расходы укладывались в определенные рамки.

5. На языке Python разработайте скрипт, который с помощью генетического алгоритма решает следующую задачу. Дано N полей и k культур для посева. Для каждого поля известна характеристика урожайности каждой из k культур, а для каждой культуры – его закупочная стоимость. Необходимо получить самый лучший урожай за наименьшую стоимость.

Нечеткая кластеризация объектов

1. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере заработной платы n людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

2. На языке Python разработайте скрипт, кластеризующий загруженные данные о возрасте n людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

3. На языке Python разработайте скрипт, кластеризующий загруженные данные о стоимости n автомобилей на определенные им

кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

4. На языке Python разработайте скрипт, кластеризующий загруженные данные о росте n людей на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

5. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на производство n продуктов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

6. На языке Python разработайте скрипт, кластеризующий загруженные данные о длительности перелетов до n пунктов на определенные им кластеры, обозначенные заданными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

7. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере затрат на связь среди n отделов на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

8. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере площадей n квартир на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

9. На языке Python разработайте скрипт, кластеризующий загруженные данные о размере урожая n сельскохозяйственных культур на определенные им кластеры, обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

10. На языке Python разработайте скрипт, кластеризующий загруженные данные о весе n людей на определенные им кластеры,

обозначенные определенными в программе лингвистическими метками. Максимальное количество меток задать самостоятельно.

Анализ временных рядов

1. *Лингвистические шкалы.* Разработать скрипт на Python, позволяющий задавать с использованием функций принадлежности лингвистическую шкалу для решения задачи по варианту (таблица 14). Количество оценок в шкале и параметры функций принадлежности должны задаваться по умолчанию, если в заданном пользователем файле не указано иное. Осуществить графическое отображение функций принадлежности меток шкалы, выделив каждую отдельным цветом. Предусмотреть загрузку файла с данными и оценку их по построенной шкале.

Таблица 14. Варианты для задачи «Лингвистические шкалы»

Вариант	Функция принадлежности	Назначение шкалы
1.	Треугольная	Оценка эффективности продаж
2.	Трапецевидная	Оценка эффективности операций с валютой
3.	Треугольная	Оценка прибыли
4.	Треугольная	Оценка затрат на оплату труда
5.	Трапецевидная	Оценка рентабельности
6.	Треугольная	Оценка объемов продаж
7.	Трапецевидная	Оценка объемов закупок
8.	Треугольная	Оценка загруженности сервера и сети
9.	Треугольная	Оценка уровня зарплаты
10.	Трапецевидная	Оценка убытков
11.	Треугольная	Оценка эффективности продаж
12.	Трапецевидная	Оценка эффективности операций с валютой
13.	Треугольная	Оценка прибыли

Вариант	Функция принадлежности	Назначение шкалы
14.	Треугольная	Оценка затрат на оплату труда
15.	Трапециевидная	Оценка рентабельности
16.	Треугольная	Оценка объемов продаж
17.	Трапециевидная	Оценка объемов закупок
18.	Треугольная	Оценка загруженности сервера и сети
19.	Треугольная	Оценка уровня зарплаты
20.	Трапециевидная	Оценка убытков
21.	Треугольная	Оценка эффективности продаж
22.	Трапециевидная	Оценка эффективности операций с валютой
23.	Треугольная	Оценка прибыли
24.	Треугольная	Оценка затрат на оплату труда
25.	Трапециевидная	Оценка рентабельности
26.	Треугольная	Оценка объемов продаж
27.	Трапециевидная	Оценка объемов закупок
28.	Треугольная	Оценка загруженности сервера и сети
29.	Треугольная	Оценка уровня зарплаты
30.	Трапециевидная	Оценка убытков

2. *Нечеткие временные ряды и ряды нечетких тенденций.* Реализовать скрипт на языке Python, позволяющий загружать четкий временной ряд величины из задачи по лингвистическим шкалам (задача «Лингвистические шкалы»). С помощью построенной в предыдущей работе шкалы (задача «Лингвистические шкалы») построить из четкого временного ряда временной ряд нечетких значений и нечетких тенденций. Осуществить графическое отображение нечетких меток шкалы и нечетких тенденций, выделив каждую отдельным цветом.

3. *Кластеризация и лингвистические шкалы.* Язык реализации Python. Построить шкалу из задачи по лингвистическим шкалам (работа «Лингвистические шкалы») с помощью алгоритма fcm-клас-

теризации. Количество кластеров задается по умолчанию, но может редактироваться пользователем посредством загрузки из файла. Каждый кластер – метка на шкале с присвоенным ему лингвистическим значением либо по умолчанию, либо пользователем. Осуществить графическое отображение меток шкалы, выделив каждую отдельным цветом.

3. *Прогнозирование значения временного ряда с помощью нейронной сети.* Реализовать прогноз следующего значения четкого временного ряда из работы «Нечеткие временные ряды и ряды нечетких тенденций» с помощью нейронной сети, но НЕ рекуррентного типа. Графически отобразить реальный ряд и прогнозное значение, а также проиллюстрировать результат тестового прогноза.

4. *Прогнозирование значения временного ряда с помощью модифицированного метода Сонга.* Реализовать прогноз следующего значения временного ряда из работы «Нечеткие временные ряды и ряды нечетких тенденций» с помощью метода, описанного в пункте «Пример моделирования временного ряда в нечетком подходе». Графически отобразить реальный ряд и прогнозное значение, а также проиллюстрировать результат тестового прогноза.

5. *F-преобразование.* Для временного ряда из работы «Нечеткие временные ряды и ряды нечетких тенденций» выделить тренд методом F-преобразования.

Работа с рекуррентными сетями

1. Модифицируйте код из листинга 91: поэкспериментируйте с разной структурой сети, разным количеством точек, по которым строится прогноз, и с разным количеством LSTM-блоков. Оцените качество работы сети в разных экспериментах. Интерпретируйте результаты.

2. Решите задачу «Прогнозирование значения временного ряда с помощью нейронной сети» из блока заданий «Анализ временных рядов» с помощью рекуррентной сети. Сравните и интерпретируйте результаты.

3. Объясните работу кода из листинга 98.

Листинг 98

```
import numpy
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.utils import np_utils

numpy.random.seed(7)
alphabet = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
char_to_int = dict((c, i) for i, c in enumerate(alphabet))
int_to_char = dict((i, c) for i, c in enumerate(alphabet))
seq_length = 1
dataX = []
dataY = []
for i in range(0, len(alphabet) - seq_length, 1):
    seq_in = alphabet[i:i + seq_length]
    seq_out = alphabet[i + seq_length]
    dataX.append([char_to_int[char] for char in seq_in])
    dataY.append(char_to_int[seq_out])
    print seq_in, '->', seq_out
X = numpy.reshape(dataX, (len(dataX), seq_length, 1))
X = X / float(len(alphabet))
y = np_utils.to_categorical(dataY)

model = Sequential()
model.add(LSTM(32, input_shape=(X.shape[1], X.shape[2])))
model.add(Dense(y.shape[1], activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
model.fit(X, y, nb_epoch=500, batch_size=1, verbose=2)
scores = model.evaluate(X, y, verbose=0)
print("Model Accuracy: %.2f%%" % (scores[1]*100))
for pattern in dataX:
    x = numpy.reshape(pattern, (1, len(pattern), 1))
    x = x / float(len(alphabet))
    prediction = model.predict(x, verbose=0)
    index = numpy.argmax(prediction)
    result = int_to_char[index]
    seq_in = [int_to_char[value] for value in pattern]
    print seq_in, "->", result
```

4. Проведите два эксперимента, модифицируя код из листинга 92 следующим образом:

1. Изменив значение `seq_length`, сделав его равным 3 и изменив форму входных данных следующим образом:

```
X = numpy.reshape(dataX, (len(dataX), 1, seq_length))
```

2. Изменив значение `seq_length`, сделав его равным 3 и изменив форму входных данных следующим образом:

```
X = numpy.reshape(dataX, (len(dataX), 1, seq_length))
```

Интерпретируйте полученные результаты.

ЗАКЛЮЧЕНИЕ

Мы рассмотрели модели и алгоритмы, используемые в сфере машинного обучения для анализа данных, а также способы их реализации на языке Python с использованием библиотек Keras, Theano и других.

В настоящее время машинное обучение и область анализа данных являются весьма перспективными направлениями в ИТ. Задачей данной книги было познакомить вас с азами этого направления, рассмотрев наиболее важные модели, алгоритмы и их реализацию. Однако, чтобы стать специалистом в сфере анализа данных, недостаточно просто знать модели и инструменты. Необходимо понимать границы их применений и преимущества одного перед другим.

Данное учебное пособие лишь приоткрывает дверцу в мир машинного обучения, освещая небольшую его часть. Если же после работы с книгой вы поняли, что анализ данных – крайне интересная вещь, то вам непременно нужно двигаться дальше. Причем не только в изучении моделей, методов, алгоритмов и инструментов их реализации (что вы можете сделать, изучив приведенные в библиографическом списке источники), но и в совершенствовании своего мышления. Ведь пока искусственный интеллект еще не изобретен, всю семантическую нагрузку задач машинного обучения берет на себя разработчик-человек. Поэтому он непременно должен обладать острым аналитическим умом.

Перефразируя Шерлока Холмса, напомним, что «развитый мозг – это то, что всегда будет в цене». В связи с этим, закончим пособие пожеланием: оттачивайте свой разум, покоряйте новые вершины, не останавливайтесь на достигнутом!

ГЛОССАРИЙ

DataMining – интеллектуальный анализ данных.

TimeSeries DataMining – интеллектуальный анализ временных рядов.

Аномалия временного ряда – новый, не типичный паттерн временного ряда.

Апостериорная вероятность – назначенная событию вероятность при условии наличия знаний, поддерживающих его наступление и полученных опытным путем.

Априорная вероятность – назначенная событию вероятность, при условии отсутствия знаний, поддерживающих его наступление.

Временной ряд (ВР) – последовательность упорядоченных в равноотстоящие моменты времени пар (момент_времени, значение_характеристики).

Генетические алгоритмы – адаптивные методы поиска, использующиеся для решения задач функциональной оптимизации.

Градиент – вектор, указывающий направление наибольшего возрастания некоторой величины, значение которой меняется в скалярном поле.

Дефаззификация – получение оптимального четкого значения по агрегированному нечеткому понятию.

Дисперсия – мера разброса элементов выборки относительно ее математического ожидания.

Задача классификации – распределение некоторого множества объектов по заданному множеству групп (классов).

Задача кластеризации – разделение некоторого множества объектов на непересекающиеся группы (кластеры) таким образом, чтобы каждая группа состояла из схожих объектов, а объекты разных кластеров существенно отличались.

Задача регрессии – приближение неизвестной целевой зависимости на некотором множестве данных.

Задача таксономии – задача построения древообразной иерархической структуры, упорядочивающей исходные данные.

Закон распределения – функция, определяющая для выборки вероятность попадания в некоторый интервал или вероятность получения определенного значения.

Знание – совокупность утверждений о закономерностях и свойствах процессов и явлений, а также связывающих их правил логического вывода и правил использования их при принятии решений.

Индексирование объектов временного ряда – построение индексов для эффективного выполнения запросов к базам данных ВР.

Классификация объектов временного ряда – назначение ВР или их паттернам одного из заранее определенных классов.

Кластеризация объектов временного ряда – поиск группировок ВР или их паттернов.

Комбинированная модель прогнозирования – модель прогнозирования, состоящая из нескольких индивидуальных (частных) моделей, называемых базовым набором моделей.

Концепт временной продолжительности – присутствие определенного паттерна или признака ВР на определенном интервале времени.

Концепт нечеткости ВР – нечеткость выраженности темпоральных событий и отношений.

Концепт одновременности ВР – совпадение во времени темпоральных событий (паттернов различных ВР).

Концепт очередности ВР – порядок следования паттернов ВР во времени.

Кроссовер – операция в генетическом алгоритме, позволяющая хромосомам обмениваться между собой своими частями.

Лингвистическая переменная – это переменная, значениями которой являются слова или высказывания естественного или искусственного языка.

Математическое ожидание – среднее значение вероятностных элементов выборки.

Медиана – число, характеризующее выборку по среднему из ее значений.

Метод градиентного спуска – нахождение локального экстремума (минимума или максимума) функции путем движения вдоль градиента в направлении наискорейшего спуска, задаваемого антиградиентом.

Мутация – операция в генетическом алгоритме, произвольным образом изменяющая хромосому.

Недообучение – ситуация, когда алгоритм при обучении с учителем не дает удовлетворительно малой средней ошибки на обучающем множестве.

Нечеткая логика – логика, оперирующая нечеткими высказываниями и рассуждениями на базе частичной истинности.

Нечеткий временной ряд (НВР) – упорядоченная в равноотстоящие моменты времени последовательность наблюдений над некоторым процессом, состояния которого изменяются во времени, если значение состояния процесса в каждый момент времени может быть выражено с помощью нечеткой метки .

Нечеткое множество – совокупность объектов, принадлежащих ему с определенной степенью.

Нормализация данных – процесс масштабирования вектора каждого признака к такому виду, что вектор будет иметь единичную норму (при этом есть разные способы оценки\подсчета нормы).

Обобщающая способность – свойство модели отражать исходные данные в требуемые результаты ($X \rightarrow Y$) на всем множестве исходных данных (во всех сценариях, а не только на тренировочных примерах).

Обучение – способность алгоритма при решении задач некоторого класса выдавать на некотором опыте лучшие результаты в смысле заданной меры качества при предъявлении нового опыта.

Ошибка – численно выраженная разница между ответом модели и требуемым (реальным) значением.

Переобучение – свойство натренированного алгоритма на объектах тренировочной выборки давать существенно меньшую вероятность ошибки, чем на объектах тестовой.

Пространство признаков – это N-мерное пространство, где N – число измеряемых характеристик объектов, выделенное для конкретной задачи.

Регуляризация – добавление некоторой дополнительной информации к условию минимизации ошибки.

Резюмирование временного ряда – формирование краткого описания ВР, содержащего существенные черты с точки зрения решаемой задачи.

Сегментация временного ряда – разбиение временного ряда на значимые сегменты.

Система нечеткого логического вывода – модель, описывающая поведение систем на естественном (или близком к естественному) языке в виде приближенных рассуждений на основе композиционного правила вывода.

Среднеквадратическое отклонение – величина, характеризующая рассеивание значений выборки относительно ее математического ожидания.

Стандартизация данных – процесс приведения вектора каждого признака к такому виду, что его математическое ожидание станет нулевым, а дисперсия – единичной.

Степень принадлежности – значение, задаваемое функцией принадлежности и находящееся в интервале от 0 до 1.

Фаззификация – процесс установки соответствия между четким значением x и нечетким f через функцию принадлежности.

Функция принадлежности – функция, отображающая базовое значение x и нечеткое f в интервал от 0 до 1.

Хромосома (в генетическом алгоритме) – битовая строка, описывающая решение.

Частотный анализ временного ряда – поиск часто проявляющихся паттернов ВР.

Энтропия – мера неупорядоченности системы.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- Градиент, 46
Дерево решений, 52
Диаграмма рассеяния, 58
Дисперсия, 48
Задача визуализации данных, 24
Задача классификации, 18
Задача кластеризации, 19
Задача регрессии, 18
Задача сокращения размерности, 24
Задача таксономии, 21
Закон распределения, 49
Кросс-валидация, 40
Математическая модель, 44
Математическое ожидание, 48
Машинное обучение, 7
Медиана, 47
Метод минимизации эмпирического риска, 37
Метрики качества кластеризации, 25
Модель МакКаллока-Питтса, 115
Недообучение, 36
Нечеткая импликация, 78
Нормализация, 33
Обобщающая способность, 16
Обучение без учителя, 23
Обучение с учителем, 21
Персептрон, 118
Пространство признаков, 13
Регрессионный анализ, 57
Регуляризация, 40
Система нечеткого логического вывода, 79, 281
Среднеквадратическое отклонение, 48
Стандартизация, 33
Теорема Байеса, 50
Функционалы качества, 22
Функция Хевисайда, 117
Энтропия, 52
accuracy_score, 205
ACL-шкала, 106
ConvolutionND, 242
DataFrame, 176
DataMining, 7
DataScience, 9
Dense, 242
Dropout, 243
FCM-кластеризация, 108
F-преобразование, 99
Keras, 240
LinearRegression, 201
LogisticRegression, 201
Matplotlib, 202
MLPClassifier, 218
Numpy, 172
Perceptron, 218
pickle, 198
PolynomialFeatures, 214
Python, 167
Samplesgenerator, 229
sklearn.linear_model.Ridge, 192
sklearn.tree.DecisionTreeClassifier, 188
s-конорма, 74
TfidfVectorizer, 192
TimeSeriesDataMining, 94
t-норма, 74

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. [Электронный ресурс]: Материалы свободной энциклопедии «Википедиа». URL: <http://ru.wikipedia.org/wiki/>, (режим доступа – свободный), (дата обращения: 28.03.2017).
2. [Электронный ресурс]: Материалы открытого курса по машинному обучению от компании ODS. URL: <https://habrahabr.ru/company/ods/blog/325654/>, (режим доступа – свободный), (дата обращения: 28.03.2017).
3. Воронина, В. В. Разработка приложений для анализа слабо-структурированных информационных ресурсов : учебное пособие / В. В. Воронина, В. С. Мошкин. – Ульяновск : УлГТУ, 2015. – 162 с.
4. [Электронный ресурс]: Материалы сайта machinelearning. URL: <http://www.machinelearning.ru/wiki/index.php?title=Переобучение>, (режим доступа – свободный), (дата обращения: 28.03.2017).
5. [Электронный ресурс]: Статья по регуляризации. URL: [https://ru.wikipedia.org/wiki/Регуляризация_\(математика\)](https://ru.wikipedia.org/wiki/Регуляризация_(математика)), (режим доступа – свободный), (дата обращения: 28.03.2017).
6. [Электронный ресурс]: Материалы сайта MSDN. URL: <https://msdn.microsoft.com/ru-ru/magazine/dn904675.aspx>, (режим доступа – свободный), (дата обращения: 28.03.2017).
7. [Электронный ресурс]: Материалы сайта machinelearning. URL: http://www.machinelearning.ru/wiki/index.php?title=Нормальное_распределение, (режим доступа – свободный), (дата обращения: 28.03.2017).
8. Паклин, Н. Б. Глава 9, // Бизнес-аналитика: от данных к знаниям : учебное пособие / Н. Б. Паклин, В. И. Орешков. – 2-е изд.. – СПб. : Питер, 2013. – С. 444-459.

9. [Электронный ресурс]: Ю. Лаходюк, Энтропия и деревья принятия решений. URL: <https://habrahabr.ru/post/171759/>, (режим доступа – свободный), (дата обращения: 28.03.2017).
10. [Электронный ресурс]: Статья по деревьям решений. URL: http://ru.wikipedia.org/wiki/Дерево_принятия_решений, (режим доступа – свободный), (дата обращения: 28.03.2017).
11. Клячкин, В. Н. Статистические методы анализа данных / В. Н. Клячкин, Ю. Е. Кувайскова, В. А. Алексеева. – М. : Финансы и статистика, 2016. – 240 с.
12. Клячкин, В. Н. Статистические методы в управлении качеством: компьютерные технологии / В. Н. Клячкин. – М. : Финансы и статистика, ИНФРА-М, 2009. – 304 с.
13. Валеев, С.Г. Регрессионное моделирование при обработке наблюдений / С. Г. Валеев. – М. : Наука, 1991. – 272 с.
14. [Электронный ресурс]: Материалы свободной энциклопедии «Википедия»: Статья под названием «Робастные методы». URL: <http://ru.wikipedia.org/wiki/Робастность>, (режим доступа – свободный), (дата обращения: 08.07.2017).
15. Zadeh, A. Lotfi. Fuzzy Sets / Lotfi A. Zadeh // Information and Control. – 1965.
16. Заде, Л. А. Основы нового подхода к анализу сложных систем и процессов принятия решений / Л. А. Заде // Математика сегодня. – М. : Знание, 1974. – С. 5-49.
17. Штовба, С. Д. Проектирование нечетких систем средствами MATLAB / С. Д. Штовба. – М. : Горячая линия – Телеком, 2007. – 288 с.
18. Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учебное пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.

19. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов : учебное пособие / Н. Г. Ярушкина, Т. В. Афанасьева., И. Г. Перфильева. – М. : ИД «ФОРУМ» : ИНФРА-М, 2012. – 160 с. – (Высшее образование).

20. Song, Q. Fuzzy time series and its models / Q. Song, B. Chissom // Fuzzy Sets and Systems. – №54 (1993) – P. 269-277.

21. [Электронный ресурс] Дегтярев, К. Ю. Применение специализированных компьютерных программ и методов, основанных на нечетких временных рядах для краткосрочного прогнозирования USD/RUB котировок / К. Ю. Дегтярев. URL: <http://www.exponenta.ru/educat/news/degtyarev/paper.pdf>, (режим доступа – свободный), (дата обращения: 08.07.2017).

22. Batyrshin, I. Perception Based Time Series Data Mining for Decision Making / I. Batyrshin // IFSA'07 Fuzzy Logic, Soft Computing and Computational Intelligence.

23. [Электронный ресурс]: Материалы IRAFM-2015. URL: <http://irafm.osu.cz/cif2015/main.php?c=Static&page=results>, (режим доступа – свободный), (дата обращения: 08.07.2017).

24. Новак, В. Математические принципы нечеткой логики / В. Новак, И. Перфильева, И. Мочкорж; пер. с англ.; под ред. А. Н. Аверкина. – М. : ФИЗМАТЛИТ, 2006.

25. Афанасьева Т. В. Модель ACL-шкалы для генерации лингвистических оценок в принятии решений / Т. В. Афанасьева // Вопросы современной науки и практики. Университет им. В. И. Вернадского. Т. 2. Серия «Технические науки». – 2008. – №4 (14). – С. 91-97.

26. [Электронный ресурс]: Саймон Хайкин - Нейронные сети. Полный курс. URL: <http://neuralnetworksanddeeplearning.com>, (режим доступа – свободный), (дата обращения: 08.07.2017).

27. [Электронный ресурс]: Материалы сайта machinelearning. URL: http://www.machinelearning.ru/wiki/index.php?title=Модель_Мак

Каллока_Питтса, (режим доступа – свободный), (дата обращения: 08.07.2017).

28. [Электронный ресурс]: Статья Логика мышления. Часть 3. Персептрон, сверточные сети. URL: <https://geektimes.ru/post/214317/>, (режим доступа – свободный), (дата обращения: 08.07.2017).

29. [Электронный ресурс]: Материалы сайта machinelearning. URL: http://www.machinelearning.ru/wiki/index.php?title=Самоорганизующаяся_карта_Кохонена, (режим доступа – свободный), (дата обращения: 08.07.2017).

30. [Электронный ресурс]: Материалы свободной энциклопедии «Википедия»: Python. URL: <https://ru.wikipedia.org/wiki/Python>, (режим доступа – свободный), (дата обращения: 08.07.2017).

31. [Электронный ресурс]: Материалы по синтаксису Python. URL: <https://habrahabr.ru/post/31180/>, (режим доступа – свободный), (дата обращения: 08.07.2017).

32. [Электронный ресурс]: Официальный сайт разработчиков Anaconda. URL: <https://www.continuum.io/Downloads>, (режим доступа – свободный), (дата обращения: 08.07.2017).

33. [Электронный ресурс]: Официальный сайт PyCharm. URL: <https://www.jetbrains.com/pycharm/download/#section=windows>, (режим доступа – свободный), (дата обращения: 08.07.2017).

34. [Электронный ресурс]: Официальный сайт SciPy. URL: <https://scipy.org/> (режим доступа – свободный), (дата обращения: 08.07.2017).

35. [Электронный ресурс]: Покоряем Python – уроки для начинающих. Функции lambda. URL: <https://pythlife.blogspot.ru/2012/11/lambda.html> (режим доступа – свободный), (дата обращения: 08.07.2017).

36. [Электронный ресурс]: Документация по DataFrame. URL: <http://pandas.pydata.org/pandas-docs/stable/generated/pandas.Data>

Frame.html (режим доступа – свободный), (дата обращения: 08.07.2017).

37. [Электронный ресурс]: Уроки по Pandas. URL: <https://bitbucket.org/hrojas/learn-pandas> (режим доступа – свободный), (дата обращения: 08.07.2017).

38. [Электронный ресурс]: Документация по Scikit-learn. URL: <http://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-normalization> (режим доступа – свободный), (дата обращения: 08.07.2017).

39. [Электронный ресурс]: Сервис Kaggle. Массив данных о пассажирах Титаника URL: <https://www.kaggle.com/prkukunoor/TitanicDataset> (режим доступа – свободный), (дата обращения: 08.07.2017).

40. [Электронный ресурс]: Документация по Scikit-learn:Ridge. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html#sklearn.linear_model.Ridge (режим доступа – свободный), (дата обращения: 08.07.2017).

41. [Электронный ресурс]: Ссылка для скачивания упомянутых в книге файлов с данными. URL: <https://drive.google.com/drive/folders/0B5yyS8oSQ0FDelpKRXg3c3lKVlU> (режим доступа – свободный), (дата обращения: 20.07.2017).

42. [Электронный ресурс]: Документация по Scikit-learn:TFIDFVectorizer. URL: http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html#sklearn.feature_extraction.text.TfidfVectorizer (режим доступа – свободный), (дата обращения: 08.07.2017).

43. [Электронный ресурс]: Документация по Scikit-learn:DictVectorizer. URL: http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.DictVectorizer.html (режим доступа – свободный), (дата обращения: 08.07.2017).

44. [Электронный ресурс]: Документация по Scikit-learn: LogisticRegression. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (режим доступа – свободный), (дата обращения: 08.07.2017).

45. [Электронный ресурс]: Документация по Scikit-learn: LinearRegression. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html (режим доступа – свободный), (дата обращения: 08.07.2017).

46. [Электронный ресурс]: Документация по Scikit-learn:Lasso. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html (режим доступа – свободный), (дата обращения: 08.07.2017).

47. [Электронный ресурс]: Документация по Scikit-learn: MinMaxScaler. URL: <http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>(режим доступа – свободный), (дата обращения: 08.07.2017).

48. [Электронный ресурс]: Функции map и zip и lambda. Python. URL: <http://ninjaside.info/blog/ru/funkcii-map-i-zip-i-lambda-python/> (режим доступа – свободный), (дата обращения: 08.07.2017).

49. [Электронный ресурс]: Документация по Scikit-learn: RandomizedLasso. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RandomizedLasso.html (режим доступа – свободный), (дата обращения: 08.07.2017).

50. [Электронный ресурс]: Документация по Scikit-learn:RFE. URL: http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html(режим доступа – свободный), (дата обращения: 08.07.2017).

51. [Электронный ресурс]: Документация по Scikit-learn: RandomForestRegressor. URL: <http://scikit-learn.org/stable/modules/>

generated/sklearn.ensemble.RandomForestRegressor.html (режим доступа – свободный), (дата обращения: 08.07.2017).

52. [Электронный ресурс]: Документация по Scikit-learn: f_regression. URL: http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_regression.html(режим доступа – свободный), (дата обращения: 08.07.2017).

53. [Электронный ресурс]: Документация по Scikit-learn: Pipeline. URL: <http://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html>(режим доступа – свободный), (дата обращения: 08.07.2017).

54. [Электронный ресурс]: Документация по Scikit-learn: Cross-validation. URL: http://scikit-learn.org/stable/modules/cross_validation.html#cross-validation(режим доступа – свободный), (дата обращения: 08.07.2017).

55. [Электронный ресурс]: Документация по Scikit-learn: Cross_val_score. URL: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html, (режим доступа – свободный), (дата обращения: 08.07.2017).

56. [Электронный ресурс]: Документация по Scikit-learn: Perceptron. URL: http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Perceptron.html, (режим доступа – свободный), (дата обращения: 08.07.2017).

57. [Электронный ресурс]: Документация по Scikit-learn: MLPClassifier. URL: http://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html, (режим доступа – свободный), (дата обращения: 08.07.2017).

58. [Электронный ресурс]: Официальный сайт NVIDIA. Проверка видеокарты. URL: <https://developer.nvidia.com/cuda-gpus> (режим доступа – свободный), (дата обращения: 08.07.2017).

59. [Электронный ресурс]: Официальный сайт NVIDIA. Скачивание CUDA. URL: <https://developer.nvidia.com/cuda-downloads> (режим доступа – свободный), (дата обращения: 08.07.2017).

60. [Электронный ресурс]: Официальный сайт NVIDIA. Скачивание CUDANN. URL: <https://developer.nvidia.com/rdp/cudnn-download> (режим доступа – свободный), (дата обращения: 08.07.2017).

61. [Электронный ресурс]: Документация по библиотеке Keras. URL: <https://keras.io/layers/core/>, (режим доступа – свободный), (дата обращения: 08.07.2017).

62. [Электронный ресурс]: Материалы сайта machinelearningmastery. URL: <https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/>, (режим доступа – свободный), (дата обращения: 08.07.2017).

63. [Электронный ресурс]: Статья по реализации генетического алгоритма на Python. URL: <http://easydan.com/arts/2016/genetic-optimization/>, (режим доступа – свободный), (дата обращения: 08.07.2017).

64. Воронина, В. В. Разработка веб-сервисов для анализа слабоструктурированных информационных ресурсов : учебное пособие / В. В. Воронина. – Ульяновск : УЛГТУ, 2016. – 166 с.

65. [Электронный ресурс]: Документация «Installing OpenCV from prebuilt binaries». URL: http://docs.opencv.org/3.2.0/d5/de5/tutorial_py_setup_in_windows.html, (режим доступа – свободный), (дата обращения: 08.07.2017).

Учебное электронное издание

ВОРОНИНА Валерия Вадимовна
МИХЕЕВ Александр Вячеславович
ЯРУШКИНА Надежда Глебовна
СВЯТОВ Кирилл Валерьевич

ТЕОРИЯ И ПРАКТИКА МАШИННОГО ОБУЧЕНИЯ

Учебное пособие

Редактор Н. А. Евдокимова

ЛР № 020640 от 22.10.97.

ЭИ № 1004. Объем данных 3,6 Мб.

Печатное издание

Подписано в печать 08.11.2017.

Усл. печ. л. 16.97. Тираж 100 экз. Заказ 932.

Ульяновский государственный технический университет,
432027, г. Ульяновск, ул. Сев. Венец, д. 32.

ИПК «Венец», УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

Тел.: (8422) 778-113

E-mail: venec@ulstu.ru

venec.ulstu.ru

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Планирование вычислительных экспериментов

Профиль подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

ЛАБОРАТОРНАЯ РАБОТА №1

«ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ»

1. Цель работы

Исследовать радиоэлектронное средство (РЭС) с применением полного факторного эксперимента и получить математическую модель данного РЭС.

2. Полный факторный эксперимент

В полном факторном эксперименте (ПФЭ) исследуется один параметр и реализуются все возможные сочетания уровней факторов.

Для каждого фактора выбираются два уровня – верхний и нижний, на которых фактор варьируется. Половина разности между верхним и нижним уровнями называется *интервалом варьирования*. Интервал варьирования должен быть больше погрешности измерения уровня фактора (ограничение снизу), а верхний и нижний уровни фактора не должны выходить за область его определения (ограничение сверху). На практике интервал варьирования составляет обычно 3–10% от области определения.

При двух уровнях для каждого из n факторов общее число опытов составляет 2^n . ПФЭ – это эксперимент типа 2^n .

ПФЭ позволяет получить математическую модель исследуемого объекта в виде уравнения множественной регрессии или по линиям

$$y = b_0 + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n \sum_{k=i+1}^n b_{ik} x_i x_k + \sum_{i=1}^n \sum_{k=i+1}^n \sum_{l=k+1}^n b_{ikl} x_i x_k x_l + \dots, \quad (2.1)$$

где b_0 – свободный член; b_i , b_{ik} , b_{ikl} – коэффициенты уравнения множественной регрессии.

Так, например, при $n = 2$

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2,$$

при $n = 3$

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + b_{123} x_1 x_2 x_3.$$

Модели (2.1) обычно называют регрессионными, а коэффициенты b_0 , b_i , b_{ik} , b_{ikl} , ... – коэффициентами уравнения регрессии.

В зависимости от объема априорной информации в ММ включают не все, а лишь некоторые взаимодействия первого порядка, иногда – взаимодействия второго порядка и очень редко – взаимодействия выше третьего порядка. Связано это с тем, что учет всех взаимодействий приводит к громоздким расчетам. Зависимость количества взаимодействий различного порядка от числа факторов приведена в табл. 2.1.

Таблица 2.1

Количество взаимодействий для числа факторов от 2 до 7

n	$N=2^n$	Число линейных эффектов	Порядок взаимодействия					
			1	2	3	4	5	6
2	4	2	1	–	–	–	–	–
3	8	3	3	1	–	–	–	–
4	16	4	6	4	1	–	–	–
5	32	5	10	10	5	1	–	–
6	64	6	15	20	15	6	1	–
7	128	7	21	35	35	21	7	1

Полное число всех возможных эффектов (включая b_0) равно числу опытов ПФЭ.

2.1. Стандартизация масштаба факторов

Для удобства расчетов масштаб факторов выбирают так, чтобы значение верхнего уровня было равно +1, а нижнего –1. С этой целью делают преобразование начала координат факторов и переходят к нормированному (стандартному) масштабу

$$x_i = \frac{(\tilde{x}_i - \tilde{x}_{i0})}{I}, \quad (2.2)$$

где x_i – нормированное значение; \tilde{x}_i – натуральное значение; \tilde{x}_{i0} – основной уровень; I – интервал варьирования.

Интеграл варьирования I равен

$$I = |\tilde{x}_i - \tilde{x}_{i0}|$$

2.2. Составление матрицы планирования ПФЭ

План ПФЭ изображают в виде таблицы, столбцы которой отражают уровни факторов, а строки – номера опытов. Эти таблицы называют *матрицами планирования* (МП) *эксперимента*. Поскольку значения уровней факторов по модулю всегда равны единице, то обычно в МП записывают только знак уровня (т. е. «+» вместо «1» и «-» вместо «-1»). В табл. 2.2 для примера приведена МП для ПФЭ типа 2^2 , которую называют *базовой*, так как с ее помощью легко построить матрицы любого порядка.

Так, для построения матрицы 2^3 сочетаем базовую матрицу с нижним и верхним уровнями x_3 (табл. 2.3). Легко заметить, что в первом столбце знаки меняются поочередно, во втором через 2, в третьем через 4 и так далее. То есть $2^0, 2^1, 2^2, 2^3, \dots$.

Таблица 2.2
МП ПФЭ типа 2^2

N	x_1	x_2	y
1	-	-	y_1
2	+	-	y_2
3	-	+	y_3
4	+	+	y_4

Таблица 2.3
МП ПФЭ типа 2^3

N	x_1	x_2	x_3	y
1	-	-	-	y_1
2	+	-	-	y_2
3	-	+	-	y_3
4	+	+	-	y_4
5	-	-	+	y_5
6	+	-	+	y_6
7	-	+	+	y_7
8	+	+	+	y_8

Геометрической интерпретацией ПФЭ 2^2 является квадрат в факторной плоскости (рис. 2.1, а), ПФЭ 2^3 – куб (рис. 2.1, б).

Здесь нормированные координаты x_1 и x_2 проходят через точку пересечения основных уровней факторов, и масштаб их осей выбран так, чтобы интервал варьирования равнялся 1. Тогда условия проведения опытов в МП эксперимента будут соответствовать вершинами квадрата, центром которого является основной уровень. Если $n > 3$, то фигуру, задающую в многомерном пространстве область эксперимента, называют *гиперкубом*.

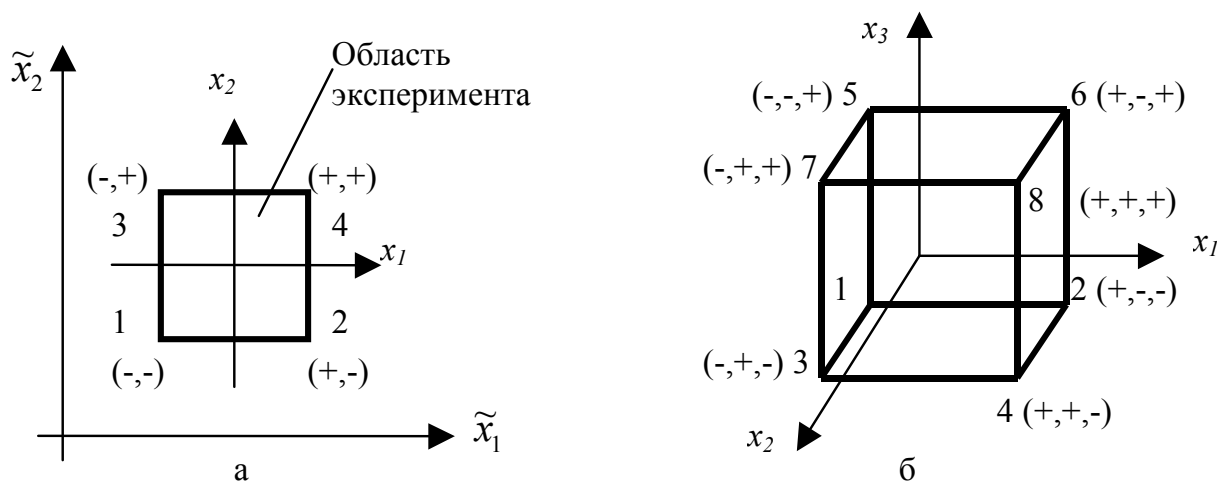


Рис. 2.1. Геометрическая интерпретация ПФЭ

Влияние факторов на выходной параметр может зависеть от уровня, на котором находится другой фактор, или от сочетания уровней нескольких факторов. Если априорно не известно, что такой зависимости между факторами нет, то строят развернутую МП, учитывающую не только факторы, но и их взаимодействия. При этом знаки в столбцах для взаимодействий получают перемножением знаков взаимодействующих факторов. Пример развернутой МП для ПФЭ дан в табл. 2.4.

Таблица 2.4

Развернутая МП для ПФЭ типа 2^3

N	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y
1	+	-	-	-	+	+	+	-	y_1
2	+	+	-	-	-	-	+	+	y_2
3	+	-	+	-	-	+	-	+	y_3
4	+	+	+	-	+	-	-	-	y_4
5	+	-	-	+	+	-	-	+	y_5
6	+	+	-	+	-	+	-	-	y_6
7	+	-	+	+	-	-	+	-	y_7
8	+	+	+	+	+	+	+	+	y_8

Фиктивный фактор x_0 вводят для удобства машинного расчета свободного члена b_0 (для идентичности формул).

Основные свойства МП эксперимента:

- а) симметричность относительно центра эксперимента

$$\sum_{j=1}^N x_{ij} = 0,$$

где i – номер фактора; j – номер опыта; N – число опытов;

б) условие нормировки

$$\sum_{j=1}^N |x_{ij}| = N;$$

в) ортогональность

$$\sum_{j=1}^N x_{ij} x_{fj} = 0,$$

если $i \neq f$.

Свойство ортогональности позволяет упростить вычисления и получить независимые оценки коэффициентов регрессии. Это означает, в частности, что замена нулем любого коэффициента в уравнении ММ не изменит оценок остальных коэффициентов. Это свойство может быть полезным, когда точный вид модели не известен и требуется по экспериментальным данным отобрать факторы, существенно влияющие на исследуемый параметр. Если условие ортогональности не выполняется, после исключения каждого незначимого коэффициента необходимо пересчитывать оценки оставшихся коэффициентов и их дисперсии. При этом могут измениться как доверительные интервалы, так и выводы относительно коэффициентов значимости;

г) рототабельность – свойство равноточного предсказания исследуемого параметра на равных расстояниях от центра эксперимента вне зависимости от направления.

Матрица, удовлетворяющая условиям симметричности, нормировки и ортогональности, называется *оптимальной*.

МП ПФЭ является оптимальной для линейных ММ. Если же ММ содержит взаимодействия, то свойство рототабельности не выполняется.

2.3. Порядок постановки ПФЭ

Для оценки точности эксперимента для каждой i -й точки факторного пространства (для каждого сочетания уровней факторов МП) проводят K опытов. В результате получают значения $y_{i1}, y_{i2}, \dots, y_{iK}$ исследуемого параметра, для которых находят среднее значение

$$\bar{y} = \frac{1}{K} \sum_{t=1}^K y_{it} \quad (2.3)$$

При этом опыты в одной точке проводят не подряд, а обходят все точки в первой серии опытов, затем во второй, и так далее до K -й. Для уменьшения влияния внешней среды и неконтролируемых факторов внутри каждой серии точки факторного пространства обходят случайным образом – рандомизируют последовательность опытов. Рандомизацию опытов можно провести с помощью генератора случайных чисел или таблицы случайных чисел (см. приложение А).

Например, в случае постановки двух серий опытов для экспериментов 2^3 получим с учетом данных таблицы такие последовательности:

1 серия	2 серия
4, 2, 3, 7, 8, 1, 5, 6	2, 4, 6, 8, 5, 7, 3, 1

Это означает, что в первой серии опытов первым выполняется опыт в точке факторного пространства № 4, вторым – в точке № 2 и т. д. Во второй серии первым выполняется опыт в точке № 2, вторым – в точке № 4 и т. д. (см. табл. 2.5).

2.4. Проверка воспроизводимости опытов (однородности дисперсий)

Опыт считается воспроизводимым, если дисперсия D_{y_i} выходного параметра y_i однородна в каждой точке факторного пространства. Оценка S_{y_i} дисперсии D_{y_i} определяется для каждой точки факторного пространства по формуле:

$$S_{y_i}^2 = \frac{1}{(K-1)} \sum_{t=1}^K (y_{it} - \bar{y}_i)^2 \quad (2.4)$$

МП для двух серий опытов ПФЭ типа 2

Номер точки факторного пространства	Номер опыта		x_0	x_1	x_2	x_3			y_{j1}	y_{j2}	y_j
	Серия один	Серия два									
1	4	2	+	-	-	-			y_{11}	y_{12}	y_{13}
2	2	4	+	+	-	-			y_{21}	y_{22}	y_{23}
3	3	6	+	-	+	-			y_{31}	y_{32}	y_{33}
4	7	8	+	+	+	-			y_{41}	y_{42}	y_{43}
5	8	5	+	-	-	+			y_{51}	y_{52}	y_{53}
6	1	7	+	+	-	+			y_{61}	y_{62}	y_{63}
7	5	3	+	-	+	+			y_{71}	y_{72}	y_{73}
8	6	1	+	+	+	+			y_{81}	y_{82}	y_{83}

Гипотезу однородности (равенства) дисперсий проверяют с помощью критерия Кохрена. Расчетное значение этого критерия определяют по формуле:

$$G_p = \frac{\max S_{y_j}^2}{\sum_{i=1}^N S_{y_j}^2}, \quad (2.5)$$

а его критическое значение $G_{кр}$ находят из таблицы распределения Кохрена по числу степеней свободы числителя $f=K-1$, знаменателя $f=N$ и уровню значимости q (см. приложение Б). Если $G_p < G_{кр}$, гипотеза об однородности дисперсий принимается, в противном случае – отвергается, и тогда эксперимент необходимо повторить, изменив условия его проведения (набор факторов, интервал их варьирования, точность измерительных приборов и пр.). Например, если при варьировании какого-то фактора изменение исследуемого параметра сравнимо с погрешностью эксперимента, то интервал варьирования необходимо увеличивать примерно на порядок.

2.5. Расчет оценок коэффициентов регрессионного уравнения

Расчет оценок коэффициентов уравнения регрессии производится по методу наименьших квадратов, при этом минимизируется сумма квадратов отклонений между экспериментальными значениями исследуемого параметра и значениями, вычисленными для тех же точек факторного пространства по уравнению

регрессии. Благодаря предварительной стандартизации масштаба факторов и ортогональности МП, расчет оценок коэффициентов регрессии в ПФЭ превращается в простую арифметическую процедуру

$$b_i = \frac{1}{N} \sum_{j=1}^N x_i y_j, \quad (2.6)$$

$$b_k = \frac{1}{N} \sum_{j=1}^N x_i \cdot x_k \cdot \overline{y_j}, \quad (2.7)$$

$$b_0 = \frac{1}{N} \sum_{j=1}^N x_0 \cdot \overline{y_j}. \quad (2.8)$$

2.6. Проверка значимости коэффициентов регрессии

Гипотезу о статистической значимости (отличии от нуля) коэффициентов регрессии проверяют по критерию Стьюдента. Расчетное значение t_p этого критерия определяют как частное от деления модуля коэффициента b_i на оценку его среднеквадратического отклонения S_b :

$$t_p = \frac{|b_i|}{S_b} \quad (2.9)$$

В ПФЭ, благодаря одинаковой удаленности всех экспериментальных точек факторного пространства от центра эксперимента, оценки всех коэффициентов уравнения регрессии независимо от их величины вычисляются с одинаковой погрешностью (при выполнении условия воспроизводимости опытов):

$$S_b = \frac{S_y}{N}, \quad (2.10)$$

где S_y – оценка дисперсии воспроизводимости эксперимента,

$$S_y^2 = \frac{\sum_{j=1}^N S y_j}{N}. \quad (2.11)$$

Критическое значение критерия $t_{кр}$ находят из таблицы распределения Стьюдента по числу степеней свободы $f=N(K-1)$ и уровню значимости q (см. приложение В). Если $t_p > t_{кр}$, гипотеза о значимости коэффициента b_i принимает-

ся, в противном случае коэффициент считается незначимым и приравнивается нулю.

Необходимо помнить, что незначимость коэффициента может быть обусловлена и неверным выбором интервала варьирования фактора. Поэтому иногда бывает полезным расширить интервал варьирования и провести новый эксперимент.

2.7. Проверка адекватности полученной ММ

Для проверки гипотезы об адекватности ММ необходимо сравнить две дисперсии:

а) дисперсию неадекватности, зависящую от разности между значениями y_{ip} , рассчитанными по ММ, и экспериментальными результатами y_{it} :

$$S_a^2 = \frac{1}{K(N-L)} \sum_{j=1}^N \sum_{i=1}^K (y_{jp} - y_{it})^2 \quad (2.12)$$

или

$$S_a^2 = \frac{1}{(N-L)} \sum_{j=1}^N (y_{jp} - \bar{y}_j)^2, \quad (2.13)$$

где L – число значимых коэффициентов исследуемого уравнения регрессии, не считая b_0 ;

б) дисперсию неоднородности, характеризующую погрешности наблюдений:

$$S_y^2 = \frac{1}{N} \sum_{i=1}^N S_{y_i}^2. \quad (2.14)$$

Заметим, что дисперсия погрешности наблюдений может быть оценена лишь путем сравнения результатов нескольких параллельных опытов, проводимых в каждой экспериментальной точке.

Адекватность ММ проверяется по F – критерию Фишера. Его расчетное значение находят как частное от деления оценки дисперсии неадекватности на оценку дисперсии единичного наблюдения

$$F_p = \frac{S_a^2}{S_y^2}, \quad (2.15)$$

причем $S_a^2 > S_y^2$.

Если это условие не выполняется, их нужно поменять местами.

Критическое значение $F_{кр}$ находят из таблицы распределения Фишера по числу степеней свободы числителя $f=K(N-L)$, знаменателя $f=N(K-1)$ и уровню значимости q (см. приложение Г). Если $F_p > F_{кр}$ гипотеза об адекватности отклоняется.

Как правило, вначале проверяют адекватность линейной ММ. Если предположение об адекватности подтверждается, то в качестве окончательной ММ выбирают линейную; если отклоняется – добавляют эффект взаимодействия с наибольшим коэффициентом и вновь проверяют гипотезу, и так до тех пор, пока существуют степени свободы.

Если в результате модель все же оказалась неадекватной, это говорит о том, что тип математической модели выбран неудачно и при данном шумовом уровне и классе точности измерительных приборов ММ должна быть уточнена. Для этого следует использовать более сложные модели, например, квадратичные (ортогональное и рототабельное композиционное планирование).

2.8. Переход к физическим переменным

Для записи ММ в реальных физических величинах производят обратный переход от стандартизированного масштаба к натуральному. Это можно сделать с помощью соотношения (2.2). После чего записывают окончательный вид модели.

2.9. Пример использования ПФЭ

Требуется исследовать влияние производственных факторов (U – опорное напряжение (x_1), I – ток потребления (x_2), T – конечная температура нагрева (x_3)) на качество производства магнитных дисков. Номинальное значение факторов: $U = 30$ В, $I = 18$ А, $T = 220$ °С.

Поставим ПФЭ при трех сериях опытов в точках: $U=(30\pm 2)$ В, $I=(18\pm 1)$ А, $T=(220\pm 20)$ °С. Для стандартизации масштабов факторов условия проведения опытов сведем в табл. 2.6.

После составления МП эксперимента и проведения рандомизированных опытов сведем полученные результаты в табл. 2.7, где y – количественный параметр, характеризующий качество обработанной поверхности.

Проведем статистическую обработку полученных результатов. Для проверки по критерию Кохрена (2.5) воспроизводимости опытов при выбранном

уровне значимости $q=0,05$ вычислим в каждой точке факторного пространства среднее значение (2.3) и дисперсию (2.4) исследуемого параметра. Получаемые результаты запишем в табл. 2.7.

Таблица 2.6

Условия проведения ПФЭ

Характеристика плана	$x_1=U$	$x_2=I$	$x_3=T$
Нулевой уровень	30 В	18 А	220 °С
Интервал варьирования	2 В	1 А	20 °С
Верхний уровень	32 В	19 А	240 °С
Нижний уровень	28 В	17 А	200 °С

Таблица 2.7

Пример расчета ПФЭ типа 2³

№ точки фактор. простран.	Номер опыта			x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y_{i1}	y_{i2}	y_{i3}	\bar{y}_i	S_{yi}^2
	серия один	серия два	серия три													
1	5	8	2	+	-	-	-	+	+	+	-	8,18	7,95	9,82	8,65	1,04
2	6	4	5	+	+	-	-	-	-	+	+	18,03	13,42	14,00	15,15	6,31
3	2	5	1	+	-	+	-	-	+	-	+	6,24	8,94	8,37	6,85	1,72
4	8	2	8	+	+	+	-	+	-	-	-	7,06	12,63	7,76	9,15	9,18
5	4	6	3	+	-	-	+	+	-	-	+	7,69	7,22	11,03	8,65	4,32
6	7	3	4	+	+	-	+	-	+	-	-	33,10	30,30	30,65	31,35	2,34
7	3	1	7	+	-	+	+	-	-	+	-	9,21	8,69	12,85	10,25	5,12
8	1	7	6	+	+	+	+	+	+	+	+	20,85	23,16	20,91	21,55	1,97
$\sum_{i=1}^N x_i \bar{y}_i$				121,6	32,0	-26,0	42,0	-5,6	15,2	-10,4	2,8	Критерий Кохрена $G_p=0,29$				
Критерий Стьюдента $f=16 \quad q=0,1 \quad t_{кр} = 1,81$			b_i	15,2	4,1	-3,25	5,28	-0,7	1,9	-1,3	0,35	$G_{кр}=0,52 \quad (f_1=2, \quad f_2=8, \quad q=0,05)$ Вывод: дисперсии однородны				
			t_i	21,71	5,85	4,64	7,50	1,01	2,71	1,85	0,85	Критерий Фишера $q=0,05, \quad f_2=16$				
			Вывод	зн	зн	зн	зн	нз	зн	зн	нз	f_i	F_p	$F_{кр}$	Вывод	
Линейная ММ: $y=15,2+4,1 \cdot x_1-3,25 \cdot x_2+5,25 \cdot x_3$												12	3,31	2,42	ММ не адекватна	
Нелинейная ММ: $y=15,2+4,1 \cdot x_1-3,25 \cdot x_2+5,25 \cdot x_3+1,9 \cdot x_1 \cdot x_3$												9	1,99	2,54	ММ адекватна	
ММ в натуральном масштабе $y=267,95-8,40 U [1/B]-3,25 I [1/A]-1,16 T [1/^\circ C]+0,05 UT [1/(B \cdot ^\circ C)]$ где U – опорное напряжение, В; I – ток потребления, А; T – конечная температура, °С																

Рассчитаем оценки коэффициентов регрессионного уравнения (2.6–2.8) и проверим их статистическую значимость по критерию Стьюдента (2.7) при $q=0,1$. По критерию Фишера (2.15) проверим адекватность линейной модели при $q=0,05$. Поскольку линейная модель неадекватна, будем последовательно добавлять в ММ нелинейные взаимодействия с наибольшими коэффициентами регрессии. Для достижения адекватности ММ оказалось достаточным добавить взаимодействие x_1x_3 . Таким образом, после перехода к физическим переменным получаем искомую ММ:

$$y=267.95-8.40 \cdot U [1/V] -3.25 \cdot I [1/A] -1.16 \cdot T [1/^{\circ}C]+0.05 \cdot U \cdot T [1/(V \ ^{\circ}C)],$$

где U – опорное напряжение, В; I – ток потребления, А; T – конечная температура нагрева, $^{\circ}C$.

3. Порядок выполнения работы

1. В соответствии с индивидуальным заданием необходимо перейти к стандартизированному масштабу факторов, составить МП ПФЭ и проверить ее свойства, рандомизировать опыты.
2. Провести ПФЭ.
3. Проверить воспроизводимость опытов. Если дисперсии неоднородны, повторить эксперимент.
4. Рассчитать оценки коэффициентов регрессионного уравнения.
5. Проверить статистическую значимость коэффициентов регрессии.
6. Проверить адекватность полученной ММ.
7. Перейти к исходным физическим переменным.
8. Записать полученную ММ и сделать выводы.

4. Содержание отчета

Отчет по выполненной работе должен содержать:

1. Постановку задачи и цель работы.
2. Матрицу планирования эксперимента.
3. Результаты проверки воспроизводимости опытов. Результаты расчетов коэффициентов регрессии и проверки их статистической значимости.
4. Результаты проверки адекватности полученной ММ исходными экспериментальными данными.
5. ММ исследуемого объекта в нормированных и физических переменных.

6. Выводы и предложения о ходе дальнейших исследований, составленные на основании анализа ММ.

5. Контрольные вопросы

1. В чем сущность планирования эксперимента? Поясните разницу между активным и пассивным экспериментом.
2. Какие задачи решает теория планирования эксперимента?
3. Что такое факторы оптимизации и какие требования к ним предъявляются? Как выбрать уровни варьирования факторов?
4. Какие требования предъявляются к параметрам оптимизации?
5. В чем сущность ПФЭ и какие ММ он позволяет исследовать?
6. Какую область описывает уравнение регрессии, полученное с помощью ПФЭ и в каких границах его можно использовать?
7. Что такое взаимодействие факторов и сколько их в ПФЭ?
8. В чем сущность и цели стандартизации масштаба факторов?
9. Как составляется и какими свойствами обладает МП ПФЭ?
10. Каков порядок постановки опытов при ПФЭ?
11. Как проверить воспроизводимость опытов?
12. Как рассчитать оценки коэффициентов регрессионного уравнения?
13. Как проверить статистическую значимость оценок коэффициентов регрессии?
14. Как проверить адекватность полученной ММ?
15. Как перейти к исходным физическим переменным?

ЛАБОРАТОРНАЯ РАБОТА № 2 «ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ»

1. Цель работы

Исследовать радиоэлектронное средство (РЭС) с применением дробного факторного эксперимента и получить математическую модель.

2. Дробный факторный эксперимент

Число опытов ПФЭ 2^n быстро растет с увеличением числа факторов n , и при больших n этот вид эксперимента оказывается практически неприемлемым. Для уменьшения числа опытов из множества точек факторного пространства может быть отобрана их некоторая часть, содержащая подходящее число опытов и представляющая собой дробный факторный план.

Дробный факторный эксперимент (ДФЭ), как и ПФЭ, позволяет исследовать полиномиальные ММ вида (2.1). Число оцениваемых параметров ММ и число проводимых в эксперименте опытов связано с понятием насыщенности эксперимента. Если число проводимых опытов превышает число оцениваемых параметров, эксперимент называется *ненасыщенным*, если равно – *насыщенным*, если больше – *сверхнасыщенным*.

Дробным факторным экспериментом называется система опытов, представляющая собой часть ПФЭ, позволяющая рассчитать коэффициенты уравнения регрессии и сократить объем экспериментальных данных.

2.1. Составление матрицы планирования ДФЭ

Для построения МП ДФЭ из имеющихся n факторов отбирают $(n-p)$ основных факторов, для которых строят МП ПФЭ. Эту матрицу дополняют затем p столбцами, соответствующими оставшимся факторам. Уровни дополнительных факторов определяют как поэлементное умножение уровней не менее двух и не более $(n-p)$ основных факторов. Говорят, что ДФЭ – это эксперимент типа 2^{n-p} .

Выбранное для дополнительного фактора произведение называется *генератором плана* (поскольку определяет для дополнительного фактора правило чередования уровней варьирования в МП). Очевидно, что ДФЭ типа 2^{n-p} будет иметь p генераторов.

Например, для ДФЭ типа 2^{3-1} число опытов равно четырем опытам по сравнению с 16 опытами в случае ПФЭ (см. табл. 3.1). При трех основных фак-

торах ДФЭ содержит 8 опытов, а генераторами для дробных планов могут служить произведения x_1x_2 , x_1x_3 , x_2x_3 , $x_1x_2x_3$.

При введении одного дополнительного фактора (ДФЭ типа 2^{4-1}) может использоваться любой из четырех возможных генераторов:

$$x_4 = x_1x_2;$$

$$x_4 = x_1x_3;$$

$$x_4 = x_2x_3;$$

$$x_4 = x_1x_2x_3.$$

В качестве генераторов плана используются незначимые взаимодействия.

Таблица 3.1

МП ДФЭ типа 2^{3-1}

№	x_1	x_2	$x_3 = x_1x_2$
1	–	–	+
2	+	–	–
3	–	+	–
4	+	+	+

Для нахождения математического описания процесса используются определенные части ПФЭ: 1/2, 1/4, 1/8 и т. д.

Эта система опытов называется дробными репликами, а сам метод ДФЭ – *методом дробных реплик*. Возможные дробные реплики от ПФЭ типа 2^4 приведены в табл. 3.2.

Таблица 3.2

МП ПФЭ типа 2^4

№	Факторы				y	Дробные реплики		
	x_1	x_2	x_3	x_4				
1	-	-	-	-	y_1	1/8	1/4	1/2
2	+	-	-	-	y_2			
3	-	+	-	-	y_3	1/8		
4	+	+	-	-	y_4			
5	-	-	+	-	y_5	1/8	1/4	
6	+	-	+	-	y_6			
7	-	+	+	-	y_7	1/8		
8	+	+	+	-	y_8			
9	-	-	-	+	y_9	1/8	1/4	1/2
10	+	-	-	+	y_{10}			
11	-	+	-	+	y_{11}	1/8		
12	+	+	-	+	y_{12}			
13	-	-	+	+	y_{13}	1/8	1/4	
14	+	-	+	+	y_{14}			
15	-	+	+	+	y_{15}	1/8		
16	+	+	+	+	y_{16}			

2.2. Определение смешанности оценок коэффициентов

Составим матрицу ДФЭ для трех факторов.

Таблица 3.3

МП ДФЭ типа 2^{3-1}

№	X_0	x_1	x_2	$x_3=x_1x_2$
1	+	-	-	+
2	+	+	-	-
3	+	-	+	-
4	+	+	+	+

По данному плану мы можем определить коэффициенты регрессии b_0, b_1, b_2, b_3 . Однако коэффициенты регрессии b_1, b_2, b_3 будут смешаны с парными взаимодействиями.

При значительном числе факторов и опытов определение смешанности по МП является трудоемким. Для нахождения, при каких факторах и взаимодействиях оценки коэффициентов будут смешанными, вводят понятие *контраста плана*. Контраст получают умножением обеих частей генератора плана вводимого дополнительного фактора x_j на этот фактор. Например, поскольку для ДФЭ (табл. 3.1) генератор плана $x_3=x_1x_2$, то для контраста получим $x_3^2=x_1x_2x_3$, т.к. $x_i^2=1$, окончательно имеем $1=x_1x_2x_3$. Чтобы определить, с какими факторами и взаимодействиями смешана оценка фактора x_i , необходимо умножить обе части контраста на это фактор. Например, для x_1 имеем: $x_1=x_1^2x_2x_3=x_2x_3$, т. е. b_1 оценивает одновременно β_1 и b_{23} . Записывают это так

$$b_1 \rightarrow \beta_1 + b_{23}.$$

Для x_2 : $x_2=x_1x_2x_3x_2=x_1x_3$, тогда $b_2 \rightarrow \beta_2 + b_{13}$; для x_3 : $x_3=x_1x_2x_3x_3=x_1x_2$, тогда

$$b_3 \rightarrow \beta_3 + b_{13},$$

где β_i – действительные значения коэффициентов b_i .

В зависимости от числа факторов, входящих в контраст, говорят о разрешающей способности ДФЭ. Так, если для ДФЭ типа 2^{4-1} в качестве генератора плана выбрано $x_4=x_1x_2x_3$ (контраст соответственно будет $1=x_1x_2x_3x_4$), то говорят, что у такого эксперимента разрешающая способность равна 4; если генератор $x_1x_2=x_4$ и контраст $1=x_1x_2x_4$, то разрешающая способность равна 3; генераторы плана с наибольшей разрешающей способностью называют *главными* и отдают им предпочтение.

Если вводится не один, а несколько дополнительных факторов, то получаем несколько генераторов плана (для каждого дополнительного фактора свой). В этом случае для определения смешанности оценок используют обобщающий контраст, который строится из отдельных контрастов, а также их произведений во всевозможных сочетаниях. Пусть, например, для ДФЭ 2^{5-2} в качестве генераторов выбраны соотношения $x_4 = x_1x_2$ и $x_5 = x_1x_2x_3$, контрасты будут соответственно $1 = x_1x_2x_4$ и $1 = x_1x_2x_3x_5$, а обобщающий контраст:

$$1 = x_1x_2x_4 = x_1x_2x_3x_5 = x_3x_4x_5.$$

Для определения смешанности перемножаем все составляющие обобщающего контраста на соответствующие факторы:

для x_1 : $x_1 = x_2x_4 = x_2x_3x_5 = x_1x_3x_4x_5$;

для x_4 : $x_4 = x_1x_2 = x_1x_2x_3x_4 = x_3x_5$.

Тогда для смешанности оценок получим:

$$b_1 \rightarrow \beta_1 + \beta_{24} + \beta_{234} + \beta_{1345};$$

$$b_4 \rightarrow \beta_4 + \beta_{12} + \beta_{1234} + \beta_{35}.$$

Необходимо отметить, что следствием уменьшения числа опытов по сравнению с ПФЭ является и уменьшение точности оценок, вызванное их смешанностью.

2.3. Порядок постановки ДФЭ

При ДФЭ стандартизация масштабов факторов, порядок постановки опытов, проверка воспроизводимости опытов, расчет оценок коэффициентов регрессионного уравнения и проверка их статистической значимости, проверка адекватности полученной ММ и переход к физическим переменным производится так же, как и при ПФЭ. Однако необходимо учитывать, что для насыщенного и сверхнасыщенного экспериментов невозможна проверка адекватности ММ, так как для нее уже не остается степеней свободы.

2.4. Пример использования ДФЭ

Продолжим рассмотрение примера, приведенного в подразделе 2.9 для ПФЭ. Воспользовавшись информацией, что взаимодействия факторов x_1x_2 и $x_1x_2x_3$ оказались статистически незначимыми, исследуем влияние на качество поверхности магнитных дисков дополнительных факторов: скорости нагрева V и изотермической выдержки t , поставив для этой цели ДФЭ типа 2^{5-2} . Условия проведения опытов сведем в табл. 3.4. Факторы $x_1x_2x_3$ остаются такими же, как в таблице 2.6.

Для факторов x_4 и x_5 генераторами плана выберем взаимодействия x_1x_2 и $x_1x_2x_3$, тогда контрасты будут соответственно $1 = x_1x_2x_4$ и $1 = x_1x_2x_3x_5$, а обобщающий контраст $1 = x_1x_2x_4 = x_1x_2x_3x_5 = x_3x_4x_5$.

Таблица 3.4

Условия проведения ДФЭ

Характеристика плана	$x_1=U$	$x_2=I$	$x_3=T$	$x_4=V$	$x_5=t$
Нулевой уровень	30 В	18 А	220°C	10 С/сек	80 сек
Интервал варьирования	2 В	1 А	20°C	3 С/сек	15 сек
Верхний уровень	32 В	19 А	240°C	13 С/сек	95 сек
Нижний уровень	28 В	17 А	200°C	7 С/сек	65 сек

Найдем смешанность оценок:

$$b_1 \rightarrow \beta_1 + \beta_{24} + \beta_{235} + \beta_{1345};$$

$$b_2 \rightarrow \beta_2 + \beta_{14} + \beta_{135} + \beta_{2345};$$

$$\begin{aligned}
b_3 &\rightarrow \beta_3 + \beta_{1234} + \beta_{125} + \beta_{45} ; \\
b_4 &\rightarrow \beta_4 + \beta_{12} + \beta_{12345} + \beta_{35} ; \\
b_5 &\rightarrow \beta_5 + \beta_{1245} + \beta_{123} + \beta_{34} ; \\
b_{13} &\rightarrow \beta_{13} + \beta_{234} + \beta_{25} + \beta_{145} ; \\
b_{23} &\rightarrow \beta_{23} + \beta_{134} + \beta_{15} + \beta_{245} .
\end{aligned}$$

После составления МП эксперимента и проведения рандомизированных опытов сведем полученные результаты в табл. 3.5, после чего проведем статистическую обработку полученных результатов. Для проверки по критерию Кохрена (2.5) воспроизводимости опытов при выбранном уровне значимости $q=0,05$ вычислим в каждой точке факторного пространства среднее значение (2.3) и дисперсию (2.4) исследуемого параметра. Получаемые результаты также будем заносить в табл. 3.5.

Таблица 3.5

Пример расчета ДФЭ типа 2⁵⁻²

№ точки фактор. простран.	Номер опыта			x_0	x_1	x_2	x_3	$x_4 = x_1 x_2$	$x_5 = x_1 x_2 x_3$	$x_2 x_3$	$x_1 x_3$	y_{i1}	y_{i2}	y_{i3}	\bar{y}_i	S_{yi}^2	
	серия один	серия два	серия три														
1	6	5	1	+	-	-	-	+	-	+	+	7,87	7,41	11,12	8,80	4,11	
2	1	6	6	+	+	-	-	-	+	+	-	17,23	15,42	15,64	16,09	0,97	
3	8	2	4	+	-	+	-	-	+	-	+	6,55	5,89	11,26	7,67	8,57	
4	2	3	5	+	+	+	-	+	-	-	-	8,49	10,91	8,79	9,40	1,73	
5	3	8	8	+	-	-	+	+	+	-	-	2,16	19,84	22,39	20,81	1,93	
6	7	4	2	+	+	-	+	-	-	-	+	32,14	27,59	28,16	29,32	6,17	
7	5	7	7	+	-	+	+	-	-	+	-	6,85	7,20	9,64	7,88	2,31	
8	4	1	3	+	+	+	+	+	+	+	+	27,19	23,56	23,04	24,62	5,06	
$\sum_{i=1}^N x_i y_i$				124,6	34,33	25,51	40,73	2,73	13,73	-9,75	16,7	Критерий Кохрена $G_p=0,28$					
Критерий Стьюдента $f=16$ $q=0,1$ $t_{кр} =$ 2,12				b_i	15,65	4,29	-3,18	5,09	0,34	1,72	-1,22	2,02	$G_{кр}=0,52$ ($f_1=2, f_2=8, q=0,05$) Вывод: дисперсии однородны				
				t_i	22,84	6,29	4,68	7,47	0,50	2,51	1,79	2,97	Критерий Фишера $q=0,05$, $f_2=16$				
				Вывод	зн	зн	зн	зн	нз	зн	нз	зн	f_1	F_p	$F_{кр}$	Вывод	
Линейная ММ: $y=15,65+4,29 \cdot x_1-3,18 \cdot x_2+5,09 \cdot x_3+1,72 \cdot x_5$												9	5,82	2,54	ММ не адекватна		
Нелинейная ММ: $y=15,65+4,29 \cdot x_1-3,18 \cdot x_2+5,09 \cdot x_3+1,72 \cdot x_5+2,02 \cdot x_1 x_3$												6	4,37	2,74	ММ не адекватна		
При исследовании влияния на качество поверхности магнитных дисков производственных факторов U, I, T, V, t выбранная ММ при уровне значимости 5% оказалась неадекватной исходным экспериментальным данным. Рекомендуется использование ортогонального или рототабельного центральных композиционных планов эксперимента.																	

Рассчитаем оценки коэффициентов регрессионного уравнения (2.6–2.8) и проверим их статистическую значимость по критерию Стьюдента (2.9) при $q=0,05$. По критерию Фишера (2.15) проверим адекватность линейной, а затем нелинейной ММ при $q=0,05$. Поскольку как линейная, так и нелинейная модели оказались неадекватными, делаем вывод о несоответствии выбранной ММ экспериментальным результатам. В дальнейших исследованиях следует использовать более сложные модели, например, квадратичные (ортогональное и рототабельное композиционное планирование).

3. Порядок выполнения работы

1. В соответствии с индивидуальным заданием необходимо перейти к стандартизированному масштабу факторов, составить МП ДФЭ, рандомизировать опыты.
2. Провести ДЭФ.
3. Проверить воспроизводимость опытов. Если дисперсии неоднородны, повторить эксперимент.
4. Рассчитать оценки коэффициентов регрессионного уравнения.
5. Проверить статистическую значимость коэффициентов регрессии.
6. Проверить адекватность полученной ММ.
7. Перейти к исходным физическим переменным.
8. Записать полученную ММ и сделать вывод.

4. Содержание отчета

Отчет по выполненной работе должен содержать:

1. Постановку задачи и цель работы.
2. Матрицу планирования эксперимента.
3. Результаты проверки воспроизводимости опытов.
4. Результаты расчетов коэффициентов регрессии и проверки их статистической значимости.
5. Результаты проверки адекватности полученной ММ исходными экспериментальными данными.
6. ММ исследуемого объекта в кодированных и физических переменных.
7. Генераторы плана. Контрасты плана.
8. Обобщающий контраст плана. Систему смешанности коэффициентов.
9. Выводы и предложения о ходе дальнейших исследований, составленные на основании анализа ММ.

5. Контрольные вопросы

1. В чем сущность планирования эксперимента? Поясните разницу между активным и пассивным экспериментом.
2. Какие задачи решает теория планирования эксперимента?
3. Что такое факторы оптимизации и какие требования к ним предъявляются? Как выбрать уровни варьирования факторов?
4. Какие требования предъявляются к параметрам оптимизации?
5. В чем сущность ДФЭ и какие ММ он позволяет исследовать?
6. Какую область описывает уравнение регрессии, полученное с помощью ДФЭ, и в каких границах его можно использовать?
7. Что такое взаимодействие факторов и сколько их может быть в ДЭФ?
8. В чем сущность и цели стандартизации масштаба факторов?
9. Как составляется и какими свойствами обладает МП ДФЭ?
10. Что такое генератор плана и из каких соображений он выбирается?
11. Что такое контраст плана и что такое обобщающий контраст?
12. Что такое смешанность оценок коэффициентов регрессии и как ее найти?
13. Каков порядок постановки опытов при ДФЭ?
14. Как проверить воспроизводимость опытов?
15. Как рассчитать оценки коэффициентов регрессионного уравнения?
16. Как проверить статистическую значимость оценок коэффициентов регрессии?
17. Как проверить адекватность полученной ММ?
18. Как перейти к исходным физическим переменным?
19. Проведите сравнительный анализ ПФЭ и ДФЭ.

ПРИЛОЖЕНИЕ А

ФРАГМЕНТ ТАБЛИЦЫ СЛУЧАЙНЫХ ЧИСЕЛ

4479	2569	1107	3514	1121	5096	0491	7070	8006	4401	4346	6464
2837	1991	3397	2466	4215	3095	0631	5927	9887	5987	6874	7437
0579	2692	9934	5376	2842	8667	6419	0510	5280	7089	5011	6260
3403	8295	4134	0149	3677	0722	1235	0908	4330	7802	7841	6827
7846	1636	2199	8466	1003	9815	9865	0857	6344	0347	4981	6763
9614	9945	7140	3328	5704	4267	4260	7155	4581	3091	7309	6032
8150	5238	8064	1247	4894	8137	4776	5419	9525	8375	4520	1736
1751	8940	8878	2804	6911	6229	5174	4976	3286	4930	0002	5634
5037	8095	0234	8542	9145	7989	5623	1834	0393	5846	1544	6640
6935	2050	8878	4817	8993	0605	2681	0644	9731	2589	7946	4375
2594	1396	7423	1973	5026	2397	9142	3279	7387	4808	2357	0866
9081	0673	9303	9773	4886	1352	4131	2614	8499	7469	8317	2677
4125	7915	9366	4959	5450	8072	9373	3589	7165	0692	9659	1728
6987	9112	9787	6711	2171	2623	6192	3546	5541	1333	8118	6712
8809	2014	6802	2685	4882	5124	6802	4689	6915	9284	3468	7242
5233	9264	1481	5502	9683	8578	4322	8719	3114	2004	1291	9708
7400	4720	6718	7831	6514	8601	2977	0449	5894	1326	4906	7501
0563	7764	8513	1193	0533	2458	9949	7576	5910	7269	0415	7062
3269	1868	6779	3860	2142	8108	9372	3256	5179	1768	3990	8019
1077	6569	9717	9184	7639	3177	0298	3194	6481	0137	2482	3658
4711	1805	8428	4322	0073	1539	8572	7581	8332	1762	5575	7589
3817	4713	3924	1122	1408	8349	7425	9403	9590	2902	1098	0464
3233	3278	0565	3883	8212	4322	2024	3235	1197	8055	7563	2877
9235	2709	3133	4414	8285	9981	5318	2068	4543	8641	0962	7987
3725	8920	9994	9680	8127	1636	6660	5240	1490	1776	7242	7467
0525	6540	4516	8234	8756	8429	1762	4703	2357	1813	9660	1645
2742	0119	6033	5129	6466	2634	7608	1938	3148	1446	0338	9008
4956	0676	9451	0622	8660	6365	0245	4176	2854	9536	1532	3355
7859	0069	9674	7426	7482	8055	0983	3396	1529	8606	7876	9795
9120	3798	0701	0015	3781	2549	1263	4637	6451	6973	3770	9858
2571	8469	7672	9804	9775	0410	4473	3153	8654	3547	3383	8373
3368	2110	2543	5061	9277	0117	7197	2130	8000	8827	0962	6317
8636	8566	3670	4916	6460	4515	8942	3013	7591	8429	2252	7643
4745	8157	6234	3985	7795	0908	5280	3510	3534	9610	5857	8644
7079	0152	7196	1811	6092	0254	6687	7838	6836	0468	1279	3461
4384	2996	8509	4095	7976	1004	4229	6337	9951	2677	6488	4838

При пользовании таблицей мысленно формируют столбцы, содержащие требуемое количество цифр. Так, для одноразрядного числа (от 0 до 9) столбец должен содержать одну цифру, для двухразрядного (от 0 до 99) – две цифры и т. д. Требуемую последовательность случайных чисел получают, начав с произвольного места сформированного столбца и последовательно выбирая из него числа. Если число по величине выходит за заданные границы, его пропускают. Например, при формировании последовательности из восьми одноразрядных неповторяющихся чисел (от 1 до 8), начав с первой строки первого столбца, получим 4, 2, 3, 7, 8, 1, 5, 6.

ПРИЛОЖЕНИЕ Б

ТАБЛИЦА G-РАСПРЕДЕЛЕНИЯ

G -случайная величина, распределенная по закону Кохрена с числом степеней свободы $f_1=n_1$ для числителя и $f_2=n_2$ для знаменателя. Таблица содержит значения ε , полученные из условия $P(|G|<\varepsilon)=0,95$ (верхняя строка при всех n_2) $P(|G|<\varepsilon)=0,99$ (нижняя строка при тех же n_2).

n_2	n_1	1	2	3	4	5	6	7	8	10
2		0,998	0,975	0,94	0,91	0,86	0,85	0,83	0,82	0,79
		0,999	0,995	0,99	0,98	0,96	0,94	0,92	0,90	0,88
3		0,966	0,87	0,80	0,75	0,71	0,68	0,65	0,63	0,60
		0,991	0,94	0,88	0,83	0,79	0,76	0,71	0,69	0,67
4		0,91	0,77	0,68	0,63	0,59	0,56	0,54	0,52	0,49
		0,97	0,96	0,78	0,72	0,68	0,64	0,61	0,58	0,55
5		0,84	0,68	0,60	0,54	0,51	0,48	0,46	0,44	0,41
		0,93	0,79	0,70	0,63	0,59	0,55	0,52	0,50	0,47
6		0,78	0,62	0,53	0,48	0,45	0,42	0,40	0,38	0,36
		0,88	0,72	0,63	0,56	0,52	0,49	0,46	0,44	0,40
7		0,73	0,56	0,48	0,43	0,39	0,37	0,36	0,34	0,32
		0,84	0,66	0,57	0,51	0,47	0,44	0,46	0,39	0,36
8		0,68	0,52	0,44	0,39	0,36	0,34	0,32	0,30	0,28
		0,79	0,62	0,52	0,46	0,43	0,39	0,37	0,35	0,33
9		0,64	0,48	0,40	0,36	0,33	0,31	0,29	0,28	0,26
		0,75	0,57	0,48	0,43	0,39	0,36	0,34	0,32	0,30
10		0,60	0,45	0,37	0,33	0,30	0,28	0,27	0,25	0,24
		0,72	0,54	0,45	0,39	0,36	0,33	0,31	0,30	0,27
12		0,54	0,39	0,33	0,29	0,26	0,24	0,23	0,22	0,21
		0,65	0,48	0,39	0,34	0,31	0,29	0,27	0,25	0,23
15		0,47	0,33	0,28	0,24	0,22	0,20	0,19	0,18	0,17
		0,57	0,41	0,33	0,29	0,26	0,24	0,22	0,21	0,19
20		0,39	0,27	0,22	0,19	0,17	0,16	0,15	0,14	0,130
		0,48	0,33	0,27	0,23	0,20	0,19	0,17	0,16	0,150
24		0,34	0,24	0,19	0,17	0,15	0,14	0,13	0,116	0,111
		0,42	0,29	0,23	0,20	0,18	0,16	0,15	0,142	0,133
30		0,29	0,20	0,16	0,14	0,12	0,11	0,106	0,100	0,092
		0,36	0,24	0,19	0,16	0,15	0,13	0,123	0,116	0,105
40		0,24	0,16	0,13	0,11	0,10	0,089	0,083	0,078	0,071
		0,29	0,19	0,15	0,13	0,11	0,103	0,095	0,090	0,082
60		0,17	0,11	0,09	0,08	0,068	0,062	0,058	0,055	0,050
		0,22	0,14	0,11	0,09	0,080	0,072	0,067	0,063	0,057
120		0,10	0,06	0,05	0,042	0,037	0,034	0,030	0,029	0,027
		0,027	0,08	0,06	0,049	0,043	0,039	0,036	0,033	0,030

Примечание. Допускается линейная интерполяция по аргументу n_2 . Погрешность интерполяции не превышает 0,01.

ПРИЛОЖЕНИЕ В

ТАБЛИЦА *t*-РАСПРЕДЕЛЕНИЯ

t-случайная величина, распределенная по закону Стъдента с числом степеней свободы $f=n$. Таблица содержит значения ε , полученные из условия $P(|G|<\varepsilon)=1-q$.

<i>n</i>	$1-q$	0,99	0,95	0,90	0,80	0,50	0,20
1		63,657	12,706	6,314	3,078	0,727	0,325
2		9,935	4,303	2,920	1,886	0,617	0,289
3		5,841	3,182	2,353	1,638	0,584	0,277
4		4,604	2,776	2,132	1,533	0,569	0,271
5		4,032	2,571	2,015	1,476	0,559	0,267
6		3,707	2,447	1,943	1,440	0,553	0,265
7		3,499	2,365	1,895	1,415	0,549	0,263
8		3,355	2,306	1,860	1,397	0,546	0,262
9		3,250	2,262	1,833	1,383	0,543	0,261
10		3,169	2,228	1,812	1,372	0,542	0,260
11		3,106	2,201	1,796	1,363	0,540	0,260
12		3,055	2,119	1,782	1,356	0,539	0,259
13		3,012	2,160	1,771	1,350	0,538	0,259
14		2,977	2,145	1,761	1,345	0,537	0,258
15		2,947	2,131	1,753	1,341	0,536	0,258
16		2,921	2,120	1,746	1,337	0,535	0,258
18		2,878	2,101	1,734	1,330	0,534	0,257
20		2,845	2,086	1,725	1,325	0,533	0,257
23		2,807	2,069	1,714	1,319	0,532	0,256
25		2,787	2,060	1,708	1,316	0,531	0,256
30		2,750	2,042	1,697	1,310	0,530	0,256
40		2,704	2,021	1,684	1,303	0,529	0,255
60		2,660	2,000	1,671	1,296	0,527	0,254
100		2,617	1,980	1,658	1,289	0,526	0,254
		2,576	1,960	1,645	1,282	0,524	0,253

Примечание. Допускается интерполяция только по аргументу *n*. Погрешность линейной интерполяции не превышает 0,007.

ПРИЛОЖЕНИЕ Г

ТАБЛИЦА F-РАСПРЕДЕЛЕНИЯ

F –случайная величина, распределенная по закону Фишера с числом степеней свободы $f_1=n_1$ для числителя и $f_2=n_2$ для знаменателя. Таблица содержит значения ε , полученные из условия $P(|F|<\varepsilon)=0,95$ (верхняя строка при всех n_2) $P(|F|<\varepsilon)=0,99$ (нижняя строка при тех же n_2).

n_2	n_1	2	3	4	6	9	12	24	
	1	199,5	215,7	224,0	234,0	241,0	244,9	249,0	254,3
		4999	5403	5625	5859	6022	6106	6235	6366
	2	19,00	19,16	19,25	19,33	19,38	19,41	19,55	19,50
		99,00	99,17	99,25	99,33	99,39	99,42	99,46	99,50
	3	9,55	9,28	9,12	8,94	8,81	8,74	8,64	8,53
		30,82	29,46	28,71	27,99	27,34	27,05	26,60	26,12
	4	6,94	9,59	6,39	6,16	6,00	5,91	5,77	5,63
		18,00	16,69	15,98	15,21	14,66	14,37	13,93	13,46
	5	5,79	5,41	5,19	4,95	4,77	4,68	4,53	4,36
		13,27	12,06	11,39	10,67	10,16	9,89	9,47	9,02
	6	5,14	4,76	4,53	4,28	4,10	4,00	3,84	3,67
		10,52	9,78	9,15	8,47	7,98	7,72	7,31	6,88
	7	4,74	4,35	4,12	3,87	3,68	3,57	3,41	3,23
		9,55	8,45	7,85	7,19	6,72	6,49	6,07	6,65
	8	4,46	4,07	3,84	3,58	3,39	3,28	3,12	2,93
		8,65	7,59	7,01	6,37	5,91	5,67	5,28	4,86
	9	4,26	3,86	3,63	3,37	3,18	3,07	2,90	2,71
		8,02	6,99	6,42	5,80	5,35	5,11	4,73	4,31
	10	4,10	3,71	3,48	3,22	3,02	2,91	2,74	2,54
		7,56	6,55	5,99	5,39	4,94	4,71	4,33	3,91
	11	3,98	5,59	3,36	3,09	2,90	2,79	2,51	2,40
		7,21	6,22	5,76	5,07	4,63	4,40	4,02	3,60
	12	3,88	3,49	3,26	3,00	2,80	2,69	2,50	2,30
		6,93	5,95	5,41	4,82	4,39	4,16	3,78	3,36
	13	3,80	3,41	3,18	2,92	2,71	2,60	2,42	2,21
		6,70	5,74	5,21	4,62	4,19	3,96	3,59	3,17
	14	3,74	3,34	3,11	2,85	2,65	2,53	2,35	2,13
		6,51	5,56	5,04	4,46	4,03	3,80	3,43	3,00

Окончание прил. Г

n_2	n_1	2	3	4	6	9	12	24	
16		3,63	3,24	3,01	2,74	2,54	2,42	2,24	2,01
		6,23	5,29	4,77	4,20	3,78	3,55	3,18	2,75
18		3,55	3,16	2,93	2,66	2,46	2,34	2,15	1,92
		6,01	5,09	4,58	4,01	3,60	3,37	3,00	2,57
20		3,49	3,10	2,87	2,60	2,39	2,28	2,08	1,84
		5,85	4,94	4,43	3,87	3,46	3,23	2,86	2,42
24		3,40	3,01	2,78	2,51	2,30	2,18	1,98	1,37
		5,61	4,72	4,22	3,67	3,26	3,03	2,66	2,21
32		3,29	2,90	2,67	2,40	2,19	2,07	1,86	1,59
		5,34	4,46	3,97	3,43	3,02	2,80	2,42	1,96
48		3,19	2,80	2,57	2,30	2,08	1,96	1,75	1,45
		5,08	4,22	3,74	3,20	2,80	2,58	2,20	1,70
		2,99	2,60	2,37	2,09	1,88	1,75	1,52	1,00
		4,61	3,78	3,32	2,80	2,41	2,18	1,79	1,00

Примечание. Допускается линейная интерполяция по аргументу n_2 и квадратичная по n_1 . Погрешность интерполяции не превышает 0,01.

ПРОВЕРКА ГИПОТЕЗ

Под гипотезой подразумевается некоторое предположение о случайной величине (функции распределения, математической модели и пр.). Примером может служить гипотеза об типе закона распределения.

Проверка статистических гипотез – один из разделов математической статистики. Необходимость выдвижения гипотез возникает при обработке или интерпретации результатов наблюдений. При проверке гипотезы необходимо установить, насколько экспериментальные результаты согласуются с выдвинутой гипотезой, после чего принять или отвергнуть гипотезу.

Правило, в соответствии с которым принимается или отвергается данная гипотеза, называется *статистическим критерием*. Построение критерия сводится к выбору подходящей функции T от результата наблюдений, служащей мерой расхождения между экспериментальными и гипотетическими законами.

При решении вопроса о принятии или отклонении какой-либо гипотезы с помощью какого-либо статистического критерия, основанного на результатах эксперимента, могут быть допущены ошибки двух типов. Ошибка «первого рода» совершается тогда, когда гипотеза отвергается, а на самом деле она верна; «второго рода» – когда гипотеза принимается, а на самом деле она не верна.

Результаты проверки гипотезы никогда не могут служить доказательством абсолютной справедливости и правильности гипотезы. Они означают лишь то, что гипотеза с заданной вероятностью не противоречит результатам эксперимента. Поэтому при проверке гипотезы нужно заранее допустить возможность ошибочного решения. Вероятность того, что гипотеза будет отвергнута, хотя на самом деле она верна, называют *уровнем значимости* и обозначают q . Тогда величина $P = 1 - q$, называемая *статистической надежностью*, характеризует вероятность выполнения статистического критерия при условии, что гипотеза верна. В технических задачах, как правило, выбирают $q = 0,05$ или $0,01$, что соответствует уровням значимости 5% и 1%.

По распределению функции T находят критическое значение $T_{кр}$ такое, что если гипотеза верна, то вероятность неравенства $T > T_{кр}$ равна q . Если $T > T_{кр}$, то считают, что расхождение значимо, и гипотеза отвергается. Результат $T < T_{кр}$ не противоречит гипотезе.

ОРТОГОНАЛЬНОЕ ЦЕНТРАЛЬНОЕ КОМПОЗИЦИОННОЕ ПЛАНИРОВАНИЕ

1. Цель работы

Получить математическую модель исследуемого радиоэлектронного средства (РЭС) или технологического процесса в виде уравнения множественной регрессии второго порядка.

2. Ортогональный центральный композиционный план

Разработка математической модели предусматривает принцип «от простого к более сложному». В имитационной модели в виде полинома этот принцип предусматривает переход от полинома первого порядка

$$y = b_0 + \sum_{i=1}^n b_i \cdot x_i + \sum_{i \neq j} b_{ij} x_i x_j \dots, \quad (2.1)$$

к полиному второго порядка

$$y = b_0 + \sum_{i=1}^n b_i x_i + \sum_{i \neq j} b_{ij} x_i \cdot x_j + \sum_{i=1}^n b_{ii} \cdot x_i^2 + \dots, \quad (2.2)$$

В том случае, если с помощью полного факторного эксперимента не удастся получить адекватного математического описания РЭС или процесса, то переходят к центральным композиционным планам (ЦКП). Наибольшее распространение получили ортогональный и рототабельный ЦКП. В ортогональном центральном композиционном плане (ОЦКП) к опытам полного факторного эксперимента (ПФЭ) или дробного факторного эксперимента нужно добавить опыты в «звездных точках» и опыт в центре плана.

Число опытов ортогонального центрального композиционного плана определяется по следующей формуле:

$$N = N_n + 2n + 1, \quad (2.3)$$

где N_n - число опытов в ядре плана;

n - число факторов.

Ядром плана являются полный факторный эксперимент или дробный факторный эксперимент.

Если число факторов $n \leq 4$, то ядром плана является полный факторный эксперимент и число опытов в ядре плана равно

$$N_n = 2^n. \quad (2.4)$$

Если число факторов $n > 4$, то ядром плана является дробный факторный эксперимент и число опытов в ядре плана равно

$$N_n = 2^{n-p}, \quad (2.5)$$

где p - число генераторов плана или число взаимодействий, замененных факторами, учитываемыми в эксперименте.

Важными свойствами ЦКП является то, что информация, полученная при проведении полного факторного эксперимента или дробного факторного эксперимента, не теряется, а используется в дальнейших исследованиях.

$2n$ - это число опытов в «звездных точках», имеющих координаты:

$$(\pm \alpha, 0, 0, \dots, 0)$$

$$(0, \pm \alpha, 0, \dots, 0)$$

$$(0, 0, 0, \dots, \pm \alpha),$$

где α - величина «звездного плеча».

В центре плана - один опыт. Это точка факторного пространства с координатами $(0, 0, 0, \dots, 0)$.

Для удобства расчетов и анализа результатов переходят к нормированному масштабу факторов. Для i -го фактора

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i0}}{I}, \quad (2.6)$$

где x_i - нормированное значение;

\tilde{x}_i - натуральное значение;

\tilde{x}_{i0} - основной уровень;

I – интервал варьирования.

Интервал варьирования I равен

$$I = [\bar{x}_i - x_{i0}], \quad (2.7)$$

В результате нормировки значение верхнего уровня фактора $x_{iB} = +1$, значение нижнего уровня фактора $x_{iH} = -1$

3. Составление матрицы планирования ОЦКП

Для нахождения условий, обеспечивающих ортогональность, квадратичную модель (2.2) можно записать в виде промежуточной модели

$$y = a_0 + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_{ii} (x_i^2 - \beta) + \sum_{i \neq j} b_{ij} x_{ij} \cdot x_i \cdot x_j + \dots, \quad (3.1)$$

$$\text{где } \beta = \frac{1}{N} \sum_{i=1}^n x_i^2 = \frac{2^{n-p} + 2\alpha^2}{N}. \quad (3.2)$$

Величина β вводится для обеспечения ортогональности плана

$$b_0 = a_0 - \beta \sum_{i=1}^n b_{ii}. \quad (3.3)$$

Матрицу планирования эксперимента для двух факторов ($n=2$) можно представить в виде таблицы 3.1. В качестве ядра плана используется ПФЭ, т. к. $n < 4$.

Число опытов ОЦКП будет равно

$$N = 2^2 + 2 \cdot 2 + 1 = 9.$$

Звездное плечо α можно определить по формуле

$$\alpha = \sqrt{\sqrt{2^{n-p-2}} \left(\sqrt{N} - \sqrt{2^{n-p}} \right)}. \quad (3.4)$$

Значения величин α и β , обеспечивающие ортогональность плана при различных значениях n , приведены в таблице 3.2

x_0 – фиктивный параметр, соответствующий коэффициенту b_0

Таблица 3.1

Матрица ОЦКП для двух факторов

Системы опытов	Номер опыта	x_0	x_1	x_2	$\chi_1^2 - \beta$	$\chi_2^2 - \beta$	$x_1 x_2$	y_1	...	y_k	\bar{y}_j	S_j^2	y_{jp}
Ядро плана ПФЭ	1	+1	+1	+1	$1 - \beta$	$1 - \beta$	+1						
	2	+1	-1	+1	$1 - \beta$	$1 - \beta$	-1						
	3	+1	+1	-1	$1 - \beta$	$1 - \beta$	-1						
	4	+1	-1	-1	$1 - \beta$	$1 - \beta$	+1						
Опыты в звездных точках	5	+1	$+\alpha$	0	$\alpha^2 - \beta$	$-\beta$	0						
	6	+1	$-\alpha$	0	$\alpha^2 - \beta$	$-\beta$	0						
	7	+1	0	$+\alpha$	$-\beta$	$\alpha^2 - \beta$	0						
	8	+1	0	$-\alpha$	$-\beta$	$\alpha^2 - \beta$	0						
Опыт в центре плана	9	+1	0	0	$-\beta$	$-\beta$	0						

Таблица 3.2

Параметры ортогональных ЦКП

n	Ядро плана	N	α	β	Элементы матрицы С			
					c_0	c_1	c_2	c_3
2	2^2	9	1.000	0.6667	0.1111	0.1667	0.5000	0.2500
3	2^3	15	1.215	0.7300	0.6667	0.0913	0.2298	0.1250
4	2^4	25	1.414	0.8000	0.0400	0.0500	0.1250	0.0625
5	2^{5-1}	27	1.547	0.7700	0.0371	0.0481	0.0871	0.0625
6	2^{6-1}	45	1.722	0.8430	0.0222	0.0264	0.0564	0.0313
7	2^{7-1}	79	1.885	0.9000	0.0127	0.0141	0.0389	0.0156
8	2^{8-2}	81	2.001	0.8889	0.0123	0.0139	0.0312	0.0156

Геометрическая интерпретация ортогонального ЦКП для двух факторов приведена на рис. 3.1 и представляет собой квадрат

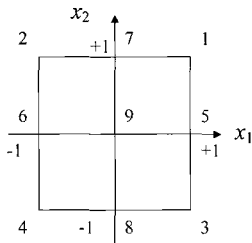


Рис. 3.1 Геометрическая интерпретация ОЦКП для двух факторов

1,2,3,4 – это опыты ПФЭ; 5,6,7,8 – это опыты в «звездных точках». Величина звездного плеча $\alpha=1$ для $n=2$. 9 – это опыт в центре плана. Для $n=3$ геометрической интерпретацией ПФЭ является куб, «звездные точки» ОЦКП лежат за пределами куба, т. к. они расположены на расстоянии большем, чем ± 1 от центра плана, и лежат на поверхности сферы диаметром 2α ($\alpha = 1,215$).

4. Порядок постановки опытов

Для оценки точности эксперимента в каждой j -й точке факторного пространства проводят K опытов. Это так называемые параллельные опыты. В результате получают значения $y_{j1}, y_{j2}, \dots, y_{jk}$ исследуемого параметра, для которых находят среднее значение

$$\bar{y}_j = \frac{1}{k} \sum_{t=1}^k y_{jt}, \quad (4.1)$$

где y_{jt} – исследуемый параметр;

t – номер параллельного опыта, $t=1,2,\dots,k$;

k – число параллельных опытов.

Чтобы исключить влияние систематических ошибок, вызванных влиянием внешней среды и неконтролируемых факторов, рекомендуется случайная последовательность при постановке опытов, которая называется *рандомизацией*. Рандомизацию опытов можно провести с помощью генератора случайных чисел или таблицы случайных чисел, а также с помощью компьютера.

5. Проверка воспроизводимости опытов (однородности дисперсий)

Опыт считается статически воспроизводимым, если дисперсия выходного параметра σ_y^2 однородная (одинаковая) в каждой точке факторного пространства. Оценка дисперсии для каждой j -й точки факторного пространства определяется по формуле

$$S_{y_j}^2 = \frac{1}{k-1} \sum_{i=1}^k (y_{ji} - \bar{y}_j)^2, \quad (5.1)$$

где \bar{y}_j – среднее значение параметра y в j - строке ;

k – число параллельных опытов;

y_{ji} – значение выходного параметра y в j - строке.

Гипотезу о воспроизводимости опытов (об однородности дисперсий) проверяют с помощью критерия Кохрена. Расчетное значение критерия Кохрена вычисляют по формуле

$$G_p = \frac{\max_j S_{y_j}^2}{\sum_{j=1}^N S_{y_j}^2}, \quad (5.2)$$

где N – число опытов.

Критическое значение критерия Кохрена $G_{кр}$ находят из таблицы распределения Кохрена по числу степеней свободы числителя $f_1=k-1$ и

знаменателя $f_2=N$ и уровню значимости α . Уровень значимости α для инженерных расчетов берется равным 0,05 или 0,1.

Таблицы распределения Кохрена приводятся в Приложении А. В таблице распределения Кохрена используется величина $p=1-\alpha$, называемая *статистической надежностью*.

Если $G_{p<} < G_{кр}$ гипотеза об однородности дисперсий принимается, в противном случае отвергается, и тогда эксперимент нужно повторить, изменив условия его проведения (набор факторов, интервал их варьирования, точность измерительных приборов и др.).

6. Расчет оценок коэффициентов регрессионного уравнения

Оценки коэффициентов уравнения регрессии рассчитываются по следующим формулам:

$$b_i = c_1 \sum_{j=1}^N x_{ij} \bar{y}_j, \quad (6.1)$$

$$b_{ii} = c_2 \sum_{j=1}^N [x_{ij}^2 - \beta] \bar{y}_j, \quad (6.2)$$

$$b_{ii} = c_3 \sum_{j=1}^N x_{ij} x_{ij} \bar{y}_j, \quad (6.3)$$

$$b_0 = \frac{1}{N} \sum_{j=1}^N \bar{y}_j - \beta \sum_{i=1}^N b_{ii}, \quad (6.4)$$

$$\alpha_0 = \frac{1}{N} \sum_{j=1}^N \bar{y}_j, \quad (6.5)$$

где c_1, c_2, c_3 – элементы дисперсионной матрицы, значения которых приведены в таблице 3.2, в зависимости от числа факторов n .

7. Проверка значимости коэффициентов регрессии

Проверка значимости коэффициентов уравнения регрессии проводится по t -критерию Стьюдента.

Гипотеза о значимости коэффициентов регрессии принимается, если выполняются следующие неравенства:

$$\begin{aligned}
 |a_0| &> t_{kp} S_{a_0}, \\
 |b_i| &> t_{kp} S_{b_i}, \\
 |b_{ii}| &> t_{kp} S_{b_{ii}}, \\
 |b_{ij}| &> t_{kp} S_{b_{ij}}, \\
 |b_{b_0}| &> t_{kp} S_{b_{b_0}},
 \end{aligned}
 \tag{7.1}$$

где $S_{a_0}, S_{b_i}, S_{b_{ii}}, S_{b_{ij}}, S_{b_{b_0}}$ – оценки дисперсий коэффициентов уравнения регрессии.

Оценки дисперсий коэффициентов уравнения регрессии определяются следующим выражением:

$$\begin{aligned}
 \text{для } a_0 - S_{a_0}^2 &= S^2 c_0, \\
 \text{для } b_i - S_{b_i}^2 &= S^2 c_1, \\
 \text{для } b_{ii} - S_{b_{ii}}^2 &= S^2 c_2, \\
 \text{для } b_{ij} - S_{b_{ij}}^2 &= S^2 c_3, \\
 \text{для } b_{b_0} - S_{b_{b_0}}^2 &= S^2 (c_0 + n\beta^2 c_3),
 \end{aligned}
 \tag{7.2}$$

где S^2 – оценка дисперсии воспроизводимости эксперимента.

Критическое значение критерия Стьюдента $t_{кр}$ находят из таблицы распределения Стьюдента по числу степеней свободы и уровню значимости α (Приложение В).

Число степеней свободы $f=N(k-1)$.

Если неравенства (7.1) не выполняются, коэффициент регрессии считается незначимым и приравнивается нулю. Так как все коэффициенты оцениваются независимо, то изменение оценки любого коэффициента (например, исключение соответствующего коэффициента из уравнения) не приводит к изменению других оценок и их дисперсий. Исключение составляет коэффициент b_0 , т. к. он связан с оценками при квадратичных членах, поэтому их исключение приводит к изменению b_0 . Необходимо помнить, что незначимость коэффициентов может быть обусловлена и неверным выбором интервала варьирования факторов. Поэтому иногда бывает полезным расширить интервал варьирования и провести новый эксперимент.

8. Проверка адекватности полученной математической модели

Проверка адекватности полученного уравнения регрессии экспериментальным данным проводится с помощью критерия Фишера, расчетное значение которого представляет собой следующее отношение

$$F_{расч} = \frac{S_{оо}^2}{S_y^2}, \quad (8.1)$$

где $S_{оо}^2$ - оценка дисперсии неадекватности; S_y^2 - оценка дисперсии воспроизводимости эксперимента.

Оценка дисперсии неадекватности определяется по следующей формуле:

$$S_{оо}^2 = \frac{1}{N-B} \sum (\bar{y}_j - y_{jp})^2, \quad (8.2)$$

где B - число значимых коэффициентов уравнения регрессии;

y_{jp} - расчетное значение функции отклика;

\bar{y}_j - экспериментальное значение функции отклика (см. табл. 3.1).

Если $F_{расч} < F_{крит}$, то гипотеза об адекватности математической модели принимается.

Критическое значение критерия Фишера $F_{крит}$ находится из таблицы распределения Фишера (см. приложение С) по числу степеней для числителя $f=k(N-B)$, знаменателя $f=N(k-1)$ и уровню значимости α .

9. Переход к физической переменной

Для записи математической модели в реальных физических величинах производят обратный переход от стандартизованного масштаба к натуральному. Это можно сделать, используя соотношение (2.6). После чего записывают окончательный вид модели.

10. Пример расчета ортогонального ЦКП

Пусть требуется исследовать влияние производственных факторов (U - опорное напряжение (x_1), I - ток потребления (x_2), T - конечная температура нагрева (x_3)) на качество производства магнитных дисков. Номинальные значения факторов: $U_{II} = 30$ В, $I_H = 18$ А, $T_{II} = 220$ °С.

Составим ортогональный ЦКП для трех серий опытов при интервалах варьирования для $U - 3$ В, $I - 2$ А, $T - 20$ °С. Для стандартизации масштабов факторов условия проведения опытов сведем в табл. 10.1.

После составления МП эксперимента и проведения рандомизированных опытов занесем полученные результаты в табл.10.2, где y_j - количественный параметр, характеризующий качество обработанной поверхности магнитных дисков.

Проведем статистическую обработку полученных материалов. После перехода к физическим переменным получаем искомую ММ:

$$y = -43.03 + 4.81 \times U - 0.08 \times I - 0.06 \times T - 0.06 \times U^2 - 0.05 \times UI + 3.12 \times 10^{-3} \times IT.$$

Условия проведения ОЦКП

Характеристика плана	Стандартный масштаб x_1	Натуральный масштаб		
		$x_1 = U, В$	$x_2 = I, А$	$x_3 = T, ^\circ C$
Нулевой уровень	0	30	18	220
Верхний уровень	+1	33	20	240
Нижний уровень	-1	27	16	200
Звездные точки	+1,215	33,645	20,43	244,3
	-1,215	26,355	15,57	195,7

Таблица 10.2

Пример расчета ортогонального ЦКП

j	x_0	x_1	x_2	x_3	β	$x_1^2 - \beta$	$x_2^2 - \beta$	$x_3^2 - \beta$	x_1x_2	x_1x_3	x_2x_3	$y(j)$	$y(j_2)$	$y(j_3)$	$y(j_4)$	$s^2(j)$
1	+1	-1	-1	-1	+0.27	+0.27	+0.27	+0.27	+1	+1	+1	6.37	6.19	6.27	6.28	0.02
2	+1	+1	-1	-1	+0.27	+0.27	+0.27	+0.27	-1	-1	+1	4.00	3.59	3.87	3.82	0.13
3	+1	-1	+1	-1	+0.27	+0.27	+0.27	+0.27	-1	+1	-1	2.96	3.96	3.75	3.56	0.83
4	+1	+1	+1	-1	+0.27	+0.27	+0.27	+0.27	+1	-1	-1	-1.16	-0.86	-1.82	-1.28	0.72
5	+1	-1	-1	+1	+0.27	+0.27	+0.27	+0.27	+1	-1	-1	5.06	4.87	4.87	4.93	0.04
6	+1	+1	-1	+1	+0.27	+0.27	+0.27	+0.27	-1	+1	-1	2.74	2.94	2.61	2.76	0.08
7	+1	-1	+1	+1	+0.27	+0.27	+0.27	+0.27	-1	-1	+1	2.96	2.44	2.80	2.73	0.21
8	+1	+1	+1	+1	+0.27	+0.27	+0.27	+0.27	+1	+1	+1	-2.46	-2.14	-2.80	-2.47	0.32
9	+1	-1.22	0	0	+0.75	-0.73	-0.73	-0.73	0	0	0	4.04	4.20	4.37	4.20	0.08
10	+1	+1.22	0	0	+0.75	-0.73	-0.73	-0.73	0	0	0	0.39	-0.73	0.76	0.14	1.81
11	+1	0	-1.22	0	-0.73	+0.75	-0.73	-0.73	0	0	0	5.88	5.93	5.68	5.83	0.05
12	+1	0	+1.22	0	-0.73	+0.75	-0.73	-0.73	0	0	0	1.41	1.14	1.07	1.21	0.10
13	+1	0	0	-1.22	-0.73	-0.73	+0.75	+0.75	0	0	0	3.43	4.14	4.39	3.99	0.74
14	+1	0	0	+1.22	-0.73	-0.73	+0.75	+0.75	0	0	0	2.30	3.05	2.61	2.65	0.43
15	+1	0	0	0	-0.73	-0.73	-0.73	-0.73	0	0	0	3.64	2.96	3.65	3.42	0.47
$\sum_{j=1}^k x(j)y(j)$	41.77	-19.6	-20.8	-6.06	-3.74	0.26	-0.33	-5.41	-0.07	-0.07	0.39	Критерий Фишера $\alpha = 0.05, q_2 = 30$				
b_j	2.89	-1.57	-1.69	-0.29	-0.67	0.014	0.014	-0.29	-0.04	0.125	Критерий Стюдента $q = 16; \alpha = 0.1; t_{q\beta} = 1.8$					
t_j	0.842	0.074	0.074	0.187	0.187	0.187	0.187	0.102	0.102	0.102	0.102	q_1	F_p	$F_{q\beta}$	Вывод	
Вывод	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	5	1.1	2.5	ММ	

$y = -43.03 + 4.81x_1 - 0.08x_2 - 0.06x_3 - 0.07x_4^2 - 0.05x_1x_2 + 3.12 \times 10^{-3} x_1x_3$

адекватна

11. Порядок выполнения работы

1. В соответствии с индивидуальным заданием необходимо перейти к стандартизованному масштабу факторов, составить матрицу ортогонального ЦКП и проверить ее свойства.

2. Провести эксперимент (или имитацию эксперимента на ЭВМ).
3. Проверить воспроизводимость опытов. Если дисперсии неоднородны, повторить эксперимент.
4. Рассчитать оценки коэффициентов регрессионного уравнения.
5. Проверить статистическую значимость коэффициентов регрессии.
6. Проверить адекватность полученной ММ.
7. Перейти к исходным физическим переменным.
8. Записать полученную ММ и сделать выводы.

12. Содержание отчета

Отчет о выполненной лабораторной работе должен содержать:

1. Постановку задачи и цель работы.
2. Матрицу планирования эксперимента.
3. Результаты проверки воспроизводимости опытов.
4. Результаты расчетов коэффициентов регрессии, проверку их статистической значимости.
5. Результаты проверки адекватности полученной ММ исходным экспериментальным данным.
6. ММ исследуемого объекта в нормированных и физических переменных.
7. Выводы и предложения о ходе дальнейших исследований, составленные на основании анализа ММ.

13. Контрольные вопросы

1. В чем сущность ортогонального ЦКП и какие ММ он позволяет построить?
2. В чем сущность и цель стандартизации масштаба факторов?
3. Как составляется и какими свойствами обладает МП ортогонального ЦКП?
4. Что такое «звездное плечо» и из каким образом выбирается его значение?
5. Как определяется число опытов ОЦКП?
6. Что является ядром плана в ОЦКП?
7. Что является геометрической интерпретацией ОЦКП для двух факторов и для трех факторов?
8. Каков порядок постановки опытов в ОЦКП?
9. Как проверить воспроизводимость опытов?
10. Как рассчитать оценки коэффициентов регрессионного уравнения?
11. Как проверить статистическую значимость оценок коэффициентов регрессии?
12. Как проверить адекватность полученной ММ?
13. Как перейти к исходным физическим переменным?

РОТОТАБЕЛЬНОЕ ЦЕНТРАЛЬНОЕ КОМПОЗИЦИОННОЕ ПЛАНИРОВАНИЕ

1. Цель работы

Получить математическую модель исследуемого радиоэлектронного средства (РЭС) или технологического процесса в виде уравнения множественной регрессии второго порядка с помощью рототабельного центрального композиционного планирования (РЦКП).

2. Рототабельный центральный композиционный план

Рототабельный центральный композиционный план позволяет получить математическое описание РЭС или технологического процесса в виде уравнения множественной регрессии второго порядка

$$y = b_0 + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_{ii} x_i^2 + \sum_{i \neq j} b_{ij} x_i x_j + \dots, \quad (2.1)$$

где b_0 , b_i , b_{ii} , b_{ij} – коэффициенты уравнения множественной регрессии.

Рототабельные планы, как и ортогональные, являются композиционными. Они позволяют сохранить экспериментальную информацию, полученную с помощью ПФЭ или ДФЭ, которую исследователь затем дополняет опытами в «звездных точках» и в центре плана.

Метод РЦКП позволяет получить более точное математическое описание, что достигается благодаря увеличению числа опытов в центре плана и специальному выбору величины «звездного плеча» α .

Число опытов РЦКП определяется по формуле

$$N = N_n + 2n + N_0,$$

где N_n – число опытов в ядре плана;

n – число факторов;

$2n$ – число опытов в «звездных точках»,

имеющих координаты

$$\begin{pmatrix} \pm\alpha, & 0, & 0, & \dots, & 0 \\ 0, & \pm\alpha, & 0, & & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0, & 0, & 0, & \dots, & \pm\alpha \end{pmatrix},$$

N_0 – число опытов в центре плана с координатами $(0, 0, \dots, 0)$.

Ядром плана является ПФЭ или ДФЭ.

Характеристики РЦКП приведены в таблице 2.1.

Таблица 2.1

Характеристики РЦКП

Число факторов, n	Число опытов в ядре плана, N_0	Величина «звездного плеча», α	Число опытов «звездных точках», $2n$	Число опытов в центре плана, N_0	Общее число опытов, N
2	2^2 (ПФЭ)	1,414	4	5	13
3	2^3 (ПФЭ)	1,682	6	6	20
4	2^4 (ПФЭ)	2,000	8	7	31
5	2^5 (ПФЭ)	2,378	10	10	52
5	2^{5-1} (ДФЭ)	2,000	10	6	32
6	2^6 (ПФЭ)	2,828	12	15	91
6	2^{6-1} (ДФЭ)	2,378	12	9	53
7	2^7 (ПФЭ)	3,333	14	21	163
7	2^{7-1} (ДФЭ)	2,828	14	14	92

Чтобы композиционный план был рототабельный, величина «звездного плеча» α выбирается из условий

$$\alpha = 2^{n/4} \text{ при } n < 5, \quad (2.2)$$

$$\alpha = 2^{(n-1)/4} \text{ при } n \geq 5. \quad (2.3)$$

3. Составление матрицы планирования эксперимента

Матрица планирования строится так же, как и при ортогональном плане. Отличие состоит только в том, что число опытов в центре плана выбирается из условия, чтобы информация о значениях выходной переменной оставалась неизменной (или почти неизменной) для точек сферы единичного радиуса, т.е. чтобы информационный профиль мало отличался внутри этой сферы. Планы, удовлетворяющие этому условию, называются рототабельными униформ-планами. Матрица РЦКП для двух факторов приведена в таблице 3.1.

Таблица 3.1

Матрица РЦКП для двух факторов

Системы опытов	Номер опыта	x_0	x_1	x_2	$x_1 \cdot x_2$	x_1^2	x_2^2	y_j
ПФЭ	1	+1	+1	+1	+1	+1	+1	y_1
	2	+1	-1	+1	-1	+1	+1	y_2
	3	+1	+1	-1	-1	+1	+1	y_3
	4	+1	-1	-1	+1	+1	+1	y_4
Опыты в «звездных точках»	5	+1	+1,41	0	0	2	0	y_5
	6	+1	-1,41	0	0	2	0	y_6
	7	+1	0	+1,41	0	0	2	y_7
	8	+1	0	-1,41	0	0	2	y_8
Опыт в центре плана	9	+1	0	0	0	0	0	y_9
	10	+1	0	0	0	0	0	y_{10}
	11	+1	0	0	0	0	0	y_{11}
	12	+1	0	0	0	0	0	y_{12}
	13	+1	0	0	0	0	0	y_{13}

Матрица, приведенная в таблице 3.1, не обладает свойством ортогональности.

Геометрическая интерпретация РЦКП для двух факторов представлена на рис. 3.1 и представляет собой квадрат

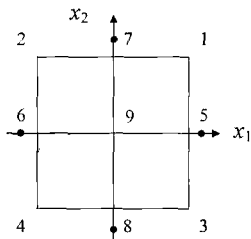


Рис. 3.1. Геометрическая интерпретация РЦКП для двух факторов

4. Порядок проведения эксперимента

При рототабельном центральном композиционном планировании стандартизация масштабов факторов, порядок постановки опытов, проверка воспроизводимости опытов проводятся так же, как и при ортогональном центральном композиционном планировании. Несколько отличаются соотношения для расчета оценок коэффициентов регрессионного уравнения и их дисперсий. При реализации рототабельных планов можно отказаться от постановки *параллельных опытов для оценки воспроизводимости* эксперимента, что уменьшает число опытов по сравнению с ОЦКП. Дисперсия воспроизводимости может быть оценена в этом случае по экспериментам в центре плана.

5. Расчет оценок коэффициентов уравнения регрессии

Формулы для расчета коэффициентов полинома и их дисперсий при РЦКП значительно сложнее, чем при ОЦКП. При рототабельном планировании для вычисления оценок коэффициентов регрессии и

соответствующих оценок дисперсии находят следующие вспомогательные величины:

$$A = \frac{1}{2B[(n+2)B-n]}, \quad (5.1)$$

$$B = \frac{nN}{(n+2)(N-N_0)}, \quad (5.2)$$

$$C = \frac{N}{N-N_0}. \quad (5.3)$$

На основании результатов эксперимента вычисляют следующие вспомогательные суммы:

$$S_0 = \sum_{j=1}^N y_j, \quad (5.4)$$

$$S_i = \sum_{j=1}^N x_{ij} y_j \quad i=1, 2, \dots, n, \quad (5.5)$$

$$S_{il} = \sum_{j=1}^N x_{ij} x_{il} y_j \quad i \neq l, \quad (5.6)$$

$$S_{ii} = \sum_{j=1}^N x_{ij}^2 y_j. \quad (5.7)$$

Коэффициенты регрессии рассчитываются по следующим формулам:

$$b_0 = \frac{2A \cdot B}{N} \left[S_0 B(n+2) - C \sum_{i=1}^n S_{ii} \right], \quad (5.8)$$

$$b_i = \frac{CS_i}{N}, \quad (5.9)$$

$$b_{il} = \frac{C^2 S_{il}}{BN} \quad i \neq l, \quad (5.10)$$

$$b_n = \frac{AC}{N} \left\{ S_n C [(n+2)B-n] + C(1-B) \sum_{i=1}^n S_{ii} - 2BS_0 \right\}. \quad (5.11)$$

Оценку дисперсии воспроизводимости S_y^2 можно найти на основании результатов опытов, проведенных в центре плана. Вычислим среднее значение \bar{y} :

$$\bar{y} = \frac{1}{N_0} \sum_{j=1}^{N_0} y_j, \quad (5.12)$$

тогда
$$S_y^2 = \frac{1}{N_0 - 1} \sum_{j=1}^{N_0} (y_j - \bar{y})^2.$$

Эта величина найдена для числа степеней свободы $f = N_0 - 1$.

6. Проверка значимости коэффициентов регрессии

Оценка значимости коэффициентов уравнения регрессии проводится по критерию Стьюдента.

Коэффициент значим, если выполняется условие

$$|b_i| \geq S_{b_i} \cdot t_{\alpha p}, \quad (6.1)$$

где S_{b_i} — оценка среднеквадратичных отклонений соответствующих коэффициентов регрессии; $t_{\alpha p}$ — критическое значение критерия Стьюдента, которое находится по таблице Приложения В при выбранном уровне значимости α и числе степеней свободы

$$f = N(K - 1), \quad (6.2)$$

где K — число параллельных опытов. Если параллельные опыты не проводились, то число степеней свободы равно $f = N_0 - 1$. Оценки дисперсий коэффициентов регрессии определяются по следующим формулам:

$$S_{b_0}^2 = \frac{2AB(n+2)}{N} \cdot S_y^2, \quad (6.3)$$

$$S_{b_i}^2 = \frac{S_y^2}{N - N_0}, \quad (6.4)$$

$$S_{b_{ij}} = \frac{c^2 S_y^2}{N}, \quad (6.5)$$

$$S_{b_0}^2 = \frac{Ac^2 \cdot S_y^2}{N} [B(n+1) - (n-1)], \quad (6.6)$$

Коэффициенты регрессии значимы, если выполняются следующие соотношения:

$$\begin{aligned} |b_0| &\geq t_{кр} \cdot S_{b_0}, \\ |b_i| &\geq t_{кр} \cdot S_{b_i}, \\ |b_{ii}| &\geq t_{кр} \cdot S_{b_{ii}}, \\ |b_{ij}| &\geq t_{кр} \cdot S_{b_{ij}}. \end{aligned} \quad (6.7)$$

Необходимо помнить, что при рототабельном ЦКП оценки коэффициентов при линейных членах и парных взаимодействиях некоррелированы с оценками остальных коэффициентов, а при квадратичных членах – коррелированы между собой и оценкой свободного члена. Исключение любого из квадратичных членов приводит к изменению оценок остальных, а также оценки свободного члена b_0 .

7. Проверка адекватности полученной математической модели

Проверка адекватности полученной математической модели проводится по критерию Фишера.

Расчетное значение критерия Фишера определяется соотношением

$$Fp = \frac{S_a^2}{S_y^2}, \quad (7.1)$$

где S_a^2 – оценка дисперсии неадекватности;

S_y^2 – оценка дисперсии воспроизводимости.

Оценку дисперсии неадекватности рассчитывают по формуле

$$S_a^2 = \frac{\sum_{j=1}^N (y_j^3 - y_j^p)^2 - S_y^2 (N_a - 1)}{N - \frac{(n+2)(n+1)}{2} - (N_0 + 1)}. \quad (7.2)$$

где y_j^3 и y_j^p – экспериментальное и расчетное значение функции отклика в j -м опыте

Число степеней свободы, связанное с оценкой этой дисперсии,

$$f_a = N - \frac{(n+2)(n+1)}{2}. \quad (7.3)$$

Если $F_p \leq F_{sp}$, математическая модель адекватно описывает поверхность отклика и ее можно использовать для целей оптимизации процесса.

Критическое значение критерия Фишера находится из приложения С для заданного уровня значимости $\alpha = 1 - P$, где P – доверительная вероятность.

8. Переход к физическим переменным

Для записи математической модели в реальных физических величинах производят обратный переход от стандартного масштаба к натуральному, используя соотношение

$$x_i = \frac{(\tilde{x}_i - x_{i0})}{I},$$

где x_i – нормированное значение;

\tilde{x}_i – натуральное (физическое) значение;

x_{i0} – основной уровень;

I – интервал варьирования. $I = |\tilde{x}_i - x_{i0}|$.

9. Пример расчета рототабельного ЦКП

Пусть требуется исследовать влияние производственных факторов (U – опорное напряжение (x_1), I – ток потребления (x_2), T – конечная температура нагрева (x_3)) на качество производства магнитных дисков. Номинальные значения факторов: $U_H = 30$ В, $I_H = 18$ А, $T_H = 220$ °С.

Составим рототабельный ЦКП для трех серий опытов при интервалах варьирования для U – 3В, I – 2А, T – 20 °С. Для стандартизации масштабов факторов условия проведения опытов сведем в табл. 9.1.

После составления МП эксперимента и проведения рандомизированны опытов занесем полученные результаты в табл. 9.2, где y_i – количественный параметр, характеризующий качество обработанной поверхности магнитных дисков.

Таблица 9.1

Условия проведения РЦКП

Характеристика плана	Стандартный масштаб x_i	Натуральный масштаб		
		$x_1 = U, \text{ В}$	$x_2 = I, \text{ А}$	$x_3 = T, \text{ °С}$
Нулевой уровень	0	30	18	220
Верхний уровень	+1	33	20	240
Нижний уровень	-1	27	16	200
Звездные точки	+1,682	35,046	21,364	253,64
	-1,682	24,954	14,636	186,36

Проведем статистическую обработку полученных материалов. Полученные результаты сведем в табл. 9.2.

10. Порядок выполнения работы

1. В соответствии с индивидуальным заданием необходимо перейти к стандартизованному масштабу факторов, составить матрицу рототабельного ЦКП и проверить ее свойства.

2. Провести эксперимент (или имитацию эксперимента на ЭВМ).

3. Проверить воспроизводимость опытов. Если дисперсии неоднородны, повторить эксперимент.

4. Рассчитать оценки коэффициентов регрессионного уравнения.

5. Проверить статистическую значимость коэффициентов регрессии.

6. Проверить адекватность полученной ММ.

7. Перейти к исходным физическим переменным.

8. Записать полученную ММ и сделать выводы.

11. Содержание отчета

Отчет о выполненной лабораторной работе должен содержать:

1. Постановку задачи и цель работы.

2. Матрицу планирования эксперимента.

3. Результаты проверки воспроизводимости опытов.

4. Результаты расчетов коэффициентов регрессии проверки их статистической значимости.

5. Результаты проверки адекватности полученной ММ исходным экспериментальным данным.

6. ММ исследуемого объекта в нормированных и физических переменных.

7. Выводы и предложения о ходе дальнейших исследований, составленные на основании анализа ММ.

12. Контрольные вопросы

1. В чем сущность рототабельного ЦКП и какие ММ он позволяет построить?
2. В чем сущность и цель стандартизации масштаба факторов?
3. Как составляется и какими свойствами обладает МП рототабельного ЦКП?
4. Как определяется число опытов РЦКП?
5. Что такое «звездное плечо» и из каких соображений выбирается его значение?
6. Как проверить воспроизводимость опытов?
7. Как определяется оценка дисперсии воспроизводимости?
8. Как рассчитать оценки коэффициентов регрессионного уравнения?
9. Как проверить статистическую значимость оценок коэффициентов регрессии?
10. Как проверить адекватность полученной ММ?
11. Как перейти к исходным физическим переменным?
12. В чем различия планов РЦКП и ОЦКП?

Пример расчета ротатбельного ЦКП (продолжение)

Таблица 9.3

		Критерий Стьюдента $q=12; \alpha=0,1; t_{гр}=1,8$												
		x_0	x_1	x_2	x_3	x_4^1	x_5^2	x_6^3	x_7^4	x_8^5	x_9^6	x_{10}^7	x_{11}^8	x_{12}^9
\bar{b}_i		4,411	2,067	-0,987	-0,533	-0,58	0,369	-0,003	0,606	-0,985	-0,754			
t_i		629,9	483	230,6	125,1	143	91	0,63	82,9	134,8	103,3			
вывод	зн	зн	зн	зн	зн	зн	зн	вз	зн	зн	зн	зн	зн	зн

Так как коэффициент \bar{b}_3 не значим и требует исключения из ММ, значение остальных коэффициентов и оценки их дисперсий необходимо пересчитать.

Пример расчета ротатбельного ЦКП (продолжение)

Таблица 9.4

		Критерий Стьюдента $q=16; \alpha=0,1; t_{гр}=1,8$											Критерий Фишера $\alpha=0,05; q_2=45;$					
		x_0	x_1	x_2	x_3	x_4^1	x_5^2	x_6^3	x_7^4	x_8^5	x_9^6	x_{10}^7	x_{11}^8	x_{12}^9	$F_{гр}$	$F_{кр}$	Вывод	
\bar{b}_i		4,411	2,067	-0,987	-0,535	-0,58	0,369	0,606	-0,985	-0,754								
t_i		629,9	482,6	230,4	124,9	144,5	92,1	82,9	134,8	103,3								
вывод	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	зн	6	1,5	2,3	ММ адекватна

$y = 4,411x_0 + 2,067x_1 - 0,987x_2 - 0,535x_3 - 0,58x_4 + 0,369x_5 + 0,606x_6 - 0,985x_7 - 0,754x_8$

ТАБЛИЦА G-РАСПРЕДЕЛЕНИЯ

G – случайная величина, распределенная по закону Кохрена с числом степеней свободы q_1 для числителя и q_2 для знаменателя. Таблица содержит значения ϵ , полученные из условия $P(|F| < \epsilon) = 0.95$ (верхняя строка для всех q_2) и $P(|F| < \epsilon) = 0.99$ (нижняя строка при тех же q_2).

	$q_1=1$	2	3	4	5	6	7	8	10
$q_2=2$	0,998	0,975	0,94	0,91	0,86	0,85	0,83	0,82	0,79
	0,999	0,995	0,99	0,98	0,96	0,94	0,92	0,90	0,88
3	0,966	0,87	0,80	0,75	0,71	0,68	0,65	0,63	0,60
	0,991	0,94	0,88	0,83	0,79	0,76	0,71	0,69	0,67
4	0,91	0,77	0,68	0,63	0,59	0,56	0,54	0,52	0,49
	0,97	0,86	0,78	0,72	0,68	0,64	0,61	0,58	0,55
5	0,84	0,68	0,60	0,54	0,51	0,48	0,46	0,44	0,41
	0,93	0,79	0,70	0,63	0,59	0,55	0,52	0,50	0,47
6	0,78	0,62	0,53	0,48	0,45	0,42	0,40	0,38	0,36
	0,88	0,72	0,63	0,56	0,52	0,49	0,46	0,44	0,40
7	0,73	0,56	0,48	0,43	0,39	0,37	0,36	0,34	0,32
	0,84	0,66	0,57	0,51	0,47	0,44	0,46	0,39	0,36
8	0,68	0,52	0,44	0,39	0,36	0,34	0,32	0,30	0,28
	0,79	0,62	0,52	0,46	0,35	0,39	0,37	0,35	0,33
9	0,64	0,48	0,40	0,36	0,33	0,31	0,29	0,28	0,26
	0,75	0,57	0,40	0,43	0,39	0,36	0,34	0,32	0,30
10	0,60	0,45	0,37	0,33	0,30	0,28	0,27	0,25	0,24
	0,72	0,54	0,45	0,39	0,36	0,33	0,31	0,30	0,27
12	0,54	0,39	0,33	0,29	0,26	0,24	0,23	0,22	0,21
	0,65	0,48	0,39	0,34	0,31	0,29	0,27	0,25	0,23
15	0,47	0,33	0,28	0,24	0,22	0,20	0,19	0,18	0,17
	0,57	0,41	0,33	0,29	0,26	0,24	0,22	0,21	0,19
20	0,39	0,27	0,22	0,19	0,17	0,16	0,15	0,14	0,13
	0,48	0,33	0,27	0,23	0,20	0,19	0,17	0,16	0,15
24	0,34	0,24	0,19	0,17	0,15	0,14	0,13	0,116	0,11
	0,42	0,29	0,23	0,20	0,18	0,16	0,15	0,142	0,13
30	0,29	0,20	0,16	0,14	0,12	0,11	0,106	0,100	0,09
	0,36	0,24	0,19	0,16	0,15	0,13	0,123	0,116	0,10
40	0,24	0,16	0,13	0,11	0,10	0,089	0,083	0,078	0,07
	0,29	0,19	0,15	0,13	0,11	0,103	0,095	0,090	0,08
60	0,17	0,11	0,09	0,08	0,068	0,062	0,058	0,055	0,05
	0,22	0,14	0,11	0,09	0,080	0,072	0,067	0,063	0,05
120	0,10	0,06	0,05	0,042	0,037	0,034	0,030	0,029	0,02
	0,12	0,08	0,06	0,049	0,043	0,039	0,036	0,033	0,03

Примечание. Допускается линейная интерполяция по аргументу q_2 , с погрешностью, не превышающей 0,01.

ТАБЛИЦА t-РАСПРЕДЕЛЕНИЯ

t – случайная величина, распределенная по закону Стьюдента с числом степеней свободы q . Таблица содержит значения ε , полученные из условия $P(|t| < \varepsilon) = 1 - \alpha$.

	$\alpha=0.99$	0.95	0.90	0.80	0.50	0.20
$q=1$	63.657	12.706	6.314	3.078	0.727	0.325
2	9.935	4.303	2.920	1.886	0.617	0.289
3	5.841	3.182	2.353	1.638	0.584	0.277
4	4.604	2.776	2.132	1.533	0.569	0.271
5	4.032	2.571	2.015	1.476	0.559	0.267
6	3.707	2.447	1.943	1.440	0.553	0.265
7	3.499	2.365	1.895	1.415	0.549	0.263
8	3.355	2.306	1.860	1.397	0.546	0.262
9	3.250	2.262	1.833	1.383	0.543	0.261
10	3.169	2.228	1.812	1.372	0.542	0.260
11	3.106	2.201	1.796	1.363	0.540	0.260
12	3.055	2.119	1.782	1.356	0.539	0.259
13	3.012	2.160	1.771	1.350	0.538	0.259
14	2.977	2.145	1.761	1.345	0.537	0.258
15	2.947	2.131	1.753	1.341	0.536	0.258
16	2.921	2.120	1.746	1.337	0.535	0.258
18	2.878	2.101	1.734	1.330	0.534	0.257
20	2.845	2.086	1.725	1.325	0.533	0.257
21	2.807	2.069	1.714	1.319	0.532	0.256
25	2.787	2.060	1.708	1.316	0.531	0.256
30	2.750	2.042	1.697	1.310	0.530	0.256
40	2.704	2.021	1.684	1.303	0.529	0.255
60	2.660	2.000	1.671	1.296	0.527	0.254
100	2.617	1.980	1.658	1.289	0.526	0.254
120	2.576	1.960	1.645	1.282	0.524	0.253

Примечание. Допускается интерполяция только по аргументу q , погрешность линейной интерполяции не превышает 0,007.

ТАБЛИЦА F-РАСПРЕДЕЛЕНИЯ

F – случайная величина, распределенная по закону Кохрена с числом степеней свободы q_1 для числителя и q_2 для знаменателя. Таблица содержит значения R , полученные из условия $P(|F| < \varepsilon) = 0.95$ (верхняя строка для всех q_2) и $P(|F| < \varepsilon) = 0.99$ (нижняя строка при тех же q_2).

	$n_2=2$	3	4	6	9	12	24	∞
$n_2=1$	199,5	215,7	224,0	234,0	241,0	244,9	249,0	254,3
	4999	5403	5625	5859	6022	6106	6235	6366
2	19,00	19,16	19,25	19,33	19,38	19,41	19,55	19,50
	99,00	99,17	99,25	99,33	99,39	99,42	99,46	99,50
3	9,55	9,28	9,12	8,94	8,81	8,74	8,64	8,53
	30,82	29,46	28,71	27,99	27,34	27,05	26,60	26,12
4	6,94	6,59	6,39	6,16	6,00	5,91	5,77	5,63
	18,00	16,69	15,98	15,21	14,66	14,37	13,93	13,46
5	5,79	5,41	5,19	4,95	4,77	4,68	4,53	4,36
	13,27	12,06	11,39	10,67	10,16	9,89	9,47	9,02
6	5,14	4,76	4,53	4,28	4,10	4,00	3,84	3,67
	10,52	9,78	9,15	8,47	7,98	7,72	7,31	6,88
7	4,74	4,35	4,12	3,87	3,68	3,57	3,41	3,23
	9,55	8,45	7,85	7,19	6,72	6,49	6,07	5,65
8	4,46	4,07	3,84	3,58	3,39	3,28	3,12	2,93
	8,65	7,59	7,01	6,37	5,91	5,67	5,28	4,86
9	4,26	3,86	3,63	3,37	3,18	3,07	2,90	2,71
	8,02	6,99	6,42	5,80	5,35	5,11	4,73	4,31
10	4,10	3,71	3,48	3,22	3,02	2,91	2,74	2,54
	7,56	6,55	5,99	5,39	4,94	4,71	4,33	3,91
11	3,98	3,59	3,36	3,09	2,90	2,79	2,51	2,40
	7,21	6,22	5,76	5,07	4,63	4,40	4,02	3,60
12	3,88	3,49	3,26	3,00	2,80	2,69	2,50	2,30
	6,93	5,95	5,41	4,82	4,39	4,16	3,78	3,36
13	3,80	3,41	3,18	2,92	2,71	2,60	2,42	2,21
	6,70	5,74	5,21	4,62	4,19	3,96	3,59	3,17
14	3,74	3,34	3,11	2,85	2,65	2,53	2,35	2,13
	6,51	5,56	5,04	4,46	4,03	3,80	3,43	3,00
16	3,63	3,24	3,01	2,74	2,54	2,42	2,24	2,01
	6,23	5,29	4,77	4,20	3,78	3,55	3,18	2,75

Окончание прил. С

	$n_1=2$	3	4	6	9	12	24	оО
18	3,55	3,16	2,93	2,66	2,46	2,34	2,15	1,92
	6,01	5,09	4,58	4,01	3,60	3,37	3,00	2,57
20	3,49	3,10	2,87	2,60	2,39	2,28	2,08	1,84
	5,85	4,94	4,43	3,87	3,46	3,23	2,86	2,42
24	3,40	3,01	2,78	2,51	2,30	2,18	1,98	1,37
	5,61	4,72	4,22	3,67	3,26	3,03	2,66	2,21
32	3,29	2,90	2,67	2,40	2,19	2,07	1,86	1,59
	5,34	4,46	3,97	3,43	3,02	2,80	2,42	1,96
43	3,19	2,80	2,57	2,30	2,08	1,96	1,75	1,45
	5,08	4,22	3,74	3,20	2,80	2,58	2,20	1,70
120	2,99	2,60	2,37	2,09	1,88	1,75	1,52	1,00
	4,61	3,78	3,32	2,80	2,41	2,18	1,79	1,00

Примечание. Допускается линейная интерполяция по аргументу q_2 и квадратичная по q_1 с погрешностью 0.01.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

История и направления развития искусственного интеллекта

Профиль подготовки

Искусственный интеллект и бизнес-аналитика

Квалификация выпускника

Магистр

Формы обучения

очная

г. Ульяновск, 2021

ЛЕКЦИИ

Лекция 1. Предпосылки развития науки искусственного интеллекта

1. Самые популярные сферы развития искусственного интеллекта
2. Области применения искусственного интеллекта

Искусственный интеллект (ИИ) – один из главных трендов нашего времени. За последнее десятилетие компьютеры обучали решению все более сложных задач. Теперь они способны выполнять множество вещей, которые ранее казались присущими только человеку. Успех ИИ не стоит на месте. Машины активно покоряют многие области, начиная от идентификации людей в толпе, управления автомобилем на загруженной автомагистрали и заканчивая победами над лучшими игроками в го – игре, которая многие годы казалась чем-то недостижимым для ИИ, – и на этом достижения не заканчиваются. Иногда компьютеры выполняют работу лучше людей. В большинстве своем машины работают быстрее, дольше и никогда не устают.

Конечно же, идея разумных машин совершенно не нова. Без малого 75 лет мы пытались создать компьютеры, способные продемонстрировать хоть толику нашего интеллекта. А концепция автоматов, похожих на человека, и вовсе родилась столетия назад. Мы очарованы собой и своим интеллектом, и нет ничего удивительно в нашем желании наделить «искрой человечности» машины. Сравнение искусственного интеллекта с человеческим вызывает как радость, так и беспокойство. Насколько похожим на нас станет ИИ? Сможет ли он нас заменить, лишит работы, превзойти в играх и творческих начинаниях, придающих смысл нашей жизни?

1. Предпосылки развития науки искусственного интеллекта

Философия искусственного интеллекта

Философия искусственного интеллекта задаётся вопросами о «мышлении машин», эти вопросы отражают интересы различных исследователей искусственного интеллекта, философов, исследователей познавательной (когнитивной) деятельности. Ответы на эти вопросы зависят от того, что понимается под понятиями «интеллект» или «сознание», и какие именно «машины» являются предметом обсуждения.

- Может ли машина мыслить?

- Что считать интеллектом?
- Как лучше представлять и использовать знания и информацию?
- Этические проблемы создания искусственного разума

Этические проблемы создания искусственного разума

Если в будущем машины смогут рассуждать, осознавать себя и иметь чувства, то что тогда делает человека человеком, а машину — машиной?

Если в будущем машины смогут осознавать себя и иметь чувства, возможно ли будет их эксплуатировать или придется наделять их правами?

Если в будущем машины смогут рассуждать, то как сложатся отношения людей и машин? Данный вопрос был не раз рассмотрен в произведениях искусства на примере противостояния людей и машин.

Будет ли человек, которому в результате многочисленных медицинских имплантаций заменили 99 процентов тела, считаться машиной?

Интеллект (от лат. intellectus «восприятие»; «разумение», «понимание»; «понятие», «рассудок») или ум — качество психики, состоящее из способности осознавать новые ситуации, способности к обучению и запоминанию на основе опыта, пониманию и применению абстрактных концепций, и использованию своих знаний для управления окружающей человека средой. Общая способность к познанию и решению проблем, которая объединяет познавательные способности: *ощущение, восприятие, память, представление, мышление, воображение.*

Википедия

Составляющие интеллекта и его роль

Интеллект — это, прежде всего, основа целеполагания, планирования ресурсов и построение стратегии достижения цели.

Интеллект как способность обычно реализуется при помощи других способностей. Таких как:

- *способности познавать, обучаться,*
- *мыслить логически,*
- *систематизировать информацию путём её анализа,*
- *определять её применимость (классифицировать), находить в ней связи, закономерности и отличия, ассоциировать её с подобной и т. д.*

О наличии интеллекта можно говорить при совокупности всех этих способностей, в отдельности каждая из них не формирует интеллект!

К параметрам, формирующим отличительные особенности интеллектуальной системы человека относят:

- объём рабочей памяти, способность к прогнозированию, орудийной деятельности, логике,
- многоуровневую (6 слоев нейронов) иерархию системного отбора ценной информации,
- сознание,
- память.

Часть исследователей интеллекта и рабочей памяти считает, что рабочая память и подвижный интеллект находятся в сильной связи друг с другом и в значительной степени являются эквивалентными конструктами, другие, что хотя эти сущности коррелируются, но являются автономными, как рост и вес. Показано, что индивидуальные различия рабочей памяти объясняют от трети до половины всех индивидуальных различий общего интеллекта.

Искусственный интеллект — свойство интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека ; наука и технология создания интеллектуальных машин, особенно интеллектуальных компьютерных программ.

Одно из частных определений интеллекта, общее для человека и «машины», можно сформулировать так:

«Интеллект — способность системы создавать в ходе самообучения программы (в первую очередь эвристические) для решения задач определённого класса сложности и решать эти задачи»

Идея создания искусственного подобия человека для решения сложных задач и моделирования человеческого разума витала в воздухе еще в древнейшие времена. Так, в древнем Египте была создана «оживающая» механическая статуя бога Амона. У Гомера в «Илиаде» бог Гефест ковал человекоподобные существа-автоматы. В литературе эта идея обыгрывалась многократно: от Галатеи Пигмалиона до Буратино папы Карло. Однако родоначальником искусственного интеллекта считается средневековый испанский философ, математик и поэт Раймонд Луллий, который еще в XIII веке попытался создать механическую машину для решения различных задач, на основе разработанной им всеобщей классификации понятий.

В XVIII веке Лейбниц и Декарт независимо друг от друга продолжили эту идею, предложив универсальные языки классификации всех наук. Эти работы можно считать первыми теоретическими работами в области искусственного интеллекта.

Окончательное рождение искусственного интеллекта как научного направления произошло только после создания ЭВМ в 40-х годах XX века. В это же

время Нор-берт Винер создал свои основополагающие работы по новой науке — кибернетике.

История искусственного интеллекта как нового научного направления начинается в середине XX века. К этому времени уже было сформировано множество предпосылок его зарождения:

- среди философов давно шли споры о природе человека и процессе познания мира,
- нейрофизиологи и психологи разработали ряд теорий относительно работы человеческого мозга и мышления,
- экономисты и математики задавались вопросами оптимальных расчётов и представления знаний о мире в формализованном виде;
- наконец, зародился фундамент математической теории вычислений — теории алгоритмов — и были созданы первые компьютеры.

Философские предпосылки к возникновению науки

На самую возможность мыслить о понятии «Искусственный интеллект» огромное влияние оказало рождение механистического материализма, которое начинается с работы Рене Декарта «Рассуждение о методе» (1637) и сразу вслед за этим работы Томаса Гоббса «Человеческая природа» (1640).

Рене Декарт предположил, что животное — некий сложный механизм, тем самым сформулировав механистическую теорию.

И тут важно понимать, чем отличается именно механистический материализм, от античного материализма, взгляды которого запечатлены в работах Аристотеля, и последующей диалектики Гегеля, диалектического и исторического материализма (Фейербах, Карл Маркс, Фридрих Энгельс, В. И. Ленин). Дело в том, что механистический материализм направлен на механистическое происхождение организмов, в то время как античный материализм направлен на механистическое происхождение природы, а диалектический и исторический материализм относится к проявлениям механизма в обществе. Поэтому понятно, что без понимания механистичности в организмах не могла идти речь о понимании искусственного интеллекта даже в самом примитивном смысле, а наличие механистичности природы и общества выходят за рамки области об искусственном интеллекте, и строго говоря не являются необходимыми предпосылками.

Технологические предпосылки к возникновению науки

В 1623 г. Вильгельм Шикард (нем. Wilhelm Schickard) построил первую механическую цифровую вычислительную машину, за которой последовали машины Блеза Паскаля (1643) и Лейбница (1671). Лейбниц также был первым,

кто описал современную двоичную систему счисления, хотя до него этой системой периодически увлекались многие великие ученые.

В 1832 году коллежский советник С. Н. Корсаков выдвинул принцип разработки научных методов и устройств для усиления возможностей разума и предложил серию «интеллектуальных машин», в конструкции которых, впервые в истории информатики, применил перфорированные карты.

В XIX веке Чарльз Бэббидж и Ада Лавлейс работали над программируемой механической вычислительной машиной.

В 1910—1913 гг. Бертран Рассел и А. Н. Уайтхед опубликовали работу «Принципы математики», которая произвела революцию в формальной логике.

В 1941 Конрад Цузе построил первый работающий программно-управляемый компьютер.

Уоррен Маккалок и Уолтер Питтс в 1943 опубликовали *A Logical Calculus of the Ideas Immanent in Nervous Activity*, который заложил основы нейронных сетей.

Автоматы: от андроидов до роботов

Первым, кто представил чертеж человекоподобного робота, был великий Леонардо да Винчи примерно в 1495 году. Чертеж представлял модель механического рыцаря, который может сидеть, стоять, двигать руками, головой и, возможно, захватывать предметы. Но так и неизвестно, пытался ли да Винчи воплотить в реальность этот механизм.

В 16-17 веке в Западной Европе инженеры начали конструировать автоматы — заводные механизмы наподобие человека, которые могли выполнять довольно сложные действия. Самый известный из них — робот «испанский монах», который был изобретен примерно в 1560 году механиком Хуанело Турриано для императора Карла V. Автоматон был около 40 см в высоту, способный ходить, бить себя в грудь рукой, кивать головой и даже преподносить деревянный крест к губам.

Более заметный прогресс в робототехнике наблюдался в 18 веке. К примеру, в 1738 году французский инженер Жак де Вокансон собрал первого в мире андроида, способного играть на флейте.

С 19 века изобретения стали приобретать более практический смысл. В 1898 году известный физик Никола Тесла представил общественности миниатюрное радиоуправляемое судно. Первоначально это изобретение казалось немного причудливым. Но в дальнейшем его идеи стали воплощаться в жизнь и приобрели широкое применение.

1921 год — механизмы, наконец, обрели четкий термин «робот» благодаря чешскому писателю Карлу Чапеку и его пьесе под названием «Россумские

Универсальные Роботы». Примечательно, что Чапек назвал этим словом не машины, а живых людей, создаваемых на специальной фабрике. Но термин закрепился в науке и дал жизнь всем автоматизированным устройствам.

В середине 20 века, в частности, в 1950-ых стали разрабатываться механические манипуляторы для взаимодействия с радиоактивными материалами. Эти роботы копировали движения рук человека, находящегося в безопасном месте.

В 1968 году японской компанией Kawasaki Heavy Industries, Ltd был произведен первый промышленный робот. С тех пор Япония начала востребованность стать мировой столицей робототехники, и ей это удалось. Несмотря на то, что роботы изначально разрабатывались в США, они импортировались в Японию в малых количествах, где инженеры изучали их и применяли в производстве.

Коммерческое распространение роботов началось с 1980-ых годов. Технический прогресс двигался в направлении совершенствования систем управления. Такие компании как Unimate, Hitachi KUKA, Westinghouse, FANUC развивали системы датчиков для своих роботов, делая их более чувствительными к задачам, которые они выполняют.

В конце 90-ых – начале 2000-ых начался активный рост и развитие отрасли с использованием новых контроллеров, языков программирования, запуска первых роботов в космос и возникновением машин, создающих роботов.

В это время также появились новые человекоподобные роботы, такие как канадский Aiko, имитирующий человеческие чувства (осознание, слух, речь, зрение), ASIMO – гуманоид японской фирмы Honda, робот-собака AIBO, созданная компанией Sony и другие.

Классические работы

В 1943 году в своей статье «Логическое исчисление идей, относящихся к нервной активности» У. Мак-Каллок и У. Питтс предложили понятие искусственной нейронной сети. В частности, ими была предложена модель искусственного нейрона.

Д. Хебб в работе «Организация поведения» 1949 года описал основные принципы обучения нейронов. Эти идеи несколько лет спустя развил американский нейрофизиолог Фрэнк Розенблатт. Он предложил схему устройства, моделирующего процесс человеческого восприятия, и назвал его «перцептроном».

Среди советских учёных искусственный интеллект был главной областью научной деятельности Д. А. Поспелова. Здесь научные интересы Д. А. Поспелова связаны с моделированием поведения человека, формализацией рассуждений, общими проблемами моделирования жизненных процессов в

естественных и искусственных системах. В частности, Д. А. Поспеловым был впервые в мире разработан подход к принятию решений, опирающийся на семиотические (логико-лингвистические) модели, который послужил теоретической основой ситуационного управления большими системами. По истории также можно проследить интерес других советских учёных к кибернетике.

Этапы развития ИИ

Кратко всю историю искусственного интеллекта можно разбить на следующие периоды:

- *Появление предпосылок искусственного интеллекта (период с 1943 года по 1955 год)*
- *Рождение искусственного интеллекта (1956 год)*
- *Ранний энтузиазм, большие ожидания (период с 1952 года по 1969 год)*
- *Столкновение с реальностью (период с 1966 года по 1973 год)*
- *Системы, основанные на знаниях (период с 1969 года по 1979 год)*
- *Превращение искусственного интеллекта в индустрию (период с 1980 года по настоящее время)*
- *Возвращение к нейронным сетям (период с 1986 года по настоящее время)*
- *Превращение искусственного интеллекта в науку (период с 1987 года по настоящее время)*
- *Появление подхода, основанного на использовании интеллектуальных агентов*

1. Появление предпосылок искусственного интеллекта (период с 1943 года по 1955 год)

Первая работа, которая теперь по общему признанию считается относящейся к искусственному интеллекту, была выполнена Уорреном Мак-Каллоком и Уолтером Питтсом. Они черпали вдохновение из трех источников: знание основ физиологии и назначения нейронов в мозгу; формальный анализ логики высказываний, а также теория вычислений Тьюринга.

В 1951 году два аспиранта факультета математики Принстонского университета, Марвин Минский и Дин Эдмондс, создали первый сетевой компьютер на основе нейронной сети.

Кроме того, можно привести большое количество примеров других ранних работ, которые можно охарактеризовать как относящиеся к искусственному интеллекту, но именно Алан Тьюринг впервые выразил полное представление об искусственном интеллекте в своей статье *Computing Machinery and Intelligence*, которая была опубликована в 1950 году. В этой статье он описал тест Тьюринга, принципы машинного обучения, генетические алгоритмы и обучение с подкреплением.

2. Рождение искусственного интеллекта (1956 год)

Джон Маккарти с другими участниками организовали двухмесячный семинар в Дартмуте летом 1956 года. Дартмутский семинар не привел к появлению каких-либо новых крупных открытий, но позволил познакомиться всем наиболее важным деятелям в научной области исследований интеллекта. Результатом данного семинара было соглашение принять новое название для этой области, предложенное Маккарти, -- искусственный интеллект.

3. Ранний энтузиазм, большие ожидания (период с 1952 года по 1969 год)

Первые годы развития искусственного интеллекта были полны успехов, хотя и достаточно скромных. Если учесть, какими примитивными были в то время компьютеры и тот факт, что компьютеры рассматривались как устройства, способные выполнять только арифметические действия, можно лишь удивляться тому, как удалось заставить компьютер выполнять операции, хоть немного напоминающие разумные.

За первыми успешными разработками Ньюэлла и Саймона последовало создание программы общего решателя задач (*General Problem Solver-- GPS*). Программа GPS была самой первой программой, в которой был воплощен подход к "организации мышления по такому же принципу, как и у человека". Начиная с 1952 года, Артур Самюэл написал ряд программ для игры в шашки. Он опроверг утверждение, что компьютеры способны выполнять только то, чему их учили: одна из его программ быстро научилась играть лучше, чем ее создатель. Эта программа была продемонстрирована по телевидению в феврале 1956 года и произвела очень сильное впечатление на зрителей.

Джон Маккарти перешел из Дартмутского университета в Массачусетский технологический институт и здесь в течение одного исторического 1958 года внес три крайне важных вклада в развитие искусственного интеллекта. Он

привел определение нового языка высокого уровня Lisp, которому суждено было стать доминирующим языком программирования для искусственного интеллекта.

Разработав это язык, Маккарти получил необходимый для него инструмент, но доступ к ограниченным и дорогостоящим компьютерным ресурсам продолжал оставаться серьезной проблемой. В связи с этим он совместно с другими сотрудниками Массачусетского технологического института изобрел режим разделения времени. В том же 1958 году Маккарти опубликовал статью под названием *Programs with Common Sense*, в которой он описал гипотетическую программу *Advice Taker*, которая может рассматриваться как первая полная система искусственного интеллекта. Замечательной особенностью указанной статьи является то, что значительная ее часть не потеряла своего значения и в наши дни.

4. Столкновение с реальностью (период с 1966 года по 1973 год)

С самого начала исследователи искусственного интеллекта не отличались сдержанностью, высказывая прогнозы в отношении своих будущих успехов. Например, часто цитировалась приведенное ниже предсказание Герберта Саймона, опубликованное им в 1957 году.

«Я не ставлю перед собой задачу удивить или шокировать вас, но проще всего я могу подвести итог, сказав, что теперь мы живем в таком мире, где машины могут думать, учиться и создавать. Более того, их способность выполнять эти действия будет продолжать расти до тех пор, пока (в обозримом будущем) круг проблем, с которыми смогут справиться машины, будет сопоставим с тем кругом проблем, где до сих пор был нужен человеческий мозг.»

Такие выражения, как "обозримое будущее", могут интерпретироваться по-разному, но Саймон сделал также более конкретный прогноз, что через десять лет компьютер станет чемпионом мира по шахматам и что машиной будут доказаны все важные математические теоремы. Эти предсказания сбылись не через десять лет, а через сорок. Чрезмерный оптимизм Саймона был обусловлен тем, что первые системы искусственного интеллекта демонстрировали многообещающую производительность, хотя и на простых примерах. Но почти во всех случаях эти ранние системы терпели сокрушительное поражение, сталкиваясь с более широким кругом проблем или с более трудными проблемами.

Сложности первого рода были связаны с тем, что основная часть ранних программ не содержала знаний или имела лишь небольшой объем знаний о своей предметной области.

Сложности второго рода были связаны с неразрешимостью многих проблем, решение которых пытались найти с помощью искусственного интеллекта. Сложности третьего рода возникли в связи с некоторыми фундаментальными ограничениями базовых структур, которые использовались для выработки интеллектуального поведения.

5. Системы, основанные на знаниях (период с 1969 года по 1979 год)

Основной подход к решению задач, сформированный в течение первого десятилетия исследований в области искусственного интеллекта, представлял собой механизм поиска общего назначения, с помощью которого предпринимались попытки связать в единую цепочку элементарные этапы проведения рассуждений для формирования полных решений. Подобные подходы получили название слабых методов, поскольку они не позволяли увеличить масштабы своего применения до уровня более крупных или более сложных экземпляров задач, несмотря на то, что были общими.

Альтернативным по сравнению со слабыми методами стал подход, предусматривающий использование более содержательных знаний, относящихся к проблемной области, который позволяет создавать более длинные цепочки шагов логического вывода и дает возможность проще справиться с теми проблемными ситуациями, которые обычно возникают в специализированных областях знаний.

Одним из первых примеров реализации такого подхода была программа Dendral. Значение программы Dendral состояло в том, что это была первая, успешно созданная экспертная система, основанная на широком использовании знаний.

Ее способность справляться с поставленными задачами была обусловлена применением большого количества правил специального назначения. В более поздних системах также широко применялся основной принцип подхода, реализованного Маккарти в программе Advice Taker, -- четкое отделение знаний (в форме правил) от компонента, обеспечивающего проведение рассуждений.

6. Превращение искусственного интеллекта в индустрию (период с 1980 года по настоящее время)

В индустрии искусственного интеллекта произошел бурный рост, начиная с нескольких миллионов долларов в 1980 году и заканчивая миллиардами долларов в 1988 году. Однако вскоре после этого наступил период, получивший название "зимы искусственного интеллекта", в течение которого постра-

дали многие компании, поскольку не сумели выполнить своих заманчивых обещаний.

7. Возвращение к нейронным сетям (период с 1986 года по настоящее время)

Хотя основная часть специалистов по компьютерным наукам прекратила исследования в области нейронных сетей в конце 1970-х годов, работу в этой области продолжили специалисты из других научных направлений. Психологи, включая Дэвида Румельхарта и Джефа Хинтона, продолжали исследовать модели памяти на основе нейронных сетей.

8. Превращение искусственного интеллекта в науку (период с 1987 года по настоящее время)

В последние годы произошла буквально революция, как в содержании, так и в методологии работ в области искусственного интеллекта. С точки зрения методологии искусственный интеллект наконец-то твердо перешел на научные методы.

9. Появление подхода, основанного на использовании интеллектуальных агентов
Наиболее широко известным примером создания полной архитектуры агента является работа Аллена Ньюэлла, Джона Лэрда и Пола Розенблума над проектом Soar. Для того чтобы проще было разобраться в работе агентов, внедренных в реальную среду с непрерывным потоком сенсорных входных данных, были применены так называемые *ситуационные движения*. Одним из наиболее важных примеров среды для интеллектуальных агентов может служить Internet. Технологии искусственного интеллекта легли в основу многих инструментальных средств Internet, таких как машины поиска, системы, предназначенные для выработки рекомендаций, и системы создания Web-узлов.

Сторонники данного подхода считают, что феномены человеческого поведения, его способность к обучению и адаптации есть следствие именно биологической структуры и особенностей её функционирования. Отличается от понимания искусственного интеллекта по Джону Маккарти, когда исходят из положения о том, что искусственные системы не обязаны повторять в своей структуре и функционировании структуру и протекающие в ней процессы, присущие биологическим системам.

Одним из следствий попыток создания полных агентов стало понимание того, что ранее изолированные подобласти искусственного интеллекта могут потребовать определенной реорганизации, когда возникнет необходимость

снова связать воедино накопленные в них результаты. Поэтому системы проведения рассуждений и планирования должны быть приспособленными к работе в условиях неопределенности. Вторым важным следствием изменения взглядов на роль агентов является то, что исследования в области искусственного интеллекта теперь необходимо проводить в более тесном контакте с другими областями, такими как теория управления и экономика, которые также имеют дело с агентами.

История развития искусственного интеллекта в СССР и России

В 1954 г. в МГУ начал свою работу семинар «Автоматы и мышление» под руководством академика Ляпунова А. А. (1911-1973), одного из основателей российской кибернетики. В этом семинаре принимали участие физиологи, лингвисты, психологи, математики. Принято считать, что именно в это время родился искусственный интеллект в России. Как и за рубежом, выделились два основных направления — нейрокибернетики и кибернетики «черного ящика».

Среди советских учёных искусственный интеллект был главной областью научной деятельности Д. А. Поспелова. Здесь научные интересы Д. А. Поспелова связаны с моделированием поведения человека, формализацией рассуждений, общими проблемами моделирования жизненных процессов в естественных и искусственных системах. В частности, Д. А. Поспеловым был впервые в мире разработан подход к принятию решений, опирающийся на семиотические (логико-лингвистические) модели, который послужил теоретической основой ситуационного управления большими системами. По истории также можно проследить интерес других советских учёных к кибернетике.

В СССР работы в области искусственного интеллекта начались в 1960-х годах. В Московском университете и Академии наук был выполнен ряд пионерских исследований, возглавленных Вениамином Пушкиным и Д. А. Поспеловым. С начала 1960-х М. Л. Цетлин с коллегами разрабатывали вопросы, связанные с обучением конечных автоматов.

В 1964 году была опубликована работа ленинградского логика Сергея Маслова «Обратный метод установления выводимости в классическом исчислении предикатов», в которой впервые предлагался метод автоматического поиска доказательства теорем в исчислении предикатов.

В 1966 году В. Ф. Турчиным был разработан язык рекурсивных функций Рефал.

Большой вклад в становление российской школы ИИ внесли выдающиеся ученые Цетлин М.Л., Пушкин В. Н., Гаврилов М. А, чьи ученики и явились пионерами этой науки в России (например, знаменитая Гавриловская школа).

При том, что отношение к новым наукам в советской России всегда было настороженное, наука с таким «вызывающим» названием тоже не избежала этой участи и была встречена в Академии наук в штыки. К счастью, даже среди членов Академии наук СССР нашлись люди, не испугавшиеся столь необычного словосочетания в качестве названия научного направления. Двое из них сыграли огромную роль в борьбе за признание ИИ в нашей стране. Это были академики А. И. Берг и Г. С. Поспелов.

В 1980–1990 гг. проводятся активные исследования в области представления знаний, разрабатываются языки представления знаний, экспертные системы (более 300).

В 1988 г. создается АИИ — Ассоциация искусственного интеллекта. Ее членами являются более 300 исследователей. Президентом Ассоциации единогласно избирается Д. А. Поспелов, выдающийся ученый, чей вклад в развитие ИИ в России трудно переоценить.

«Официальная» история искусственного интеллекта в России началась в январе 2019 года, когда президент страны Владимир Путин дал поручение правительству разработать подходы к национальной стратегии развития искусственного интеллекта (ИИ) и представить соответствующие предложения. В середине октября президент Путин подписал указ, которым утвердил стратегию развития ИИ в стране до 2030 года. Согласно документу, Россия должна занять одну из ведущих позиций в мире в этой сфере, так как лидер в области ИИ станет, по мнению российского президента, «властелином мира».

Национальная стратегия развития искусственного интеллекта

30 мая 2019 г. на совещании по развитию цифровой экономики под председательством В. В. Путина было принято решение о подготовке национальной стратегии по искусственному интеллекту.

Затраты развитых стран, особенно США, Китая, а также Евросоюза, на технологии искусственного интеллекта растут ударными темпами. Между тем, вплоть до последнего времени Россия оставалась едва ли не последней из крупных стран, не имеющих собственной стратегии развития технологий ИИ. Ситуация начала меняться лишь в конце 2019 года, когда была утверждена **«Национальная стратегия развития искусственного интеллекта на период до 2030 года»**.

Приоритеты развития технологий ИИ в России

К приоритетам развития ИИ в России относится:

1. Ускорение технологического развития РФ за счет увеличения количества организаций, осуществляющих технологические инновации, до 50% от их общего числа;
2. Обеспечение ускоренного внедрения цифровых технологий в экономику и социальную сферу;
3. Создание в базовых отраслях экономики, прежде всего в обрабатывающей промышленности и агропромышленном комплексе, высокопроизводительного экспортоориентированного сектора, развивающегося на основе современных технологий и обеспеченного высококвалифицированными кадрами.

Задачи развития технологий ИИ в России

Стратегия также обозначает ряд задач, которые необходимо решить для успешного развития технологий ИИ в России, в числе которых:

- Создание высокопроизводительных рабочих мест;
- Обеспечение конкурентоспособных условий труда для специалистов в сфере ИИ;
- Привлечение специалистов из-за рубежа;
- Поддержка экспорта продуктов и услуг, созданных с использованием ИИ;
- Создание стимулов для развития корпоративной науки и исследований;
- Формирование комплексной системы безопасности при создании, развитии, внедрении и использовании технологий ИИ.

Первоначально на развитие проектов в области искусственного интеллекта планировалось выделить 125 млрд рублей, из которых почти 90 млрд из бюджета. Однако, существенные коррективы в планы российского правительства внесла эпидемия коронавируса, в результате чего размер федерального финансирования проектов, связанных с ИИ, заметно снизился. Как следствие на **2021–2024** гг. на развитие ИИ в России планируется потратить лишь 29,4 млрд рублей бюджетных средств и 6,9 млрд из внебюджетных источников.

Хотя в Стратегии заявлено, что «Российская Федерация обладает существенным потенциалом для того, чтобы стать одним из международных лидеров в развитии и использовании технологий искусственного интеллекта», достижение этой цели в краткосрочной и среднесрочной перспективе выглядит маловероятным. Ключевым препятствием на пути России в число мировых лидеров в сфере ИИ может стать недостаточное финансирование, особенно на фоне Китая и США.

Национальная стратегия определяет две ключевые точки развития ИИ в России — **2024** и **2030** годы. Предполагается, что к первой дате страна значительно улучшит позиции в этой сфере, а к 2030 году ликвидирует отставание от развитых стран и добьется мирового лидерства в отдельных направлениях, связанных с ИИ. Внедрять технологии ИИ российские власти планируют в том числе через государственные национальные проекты.

Между тем искусственный интеллект уже сейчас используется в России при решении самых разных задач. Например, Сбербанк применяет его при выдаче кредитов, Яндекс – в развитии беспилотного транспорта и в голосовом помощнике «Алиса», Mail.ru – в коммуникациях, Ростех – в сфере производства, а МВД – при распознавании лиц для обеспечения безопасности на улицах Москвы.

Инвестиции США и Китая

Согласно бюджетным документам, федеральное правительство США планировало инвестировать около 4,9 млрд долларов в исследования и разработки в области искусственного интеллекта и машинного обучения в 2020 финансовом году (в одном году, а не на 4 года).

А вот Китай утвердил свою масштабную стратегию развития ИИ ещё в 2017 году. Расходы на нее не раскрываются, но американский Центр национальной безопасности (CNAS) оценивает их «как минимум в десятки миллиардов долларов». Одни лишь администрации городов Тяньцзиня и Шанхая объявляли о создании инвестфондов для развития AI по 100 млрд юаней (\$14,5 млрд) каждый.

Лекция 2. Подходы к пониманию проблемы

Два альтернативных подхода

Все известные попытки создания искусственного интеллекта направлены на преодоление главной проблемы: «Как не только понять ход мыслительных процессов и природу интеллекта человека, но и воплотить все эти механизмы в одной интеллектуальной сущности?». Для решения этой задачи на протяжении многих лет существовало два альтернативных стратегических подхода к разработке ИИ :

- **нейрокибернетика** и
- **кибернетика «черного ящика».**

Эти подходы появились в середине 20 века практически сразу после выделения искусственного интеллекта в отдельную область науки.

Нейрокибернетика

Основную идею этого направления можно сформулировать следующим образом: единственный объект, способный мыслить, – это человеческий мозг. Поэтому любое «мыслящее» устройство должно каким-то образом воспроизводить его структуру.

Таким образом, нейрокибернетика ориентирована на программно-аппаратное моделирование структур, подобных структуре мозга. Физиологами давно установлено, что основой человеческого мозга является большое количество (до 10^{21}) связанных между собой и взаимодействующих нервных клеток — нейронов. Поэтому усилия нейрокибернетики были сосредоточены на создании элементов, аналогичных нейронам, и их объединении в функционирующие системы. Эти системы принято называть **нейронными сетями**, или нейросетями.

Нейрокибернетика — научное направление, изучающее основные закономерности организации и функционирования нейронов и нейронных образований. Основным методом нейрокибернетики является математическое моделирование, при этом данные физиологического эксперимента используются в качестве исходного материала для создания моделей.

Нейрокибернетика имеет широкий спектр приложений — от медико-биологических разработок до создания специализированных **нейрокомпьютеров**.

Первые нейросети были созданы Розенблаттом и Мак-Каллоком в 1956–1965г. Это были попытки создать системы, моделирующие человеческий глаз и его взаимодействие с мозгом. Устройство, созданное ими тогда, получило название **персептрона** (perceptron). Оно умело различать буквы алфавита, но было чувствительно к их написанию. Постепенно в 70–80 годах количество работ по этому направлению искусственного интеллекта стало снижаться. Слишком неутешительны были первые результаты. Авторы объясняли неудачи малой памятью и низким быстродействием существующих в то время компьютеров.

Кибернетика «черного ящика»

В основу этого подхода был положен принцип, противоположный нейрокибернетике: не имеет значения, как устроено «мыслящее» устройство. Главное, чтобы на заданные входные воздействия оно реагировало так же, как человеческий мозг.

В рамках этого направления созданы модели и алгоритмы, которые при решении интеллектуальных задач дают результаты сравнимые с результатами, которые получает человек. При этом модели и алгоритмы могут, как воспро-

изводить процесс принятия решения человеком, так и совершенно отличаться от него.

Модель лабиринтного поиска (конец 50-х годов)

Этот подход представляет задачу как некоторое пространство состояний в форме графа, и в этом графе проводится поиск оптимального пути от входных данных к результирующим. Была проделана большая работа по разработке этой модели, но для решения практических задач эта идея не нашла широкого применения. В первых учебниках по искусственному интеллекту [Хант, 1986; Эндрю, 1985] описаны эти программы — они играют в игру «15», собирают «Ханойскую башню», играют в шашки и шахматы.

Эпоха эвристического программирования (начало 60-х)

Эвристический алгоритм (эвристика) — алгоритм решения задачи, включающий практический метод, не являющийся гарантированно точным или оптимальным, но достаточный для решения поставленной задачи. Позволяет ускорить решение задачи в тех случаях, когда точное решение не может быть найдено. **Эвристический алгоритм** — это алгоритм решения задачи, правильность которого для всех возможных случаев не доказана, но про который известно, что он даёт достаточно хорошее решение в большинстве случаев.

Эвристическое программирование — разработка стратегии действий на основе известных, заранее заданных эвристик.

Использование методов математической логики (1963-1970)

Робинсон разработал метод резолюций, который позволяет автоматически доказывать теоремы при наличии набора исходных аксиом. Примерно в это же время выдающийся отечественный математик Ю. С. Маслов предложил так называемый обратный вывод, впоследствии названный его именем, решающий аналогичную задачу другим способом. На основе метода резолюций француз Альбер Кольмероз в 1973 г. создает язык логического программирования **Пролог**. Большой резонанс имела программа «Логик-теоретик», созданная Ньюэллом, Саймоном и Шоу, которая доказывала школьные теоремы. Однако большинство реальных задач не сводится к набору аксиом, и человек, решая производственные задачи, не использует классическую логику, поэтому логические модели при всех своих преимуществах имеют существенные ограничения по классам решаемых задач.

Системы, основанные на знаниях, или экспертные системы

К середине 1970-х на смену поискам универсального алгоритма мышления пришла идея моделировать конкретные знания специалистов-экспертов. В США появились первые коммерческие системы, основанные на знаниях, или экспертные системы (ЭС). Стал применяться новый подход к решению задач искусственного интеллекта — представление знаний. Созданы MYCIN и DENDRAL, ставшие уже классическими, две первые экспертные системы для медицины и химии.

Начиная с середины 1980-х годов, повсеместно происходит коммерциализация искусственного интеллекта. Растут ежегодные капиталовложения, создаются промышленные экспертные системы. Растет интерес к самообучающимся системам. Издаются десятки научных журналов, ежегодно собираются международные и национальные конференции по различным направлениям ИИ. Искусственный интеллект становится одной из наиболее перспективных и престижных областей информатики (computer science).

Эволюционный подход

В настоящий момент дополнение к двум выше указанным стратегическим подходам появился ещё один, который называется **эволюционный**. Данный подход связан с моделированием процесса эволюции человеческого мозга, что позволяет разбить задачу создания целостного интеллекта на подпроблемы. Как известно, различные виды животных в процессе эволюции получали нервные системы разной сложности. Идея заключается в том, чтобы развить процесс создания искусственного интеллекта постепенно, разбираясь сначала в том, какие механизмы отвечают за тот или иной скачок в интеллектуальности животного, а потом уже использовать полученные результаты для создания искусственного интеллекта человека.

Тест Тьюринга и интуитивный подход

Эмпирический тест был предложен Аланом Тьюрингом в статье «Вычислительные машины и разум», опубликованной в 1950 году в философском журнале «Mind». Целью данного теста является определение возможности искусственного мышления, близкого к человеческому.

Стандартная интерпретация этого теста звучит следующим образом: «Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека в заблуждение, заставив сделать неверный выбор».

Все участники теста не видят друг друга. Если судья не может сказать определённо, кто из собеседников является человеком, то считается, что машина прошла тест. Чтобы протестировать именно интеллект машины, а не её возможность распознавать устную речь, беседа ведётся в режиме «только текст», например, с помощью клавиатуры и экрана (компьютера-посредника). Переписка должна производиться через контролируемые промежутки времени, чтобы судья не мог делать заключения, исходя из скорости ответов. Во времена Тьюринга компьютеры реагировали медленнее человека. Сейчас это правило тоже необходимо, потому что они реагируют гораздо быстрее, чем человек.

Почти все разработанные программы и близко не подошли к прохождению теста. Хотя такие программы, как Элиза (ELIZA), иногда заставляли людей верить, что они говорят с человеком, как, например, в неформальном эксперименте, названном AOLiza, но эти случаи нельзя считать корректным прохождением теста Тьюринга по целому ряду причин.

Машина может избежать лишних вопросов, например, притворившись параноиком, подростком или иностранцем с недостаточным знанием местного языка. Победитель одного из последних конкурсов, организованных по принципу теста Тьюринга, — бот по имени Женя Густман — сумел объединить все три приема, притворяясь тринадцатилетним мальчишкой из Одессы. Конкурс был организован в 2014 году университетом Рединга (Великобритания).

Был создан группой из трёх программистов: Владимира Веселова (родом из России, живёт в Нью-Джерси), Евгения Демченко (родом из Украины) и Сергея Уласеня (родом из России). Разработка программы была начата в Санкт-Петербурге в 2001 году. Чтобы характер и знания Густмана казались более правдоподобными, он представляется пользователям 13-летним мальчиком из Одессы.

Густман с момента его создания принимал участие в ряде соревнований на прохождение теста Тьюринга и несколько раз занимал второе место в соревновании на премию Лёбнера. В июне 2012 года Густман выиграл соревнование в честь 100-летия Алана Тьюринга, сумев убедить 29 % судей, что он человек. 7 июня 2014 года, на конкурсе, посвященном 60-летию со дня смерти Тьюринга, Густман убедил 33 % судей, что он человек, и, по словам Кевина Уорика, стал первым в истории компьютером, прошедшим тест Тьюринга.

Цель искусственного интеллекта, очевидно, заключается не в том, чтобы обмануть людей, а в том, чтобы достигнуть понимания мира и научиться действовать в нем способами, сравнимыми по своей полезности, эффективности

и надежности с человеческой деятельностью. Специалисты отмечают, что тест Тьюринга для этого недостаточен.

Более корректное тестирование подразумевает широкий спектр задач, таких как понимание человеческого языка, способность делать выводы о физическом и умственном состоянии людей, анализ видео на YouTube, владение элементарными научными знаниями и способность к автономному выполнению роботизированных операций.

Гипотеза Ньюэлла — Саймона

Гипотеза Ньюэлла — Саймона (гипотеза о физической символической системе) — предположение, сформулированное Алленом Ньюэллом и Гербертом Саймоном в 1976 году, согласно которому «физическая символическая система имеет необходимые и достаточные средства для производства основных интеллектуальных операций» (под интеллектуальными операциями подразумеваются действия сильного искусственного интеллекта).

Другими словами, предполагается, что без символических вычислений невозможно выполнять осмысленные действия, а способность выполнять символические вычисления вполне достаточна для того, чтобы стать способным выполнять осмысленные действия.

Таким образом, если предполагать, что животное или человек или машина действуют осмысленно, то значит, они каким-то образом выполняют символические вычисления. И наоборот, так как компьютер способен к подобным вычислениям, то на его основе может быть создан искусственный интеллект. Основанием для гипотезы стало успешное применение созданной Ньюэллом и Саймоном программы — универсального решателя задач — для моделирования рассуждений человека.

Символьный подход

Исторически символический подход был первым в эпоху цифровых машин, так как именно после создания Лисп, первого языка символических вычислений, у его автора возникла уверенность в возможности практически приступить к реализации этими средствами интеллекта. Символьный подход позволяет оперировать слабоформализованными представлениями и их смыслами.

Успешность и эффективность решения новых задач зависит от умения выделять только существенную информацию, что требует гибкости в методах абстрагирования. Тогда как обычная программа устанавливает один свой способ интерпретации данных, из-за чего её работа и выглядит предвзятой и чисто механической. Интеллектуальную задачу в этом случае решает только человек, аналитик или программист, не умея доверить этого машине.

В результате создается единственная модель абстрагирования, система конструктивных сущностей и алгоритмов. А гибкость и универсальность выливается в значительные затраты ресурсов для не типичных задач, то есть система от интеллекта возвращается к грубой силе.

Основная особенность символьных вычислений — создание новых правил в процессе выполнения программы. Тогда как возможности не интеллектуальных систем завершаются как раз перед способностью хотя бы обозначать вновь возникающие трудности. Тем более эти трудности не решаются и наконец компьютер не совершенствует такие способности самостоятельно.

Недостатком символьного подхода является то, что такие открытые возможности воспринимаются не подготовленными людьми как отсутствие инструментов. Эту, скорее культурную проблему, отчасти решает логическое программирование.

Логический подход

Логический подход к созданию систем искусственного интеллекта основан на моделировании рассуждений. Теоретической основой служит логика.

Логический подход может быть проиллюстрирован применением для этих целей языка и системы логического программирования **Пролог**. Программы, записанные на языке Пролог, представляют наборы фактов и правил логического вывода без жесткого задания алгоритма как последовательности действий, приводящих к необходимому результату.

Агентно-ориентированный подход

Последний подход, развиваемый с начала 1990-х годов, называется агентно-ориентированным подходом, или подходом, основанным на использовании интеллектуальных (рациональных) агентов. Согласно этому подходу, интеллект — это вычислительная часть (грубо говоря, планирование) способности достигать поставленных перед интеллектуальной машиной целей. Сама такая машина будет интеллектуальным агентом, воспринимающим окружающий его мир с помощью датчиков, и способной воздействовать на объекты в окружающей среде с помощью исполнительных механизмов.

Этот подход акцентирует внимание на тех методах и алгоритмах, которые помогут интеллектуальному агенту выживать в окружающей среде при выполнении его задачи. Так, здесь значительно тщательнее изучаются алгоритмы поиска пути и принятия решений.

Точно так же, как объектно-ориентированное программирование сдвинуло парадигму с написания процедур к созданию объектов, рациональное про-

граммирование сдвинуло парадигму с создания информационных объектов к созданию мотивированных агентов.

Определение парадигмы, данное автором

Агентом является всё, что может рассматриваться как воспринимающее свою среду с помощью **датчиков** и воздействующее на эту **среду** с помощью **исполнительных механизмов**.

Понятие агента, в отличие от простого объекта, наделяется рядом ментальных конструкций, таких как вера, обязанности и способности. Поэтому в языке программирования будут появляться различные ментальные категории, а семантика программирования будет связана с семантикой ментальных конструкций.

Гибридный подход

Гибридный подход предполагает, что только синергичная комбинация нейронных и символьных моделей достигает полного спектра когнитивных и вычислительных возможностей. Например, экспертные правила умозаключений могут генерироваться нейронными сетями, а порождающие правила получают с помощью статистического обучения. Сторонники данного подхода считают, что гибридные информационные системы будут значительно более сильными, чем сумма различных концепций по отдельности.

Под гибридной интеллектуальной системой (ГиИС) принято понимать систему, в которой для решения задачи используется более одного метода имитации интеллектуальной деятельности человека. Таким образом ГиИС — это совокупность:

- *аналитических моделей*
- *экспертных систем*
- *искусственных нейронных сетей*
- *нечётких систем*
- *генетических алгоритмов*
- *имитационных статистических моделей*

Междисциплинарное направление «гибридные интеллектуальные системы» объединяет ученых и специалистов, исследующих применимость не одного, а нескольких методов, как правило, из различных классов, к решению задач управления и проектирования.

Современный искусственный интеллект

Можно выделить два направления развития ИИ:

- решение проблем, связанных с приближением специализированных систем ИИ к возможностям человека, и их интеграции, которая реализована природой человека (см. Усиление интеллекта);
- создание искусственного разума, представляющего интеграцию уже созданных систем ИИ в единую систему, способную решать проблемы человечества (см. Сильный и слабый искусственный интеллект).

Усиление интеллекта

Усиление интеллекта (УИ) (Intelligence amplification, Cognitive augmentation, Machine augmented intelligence) — совокупность средств и методов, обеспечивающих максимально возможную производительность интеллекта человека; эффективное использование информационных технологий для усиления человеческого интеллекта. Теория УИ активно разрабатывалась в 1950-е и 1960-е годы пионерами кибернетики и информатики.

УИ иногда противопоставляется ИИ, то есть, проекту построения человекоподобного интеллекта в форме автономной технической системы, такой как компьютер или робот. ИИ столкнулся со многими фундаментальными препятствиями, практическими и теоретическими, с которыми вряд ли столкнется УИ, так как УИ требует технологию просто как дополнительную поддержку для автономного интеллекта, который уже существует.

Слабый и сильный ИИ

Сильный и слабый искусственные интеллекты — гипотеза в философии искусственного интеллекта, согласно которой некоторые формы искусственного интеллекта могут действительно обосновывать и решать проблемы.

- **теория сильного искусственного интеллекта** предполагает, что компьютеры могут приобрести способность мыслить и осознавать себя как отдельную личность (в частности, понимать собственные мысли), хотя и не обязательно, что их мыслительный процесс будет подобен человеческому.
- **теория слабого искусственного интеллекта** отвергает такую возможность. Сторонники слабого ИИ предпочитают рассматривать программы лишь как инструмент, позволяющий решать те или иные задачи, которые не требуют полного спектра человеческих познавательных способностей.

Для создания сильного искусственного интеллекта необходимо, чтобы система не работала по какому-то заданному алгоритму, а проявляла признаки интеллектуальности и осознанности, свойственные человеку.

При общении с интеллектуальным устройством посредством анонимного канала человек не должен понять, что он общается с машиной. На сегодняшний день процесс создания устройств сильного искусственного интеллекта находится лишь в стадии становления. Это подтверждает неудачный пример чат-бота Тэй от компании Майкрософт. Уже через день общения пользователями сети Твиттер Тэй стал агрессивен, восхваляя Гитлера и проявляя признаки расизма. Этот пример ярко иллюстрирует то, что созданному боту не хватает элемента осознанности человека. По мнению ряда учёных, наиболее перспективным направлением создания сильного искусственного интеллекта, является эволюционный подход. Если говорить о развитии слабого искусственного интеллекта, то данное направление уже охватило целый спектр научных областей, начиная от задач общего характера, таких как обучение и восприятие, и заканчивая узкоспециализированными сферами, связанными с игрой в шахматы, доказательством теорем, написании литературных произведений, управлении автомобилем и диагностикой заболеваний.

Формы ИИ

- **Узконаправленный (слабый) искусственный интеллект (УИИ).** В последние годы интерес к системам искусственного интеллекта значительно увеличился. При этом уровень развития современных технологий позволяет создавать системы, лишь добавляющие интеллектуальности в нашу жизнь (система автопилотирования, робот — пылесос, стиральная машина с нечеткой логикой и др.), а не воспроизводящие интеллект человека полностью. Разработки такого класса входят в группу «слабого» искусственного интеллекта. УИИ специализируется в одной области. Среди таких ИИ есть те, кто может обыграть чемпиона мира по шахматам, но на этом все. Есть такой, который может предложить лучший способ хранения данных на жестком диске, и все.
- **Общий (сильный) искусственный интеллект.** Иногда также называют ИИ человеческого уровня. ОИИ относят к компьютеру, который умен, как человек — машина, которая способна выполнять любое интеллектуальное действие, присущее человеку. Создать ОИИ намного сложнее, чем УИИ, и мы пока до этого не дошли. Профессор Линда Готтфредсон описывает интеллект как «в общем смысле психический потенциал, который, наряду с другими вещами, включает способность рассуждать, планировать, решать проблемы, мыслить абстрактно, понимать сложные идеи, быстро учиться и извлекать опыт». ОИИ должен уметь делать все это так же просто, как делаете вы.

- **Искусственный сверхинтеллект (ИСИ).** Оксфордский философ и теоретик ИИ Ник Бостром определяет сверхинтеллект как «интеллект, который гораздо умнее лучших человеческих умов в практически любой сфере, включая научное творчество, общую мудрость и социальные навыки». Искусственный сверхинтеллект включает в себя как компьютер, который немного умнее человека, так и тот, который в триллионы раз умнее в любом направлении. ИСИ и есть причина того, что растет интерес к ИИ, а также того, что в таких обсуждениях часто появляются слова «вымирание» и «бессмертие».

Известные ИИ-системы

В настоящий момент в области искусственного интеллекта наблюдается вовлечение многих предметных областей, имеющих скорее практическое отношение к ИИ, а не фундаментальное. Многие подходы были опробованы, но к возникновению искусственного разума ни одна исследовательская группа пока так и не подошла. Ниже представлены лишь некоторые наиболее известные разработки в области ИИ.

Deep Blue — разработанный IBM шахматный суперкомпьютер, победил чемпиона мира по шахматам.

AlphaGo — разработанный Google DeepMind, выиграл матч в Го у корейским профессионала 9 дана Ли Седоля.

Watson — перспективная разработка IBM, способная воспринимать человеческую речь и производить вероятностный поиск, с применением большого количества алгоритмов. Для демонстрации работы Watson принял участие в американской игре «Jeopardy!», аналога «Своей игры» в России, где системе удалось выиграть в обеих играх.

MYCIN — одна из ранних экспертных систем, которая могла диагностировать небольшой набор заболеваний, причем часто так же точно, как и доктора.

20Q — проект, основанный на идеях ИИ, по мотивам классической игры «20 вопросов». Стал очень популярен после появления в Интернете на сайте 20q.net.

Распознавание речи. Такие системы как *ViaVoice* способны обслуживать потребителей.

Лекция 3. Самые популярные сферы развития ИИ

1. Моделирование рассуждений
2. Машинное обучение
3. Обработка естественного языка

4. Представление и использование знаний
5. Интеллектуальная робототехника
6. Биологическое моделирование искусственного интеллекта
7. Другие области исследований

1. Моделирование рассуждений

Анализируя историю ИИ, можно выделить такое обширное направление как моделирование рассуждений. Долгие годы развитие этой науки двигалось именно по этому пути, и теперь это одна из самых развитых областей в современном ИИ. Моделирование рассуждений подразумевает создание **символьных систем**, на входе которых поставлена некая задача, а на выходе требуется её решение. Как правило, предлагаемая задача уже формализована, то есть переведена в математическую форму, но либо не имеет алгоритма решения, либо он слишком сложен, трудоёмок и т. п. В это направление входят: *доказательство теорем, принятие решений и теория игр, планирование и диспетчеризация, прогнозирование.*

Рассуждение - один из важнейших видов мыслительной деятельности человека, в результате которого он формулирует на основе некоторых предложений, высказываний, суждений новые предложения, высказывания, суждения. Действительный механизм рассуждений человека остается пока недостаточно исследованным. Человеческим рассуждениям присущи: неформальность, нечеткость, нелогичность, широкое использование образов, эмоций и чувств, что делает чрезвычайно трудными их исследование и моделирование. К настоящему времени лучше всего изучены логические рассуждения и разработано много механизмов дедуктивных выводов, реализованных в различных интеллектуальных системах, основанных на представлении знаний с помощью логики предикатов 1-го порядка.

Автоматическое доказательство

Автоматическое доказательство (Automated Theorem Proving, ATP, а также Automated deduction) — доказательство, реализованное программно. В основе лежит аппарат математической логики. Используются идеи теории искусственного интеллекта. Процесс доказательства основывается на логике высказываний и логике предикатов.

В силу неразрешимости даже достаточно простых теорий практическое применение имеет лишь полуавтоматическое человеко-машинное доказательство. К тому же после полной автоматизации доказательство называют уже вычислением. Полностью автоматической может быть лишь проверка доказательства теорий посложнее (если его для этого подготовить).

В настоящее время автоматическое доказательство теорем в промышленности применяется в основном при разработке и верификации интегральных схем и программного обеспечения. После того, как была обнаружена ошибка деления в процессорах Пентиум, сложные модули операций с плавающей запятой современных микропроцессоров разрабатываются с особой тщательностью. В новых процессорах AMD, Intel и других фирм автоматическое доказательство теорем используется для проверки того, что деление и другие операции выполняются корректно.

Корпорация Microsoft использует автоматический доказатель теорем Z3 для верификации кода операционной системы Windows 7 и других программных продуктов

Теория принятия решений

Теория принятия решений — область исследования, вовлекающая понятия и методы математики, статистики, экономики, менеджмента и психологии с целью изучения закономерностей выбора людьми путей решения проблем и задач, а также способов достижения желаемого результата.

Различают **нормативную теорию**, которая описывает рациональный процесс принятия решения и **дескриптивную теорию**, описывающую практику принятия решений.

Процесс решения проблем и задач

Рациональный процесс решения проблем и задач включает следующие этапы, при необходимости, выполняемые одновременно, параллельно, итерационно, с возвратом к исполнению предыдущих этапов:

- Ситуационный анализ (анализ проблемной ситуации);
- Идентификация проблемы и постановка цели;
- Поиск необходимой информации;
- Формирование множества возможных решений;
- Формирование критериев оценки решений;
- Разработка индикаторов и критериев для мониторинга реализации решений;
- Проведение оценки решений;
- Выбор наилучшего решения;
- Планирование;
- Реализация;
- Мониторинг реализации;
- Оценка результата.

При этом выполнение всего процесса и этапов осуществляется рационально обоснованным способом.

Теория игр

Теория игр — математический метод изучения оптимальных стратегий в играх. Под игрой понимается процесс, в котором участвуют две и более сторон, ведущие борьбу за реализацию своих интересов. Каждая из сторон имеет свою цель и использует некоторую стратегию, которая может вести к выигрышу или проигрышу — в зависимости от поведения других игроков. Теория игр помогает выбрать лучшие стратегии с учётом представлений о других участниках, их ресурсах и их возможных поступках.

Теория игр — раздел прикладной математики, точнее исследования операций. Чаще всего методы теории игр находят применение в международных отношениях, экономике, чуть реже в других общественных науках — социологии, политологии, психологии, этике, юриспруденции и других. Начиная с 1970-х годов, её взяли на вооружение биологи для исследования поведения животных и теории эволюции. Очень важное значение она имеет для искусственного интеллекта и кибернетики, особенно с проявлением интереса к интеллектуальным агентам.

Автоматическое планирование и диспетчеризация

Автоматическое планирование и диспетчеризация (*Automated planning and scheduling, APS*) — область задач искусственного интеллекта, касающаяся выполнения стратегии или последовательности действий, обычно для интеллектуальных агентов, автономных роботов и беспилотных аппаратов. В отличие от классических проблем управления и классификации, решения задач данной области комплексны, неизвестны и должны разрабатываться и оптимизироваться в многомерном пространстве.

Решения в основном используют процессы проб и ошибок присущие области искусственного интеллекта, такие как динамическое программирование, обучение с подкреплением и комбинаторная оптимизация.

У типичного планировщика три входа:

- описание начальных условий,
- описание желаемой цели и
- множество возможных действий, заданных формальным языком наподобие STRIPS.

Планировщик создаёт последовательность действий, которые ведут систему из начального состояния в состояние, удовлетворяющее поставленной цели. Альтернативным способом описания проблем планирования является иерар-

хическая сеть задач, в которой из данного множества задач, каждая задача может быть либо выполнена с помощью примитивного действия, либо разбита на аналогичное подмножество задач.

Прогнозирование

Прогноз (от греч. πρόβωσις «предвидение, предсказание») — это научно обоснованное суждение о возможных состояниях объекта в будущем и (или) об альтернативных путях и сроках их осуществления. В узком смысле это вероятностное суждение о будущем состоянии объекта исследования.

Прогнозирование — это разработка прогноза; в узком значении — специальное научное исследование конкретных перспектив дальнейшего развития какого-либо процесса.

Необходимость прогноза обусловлена желанием знать события будущего, что достоверно — невозможно в принципе, исходя из статистических (ошибки текущих оценок), вероятностных (многовариантность следствий), эмпирических (методологические ошибки моделей), философских (ограниченность текущих знаний) принципов.

Точность любого прогноза обусловлена:

объёмом «истинных» (верифицированных с известной погрешностью) исходных данных и периодом их сбора;

объёмом неверифицированных исходных данных и периодом их сбора;

свойствами объекта прогнозирования и системы его взаимодействия с субъектом прогноза;

методиками и моделями прогнозирования.

При возрастании совокупности факторов, влияющих на точность прогноза, он практически замещается рутинным расчётом с некоторой установившейся погрешностью.

К основным методам прогнозирования относят:

- статистические методы;
- экспертные оценки (например, метод Дельфи);
- методы моделирования, в том числе имитационного;
- интуитивные (то есть выполненные без применения технических средств, экспромтом, «в уме» специалистом, имеющим опыт ранее применяемых научных методов в данном типе прогнозов).

2. Машинное обучение (Machine Learning)

Единственным способом заставить компьютер что-то делать — от сложения двух чисел до управления самолетом — было составление некоего алгоритма, скрупулезно объясняющего машине, что именно от нее требуется. Одна-

ко алгоритмы машинного обучения — совсем другое дело: они угадывают все сами, делая выводы на основе данных, и чем больше данных, тем лучше у них получается. Это значит, что компьютеры не надо программировать: они программируют себя сами.

Машинное обучение — технология, которая строит саму себя. Это новое явление в нашем мире.

Машинное обучение — это тренировка математической модели на исторических данных для того, чтобы прогнозировать какое-то событие или явление на новых данных. То есть попытка заставить алгоритмы программ совершать действия на основе предыдущего опыта, а не только на основе имеющихся данных. Для обучения нужны исторические данные (**обучающая выборка**) и значение целевой переменной (то, что прогнозируем), которое соответствует заданным историческим данным. Модель наблюдает и находит зависимости между данными и целевой переменной. Эти зависимости используются моделью для нового набора данных, чтобы прогнозировать целевую переменную, которая неизвестна.

Машинное обучение включает в себя целый набор методов и алгоритмов, которые могут предсказать какой-то результат по входным данным. Например, у вас есть какая-то информация по тому, сколько стоили ценные бумаги в каждый момент из какого-то длинного промежутка времени, алгоритмы машинного обучения могут предсказать, сколько эти бумаги будут стоить в будущем.

Типы машинного обучения

Существует множество моделей для машинного обучения, но они, как правило, относятся к одному из трех типов:

1. *обучение с учителем (supervised learning);*
2. *обучение без учителя, или самообучение (unsupervised learning);*
3. *обучение с подкреплением (reinforcement learning).*

В зависимости от выполняемой задачи одни модели могут быть более подходящими и более эффективными, чем другие.

Обучение с учителем

В этом типе корректный результат при обучении модели явно обозначается для каждого идентифицируемого элемента в наборе данных. Это означает, что при считывании данных у алгоритма уже есть правильный ответ. Поэтому вместо поисков ответа он стремится найти связи, чтобы в дальнейшем, при

введении необозначенных данных, получались правильные классификация или прогноз.

В контексте классификации алгоритм обучения может, например, снабжаться историей транзакций по кредитным картам, каждая из которых помечена как безопасная или подозрительная. Он должен изучить отношения между этими двумя классификациями, чтобы затем суметь соответствующим образом маркировать новые операции в зависимости от параметров классификации (например, место покупки, время между операциями и т. д.).

В случае, когда данные непрерывно связаны друг с другом, как, например, изменение курса акций во времени, регрессионный алгоритм обучения может использоваться для прогнозирования следующего значения в наборе данных.

Машинное обучение с учителем — это прямая имитация закономерностей, имеющих место между двумя наборами данных. В нем всегда входной набор данных преобразуется в выходной. Часто это невероятно мощный и полезный метод. Рассмотрим следующие примеры (**входные** данные выделены жирным шрифтом, а выходные — *курсивом*):

- Использование **пикселей** изображения для определения *присутствия или отсутствия кота*.
- Использование списка **понравившихся фильмов** для выбора *фильмов, которые могут понравиться*.
- Использование **слов** в сообщении, чтобы предсказать, *счастливы ли их автор или расстроен*.
- Использование **данных** с метеорологических приборов для *предсказания вероятности дождя*.
- Использование **датчиков** автомобильного двигателя для *определения оптимальных настроек*.
- Использование **новостей** для предсказания *завтрашних котировок на бирже*.
- Использование входного **числа** для предсказания *удвоенного числа*.
- Использование **аудиофайла** для получения *транскрипции речи, содержащейся в нем*.

Все это — задачи машинного обучения с учителем. Во всех случаях алгоритм машинного обучения пытается выявить такие закономерности между двумя наборами данных, чтобы по одному можно было спрогнозировать другой. Его удобно использовать, когда на входе имеется нечто известное и требуется быстро преобразовать его в то, что хотелось бы знать.

Обучение без учителя

В этом случае у алгоритма в процессе обучения нет заранее установленных ответов. Его цель — найти смысловые связи между отдельными данными, выявить шаблоны и закономерности. Например, **кластеризация** — это использование неконтролируемого обучения в рекомендательных системах (например, люди, которым понравился этот смартфон, также положительно оценили вот этот).

Обучение с подкреплением

Этот тип обучения представляет собой смесь первых двух. Обычно он используется для решения более сложных задач и требует взаимодействия с окружающей средой. Данные предоставляются средой и позволяют алгоритму реагировать и учиться.

Область применения такого метода обширна: от контроля роботизированных рук и поиска наиболее эффективной комбинации движений, до разработки систем навигации роботов, где поведенческий алгоритм «избежать столкновения» обучается опытным путем, получая обратную связь при столкновении с препятствием.

Логические игры также хорошо подходят для обучения с подкреплением, так как они традиционно содержат логическую цепочку решений. Этот метод обучения также часто применяется в логистике, составлении графиков и тактическом планировании задач.

Задачи машинного обучения

Машинное обучение базируется на идее о том, что аналитические системы могут учиться выявлять закономерности и принимать решения с минимальным участием человека. В машинном обучении выделяют следующие ключевые задачи:

регрессия — предсказание числовых значений признаков, например, предсказание будущих объемов продаж на основании известных данных о продажах в прошлом;

классификация — предсказание того, к какому из известных классов относится объект, например, предсказание того, вернет ли заемщик кредит, на основании данных о том, как возвращали кредиты заемщики в прошлом;

кластеризация — разделение большого множества объектов на кластеры — классы, внутри которых объекты похожи между собой, например, сегментирование рынка, разделение всех потребителей на классы так, что внутри классов потребители похожи между собой, а в разных классах — отличаются;

уменьшение размерности – сведение большого числа признаков к меньшему (обычно 2–3) для удобства их последующей визуализации (например, сжатие данных);

поиск аномалий — поиск редких и необычных объектов, существенно отличающихся от основной массы, например поиск мошеннических транзакций.

Методы машинного обучения

Сегодня чаще всего для создания программ машинного обучения используются языки R, Python, Scala и Julia. Они поддерживаются многими интегрированными средами разработки, в частности, R-Studio, R-Brain, Visual Studio, Eclipse, PyCharm, Spyder, IntelliJ IDEA, Jupyter Notebooks, Juno и др.

Как работает машинное обучение

Машинное обучение часто называют волшебным или черным ящиком:

Вводишь данные → «волшебный черный ящик» → Миссия выполнена

Давайте посмотрим на сам процесс обучения, чтобы лучше понять, как машинное обучение справляется с данными.

Модель машинного обучения

Сбор

Машинное обучение основывается на данных. Первый шаг — убедиться, что имеющиеся данные верны и относятся именно к той задаче, которую вы пытаетесь решить. Оцените свои возможности для сбора данных, обдумайте их источник, необходимый формат и т. д.

Очистка

Данные зачастую формируются из различных источников, отображаются в различных форматах и языках. Соответственно, среди них могут оказаться нерелевантные или ненужные значения, которые потребуется удалить. И наоборот, каких-то данных может не хватать, и потребуется их добавить. От правильной подготовки базы данных прямым образом зависит и пригодность к использованию, и достоверность результатов.

Разделение

В зависимости от размера набора данных в некоторых случаях может потребоваться только небольшая их часть. Обычно это называется выборкой. Из

выбранной части данные надо разделить на две группы: одна для использования алгоритмом, а другая для оценки его действий.

Обучение

Этот этап фактически направлен на поиск математической функции, которая точно выполнит указанную задачу. Обучение разнится в зависимости от типа используемой модели. Построение линий в простой линейной модели — это обучение; генерация дерева принятия решений для алгоритма случайного леса — это также обучение. Изменение ответов при построении дерева решений поможет скорректировать алгоритм.

Чтобы было проще, сосредоточимся на нейронных сетях.

Суть в том, что алгоритм использует часть данных, обрабатывает их, замеряет эффективность обработки и автоматически регулирует свои параметры (также называемый метод обратного распространения ошибки) до тех пор, пока не сможет последовательно производить желаемый результат с достаточной достоверностью.

Оценка

После того как алгоритм хорошо показал себя на учебных данных, его эффективность оценивается на данных, с которыми он еще не сталкивался. Дополнительная корректировка производится при необходимости. Этот процесс позволяет предотвратить переобучение — явление, при котором алгоритм хорошо работает только на учебных данных.

Оптимизация

Модель оптимизируется, чтобы при интеграции в приложение весить как можно меньше и как можно быстрее работать.

Для чего можно использовать машинное обучение

В бизнесе можно рассматривать три сферы применения машинного обучения:

1. *описательную*
2. *прогнозирующую*
3. *нормативную.*

Описательное применение относится к записи и анализу статистических данных для расширения возможностей бизнес-аналитики. Руководители получают описание и максимально информативный анализ результатов и последствий прошлых действий и решений. Этот процесс в настоящее время обычен

для большинства крупных компаний по всему миру — например, анализ продаж и рекламных проектов для определения их результатов и рентабельности.

Второе применение машинного обучения — прогнозирование. Сбор данных и их использование для прогнозирования конкретного результата позволяет повысить скорость реакции и быстрее принимать верные решения. Например, прогнозирование оттока клиентов может помочь его предотвратить. Сегодня этот процесс применяется в большинстве крупных компаний.

Третье и наиболее продвинутое применение машинного обучения внедряется уже существующими компаниями и совершенствуется усилиями недавно созданных. Простого прогнозирования результатов или поведения уже недостаточно для эффективного ведения бизнеса. *Понимание причин, мотивов и окружающей ситуации — вот необходимое условие для принятия оптимального решения.* Этот метод наиболее эффективен, если человек и машина объединяют усилия. Машинное обучение используется для поиска значимых зависимостей и прогнозирования результатов, а специалисты по данным интерпретируют результат, чтобы понять, почему такая связь существует. В результате становится возможным принимать более точные и верные решения.

Глубокое обучение (Deep learning)

Глубокое обучение может быть как с учителем, так и без, но оно подразумевает под собой анализ **Big Data** — настолько большого массива информации, что одного компьютера будет недостаточно. Поэтому Deep Learning использует для работы нейронные сети. Нейронные сети позволяют разделить одну большую задачу на несколько маленьких и делегировать их другим устройствам. Например, один процессор собирает информацию и передает ее двум другим. Те, в свою очередь, анализируют ее и передают еще четверем, которые выполняют еще какие-то задачи и передают следующим процессорам.

3. Обработка естественного языка

Немаловажным направлением является обработка естественного языка, в рамках которого проводится анализ возможностей понимания, обработки и генерации текстов на «человеческом» языке. В рамках этого направления ставится цель такой обработки естественного языка, которая была бы в состоянии приобрести знание самостоятельно, читая существующий текст, доступный по Интернету. Некоторые прямые применения обработки естествен-

ного языка включают **информационный поиск** (в том числе, глубокий анализ текста) и **машинный перевод**.

Понимание естественного языка иногда считают **AI-полной** задачей, потому как распознавание живого языка требует огромных знаний системы об окружающем мире и возможности с ним взаимодействовать. Само определение смысла слова «понимать» — одна из главных задач искусственного интеллекта.

Качество понимания зависит от множества факторов: от языка, от национальной культуры, от самого собеседника и т. д. Вот некоторые примеры сложностей, с которыми сталкиваются системы понимания текстов.

Свободный порядок слов может привести к совершенно иному толкованию фразы: «Бытие определяет сознание» — что определяет что?

В русском языке свободный порядок компенсируется развитой морфологией, служебными словами и знаками препинания, но в большинстве случаев для компьютера это представляет дополнительную проблему.

В речи могут встретиться неологизмы, например глагол «Пятидесятирублёруй» — то есть высылай 50 рублей. Система должна уметь отличать такие случаи от опечаток и правильно их понимать.

Сложности с раскрытием анафор (распознаванием, что имеется в виду при использовании местоимений): предложения «Мы отдали бананы обезьянам, потому что **они** были голодные» и «Мы отдали бананы обезьянам, потому что **они** были перезрелые» похожи по синтаксической структуре. В одном из них местоимение **они** относится к обезьянам, а в другом — к бананам. Правильное понимание зависит от знаний компьютера, какими могут быть бананы и обезьяны.

Правильное понимание омонимов — ещё одна проблема. При распознавании речи, помимо прочих, возникает проблема фонетических омонимов. Во фразе «Серый волк в глухом лесу встретил рыжую лису» выделенные слова слышатся одинаково, и без знания, кто глухой, а кто рыжий, не обойтись.

Популярные задачи

- Распознавание речи
- Анализ текста
- Извлечение информации
- Информационный поиск
- Анализ высказываний
- Анализ тональности текста
- Вопросно-ответные системы
- Генерирование текста
- Синтез речи

- Общая классификация
- Категоризация текстов
- Классификация последовательностей символов
 - Распознавание именованных сущностей
 - Определение частей речи слов
- Распознавание фраз
- Извлечение информации из текста
- Синтаксическая аннотация
- Семантическая аннотация
- Генерирование текста
- Генерация текста на основе распознанной речи
- Машинный перевод
- Обобщение текста

Информационный поиск

Информационный поиск (information retrieval) — процесс поиска неструктурированной документальной информации, удовлетворяющей информационные потребности, и наука об этом поиске.

Поиск информации представляет собой процесс выявления в некотором множестве документов (текстов) всех тех, которые посвящены указанной теме (предмету), удовлетворяют заранее определенному условию поиска (запросу) или содержат необходимые (соответствующие информационной потребности) факты, сведения, данные.

Процесс поиска включает последовательность операций, направленных на сбор, обработку и предоставление информации.

В общем случае поиск информации состоит из четырех этапов:

1. определение (уточнение) информационной потребности и формулировка информационного запроса;
2. определение совокупности возможных держателей информационных массивов (источников);
3. извлечение информации из выявленных информационных массивов;
4. ознакомление с полученной информацией и оценка результатов поиска.

Методы поиска

Адресный поиск

Процесс поиска документов по чисто формальным признакам, указанным в запросе.

Семантический поиск

Процесс поиска документов по их содержанию.

Документальный поиск

Процесс поиска в хранилище информационно-поисковой системы первичных документов или в базе данных вторичных документов, соответствующих запросу пользователя.

Фактографический поиск

Процесс поиска фактов, соответствующих информационному запросу.

Машинный перевод

Машинный перевод — процесс перевода текстов (письменных, а в идеале и устных) с одного естественного языка на другой с помощью специальной компьютерной программы. Так же называется направление научных исследований, связанных с построением подобных систем.

Формы организации взаимодействия ЭВМ и человека при машинном переводе:

С постредактированием: исходный текст перерабатывается машиной, а человек-редактор исправляет результат.

С предредактированием: человек приспособливает текст к обработке машиной (устраняет возможные неоднозначные прочтения, упрощает и размечает текст), после чего начинается программная обработка.

С интерредактированием: человек вмешивается в работу системы перевода, разрешая трудные случаи.

Смешанные системы (например, одновременно с пред- и постредактированием).

4. Представление и использование знаний

Направление инженерия знаний объединяет задачи получения знаний из простой информации, их систематизации и использования. Это направление исторически связано с созданием **экспертных систем** — программ, использующих специализированные базы знаний для получения достоверных заключений по какой-либо проблеме.

Производство знаний из данных — одна из базовых проблем **интеллектуального анализа данных**. Существуют различные подходы к решению этой проблемы, в том числе — на основе нейросетевой технологии, использующие процедуры вербализации нейронных сетей.

Модели представления знаний

Существуют десятки моделей (или языков) представления знаний для различных предметных областей. Большинство из них может быть сведено к следующим классам:

- производственные модели;
- семантические сети;
- фреймы;
- формальные логические модели.

Производственные правила

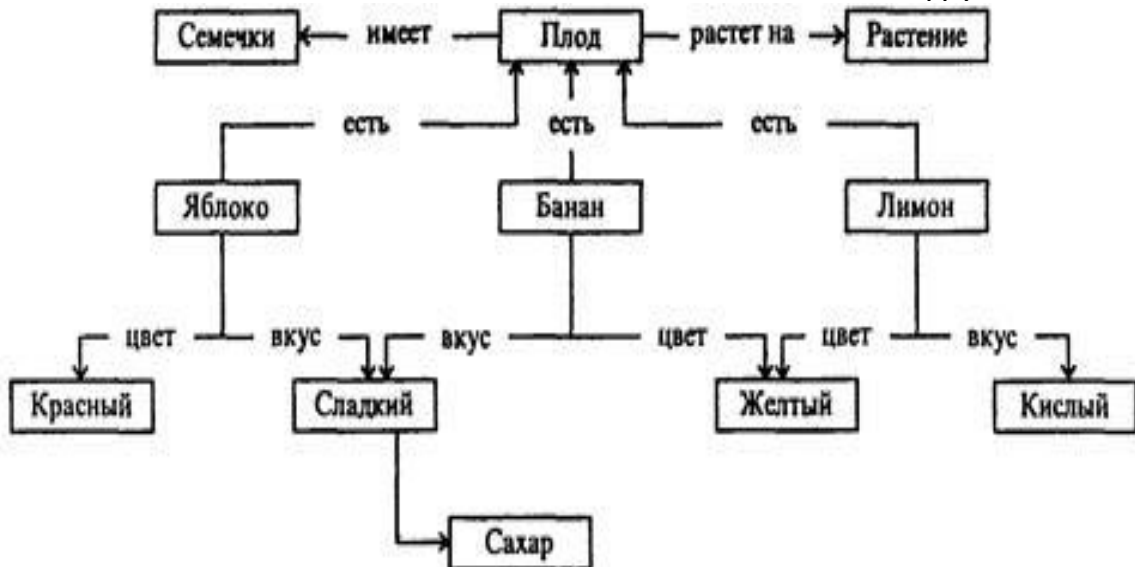
Производственные правила - наиболее простой способ представления знаний. Он основан на представлении знаний в форме правил, структурированных в соответствии с образцом «ЕСЛИ - ТО». Часть правила «ЕСЛИ» называется посылкой, а «ТО» - выводом или действием. Правило в общем виде записывается так:

ЕСЛИ A_1, A_2, \dots, A_n , ТО B .

Такая запись означает, что «если все условия от A_1 до A_n являются истинными, то B также истинно» или «когда все условия от A_1 до A_n выполняются, то следует выполнить действие B ».

Семантическая сеть

Семантическая сеть - иной подход к представлению знаний, который основан на изображении понятий (сущностей) с помощью точек (узлов) и отношений между ними с помощью дуг на плоскости. Семантические сети способны отображать структуру знаний во всей сложности их взаимосвязей, увязать в единое целое объекты и их свойства. В качестве примера может быть приведена часть семантической сети, относящейся к понятию «фрукты»



Фреймовая система

Фреймовая система имеет все свойства, присущие языку представления знаний, и одновременно является новым способом обработки информации.

Слово «фрейм» в переводе с английского языка означает «рамка». Фрейм является единицей представления знаний об объекте, которую можно описать некоторой совокупностью понятий и сущностей. Фрейм имеет определенную внутреннюю структуру, состоящую из множества элементов, называемых **слотами**. Каждый слот, в свою очередь, представляется определенной структурой данных, процедурой, или может быть связан с другим фреймом.

- *Фрейм*: человек
- *Класс*: Животное
- *Структурный элемент*: Голова, шея, руки, ноги,...
- *Рост*: 30–220 см
- *Масса*: 1–200 кг
- *Хвост*: Нет
- *Фрейм аналогии*: Обезьяна

Существуют и другие, менее распространенные подходы к представлению знаний в интеллектуальных системах, в том числе гибридные, на основе уже описанных подходов.

Экспертная система

Экспертная система (ЭС, англ. expert system) — компьютерная система, способная частично заменить специалиста-эксперта в разрешении проблемной ситуации. Современные экспертные системы начали разрабатываться исследователями искусственного интеллекта в 1970-х годах, а в 1980-х годах получили коммерческое подкрепление.

Важнейшей частью экспертной системы являются базы знаний как модели поведения экспертов в определённой области знаний с использованием процедур логического вывода и принятия решений, иными словами базы знаний — совокупность фактов и правил логического вывода в выбранной предметной области деятельности.

В настоящее время «классическая» концепция экспертных систем, сложившаяся в 1970—1980 годах, переживает кризис, по всей видимости связанный с её глубокой ориентацией на общепринятый в те годы текстовый человеко-машинный интерфейс, который в настоящее время в пользовательских приложениях почти полностью вытеснен графическим (GUI). Кроме того, «классический» подход к построению экспертных систем плохо согласуется с реляционной моделью данных, что делает невозможным эффективное использование современных промышленных СУБД для организации баз знаний таких систем.

Интеллектуальный анализ данных

Data mining (рус. *добыча данных, интеллектуальный анализ данных, глубокий анализ данных*) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Термин введён Пятецким-Шапиро в 1989 году.

Английское словосочетание «data mining» пока не имеет устоявшегося перевода на русский язык. При передаче на русском языке используются следующие словосочетания: просев информации, добыча данных, извлечение данных, а также интеллектуальный анализ данных. Более полным и точным является словосочетание «обнаружение знаний в базах данных» (англ. knowledge discovery in databases, KDD).

Основу методов data mining составляют всевозможные методы классификации, моделирования и прогнозирования, основанные на применении деревьев решений, искусственных нейронных сетей, генетических алгоритмов, эволюционного программирования, ассоциативной памяти, нечёткой логики. К методам data mining нередко относят статистические методы (дескриптивный анализ, корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ, компонентный анализ, дискриминантный анализ, анализ временных рядов, анализ выживаемости, анализ связей). Такие методы, однако, предполагают некоторые априорные представления об анализируемых данных, что несколько расходится с целями data mining (обнаружение ранее неизвестных нетривиальных и практически полезных знаний).

Методы data mining лежат на стыке баз данных, статистики и искусственно-го интеллекта.

5. Интеллектуальная робототехника

Роботы — часть стремительно надвигающегося будущего высоких технологий. В настоящее время на планете Земля в сфере робототехники революции происходят чуть ли не каждую неделю. Роботы спасают людей, работают в экстремальных условиях, заменяют живое общение, исследуют планеты Солнечной системы и многое другое. От слуг до наставников роботы развиваются чрезвычайно быстрыми темпами. Работа в сфере робототехники на данный момент борется с самой важной задачей: **как оснастить робота искусственным интеллектом**.

Интеллектуальность требуется роботам, чтобы манипулировать объектами, выполнять навигацию с проблемами локализации (определять местонахож-

дение, изучать ближайшие области) и планировать движение (как добраться до цели)

Основное предназначение роботов – сделать нашу жизнь более комфортной, улучшить условия труда, освободить «руки» от сложных рабочих процессов и увеличить производительность.

Роботы чаще всего встречаются в промышленности, где с их помощью удалось полностью автоматизировать большинство производственных задач. Но, кроме того, умные машины все больше задействуются в военной отрасли, медицине, сфере обслуживания и потребительском секторе.

Примеры интеллектуальной робототехники

1. *Pleo — робот-динозавр.*

Сложный искусственный интеллект дает возможность формировать уникальную индивидуальность каждому отдельно взятому роботу Pleo.

Технические возможности Pleo:

- зрение, основанное на видеокамере (для реакции на освещенность и навигации)
- инфракрасное определение внешних объектов
- два микрофона для реакции на звук
- восемь датчиков прикосновений (голова, подбородок, плечи, хвост, лапы)
- датчики на ступнях (для определения типа поверхности)
- датчик наклона для определения позиции тела
- 14 сервомоторов, по одному в каждом суставе
- инфракрасный датчик для обнаружения попадания объекта в ротовую полость ...

2. *Aibo - серия собак-роботов, разработанная компанией Sony.*

Aibo умеет ходить, «видеть» окружающие его предметы с помощью видеокамеры и инфракрасных датчиков расстояния, распознавать команды и лица. Робот является полностью автономным: он может учиться и развиваться, основываясь на побуждениях своего хозяина, обстановки, или другого AIBO. Несмотря на это, он поддается настройкам с помощью специальных программ. «Настроение» Aibo может меняться в зависимости от окружающей обстановки и влиять на поведение. Инстинкты позволяют Aibo двигаться, играть со своими игрушками, удовлетворять своё любопытство, играть и общаться с хозяином, самостоятельно подзаряжаться и просыпаться после сна. Разработчики утверждают, что у AIBO есть симу-

лирование шести эмоций: счастья, грусти, страха, антипатии, удивления и гнева.

Виды роботов по сфере применения

- Медицинские помощники
- Бытовые ассистенты
- Роботы-игрушки
- Сервисные
- Военные роботы
- Промышленные машины
- Развлекательные

Примеры

- автомобили с автономным управлением
- разгрузочно-погрузочные роботы, упаковщики, сортировщики, формовщики
- радиоуправляемые тракторы и плуги
- беспилотные летательные аппараты
- роботы-пылесосы
- системы умного дома

...

6. Биологическое моделирование искусственного интеллекта

Сюда можно отнести несколько направлений.

- **Нейронные сети** используются для решения нечётких и сложных проблем, таких как распознавание геометрических фигур или кластеризация объектов.
- **Генетический подход** основан на идее, что некий алгоритм может стать более эффективным, если позаимствует лучшие характеристики у других алгоритмов («родителей»).
- Относительно новый подход, где ставится задача создания автономной программы — агента, взаимодействующей с внешней средой, называется **агентным подходом**.

Биокомпьютинг

Биокомпьютинг (или квазибиологическая парадигма — биологическое направление в искусственном интеллекте, сосредоточенное на разработке и использовании компьютеров, которые функционируют как живые организмы или содержат биологические компоненты, так называемые **биокомпьютеры**).

Квазибиологическая парадигма сегодня по своему содержанию и возможным приложениям значительно богаче, чем первоначальный подход Мак-Каллоха и Питса. Она находится в процессе развития и изучения возможностей создания на её основе эффективных средств обработки информации.

Направления в исследованиях

- Биокompьютинг позволяет решать сложные вычислительные задачи, организуя вычисления при помощи живых тканей, клеток, вирусов и биомолекул. Часто используют молекулы дезоксирибонуклеиновой кислоты, на основе которого создают ДНК-компьютер. Кроме ДНК, в качестве биопроцессора могут использоваться также белковые молекулы и биологические мембраны.
- Молекулярные вычисления
- Биомолекулярная электроника
- Искусственные нейронные сети
- Эволюционные вычисления
- Нейрокомпьютинг

7. Другие области исследований

Машинное творчество

Природа человеческого творчества ещё менее изучена, чем природа интеллекта. Тем не менее, эта область существует, и здесь поставлены проблемы написания компьютером музыки, литературных произведений (часто — стихов или сказок), художественное творчество. Создание реалистичных образов широко используется в кино и индустрии игр.

Отдельно выделяется изучение проблем технического творчества систем искусственного интеллекта. Теория решения изобретательских задач, предложенная в 1946 году Г. С. Альтшуллером, положила начало таким исследованиям.

Добавление данной возможности к любой интеллектуальной системе позволяет продемонстрировать, что именно система воспринимает и как это понимает. Добавлением шума вместо недостающей информации или фильтрация шума имеющимися в системе знаниями производит из абстрактных знаний конкретные образы, легко воспринимаемые человеком, особенно это полезно для интуитивных и малоценных знаний, проверка которых в формальном виде требует значительных умственных усилий.

Компьютерное зрение (Computer Vision, CV)

Компьютерное зрение (иначе техническое зрение) — теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов.

Как научная дисциплина, компьютерное зрение относится к теории и технологии создания искусственных систем, которые получают информацию из изображений. Видеоданные могут быть представлены множеством форм, таких как видеопоследовательность, изображения с различных камер или трехмерными данными, или медицинского сканера.

Как технологическая дисциплина, компьютерное зрение стремится применить теории и модели компьютерного зрения к созданию систем компьютерного зрения. Примерами применения таких систем могут быть:

- Системы управления процессами (промышленные роботы, автономные транспортные средства).
- Системы видеонаблюдения.
- Системы организации информации (например, для индексации баз данных изображений).
- Системы моделирования объектов или окружающей среды (анализ медицинских изображений, топографическое моделирование).
- Системы взаимодействия (например, устройства ввода для системы человеко-машинного взаимодействия).
- Системы дополненной реальности.
- Вычислительная фотография, например, для мобильных устройств с камерами.

Одним из наиболее важных применений является обработка изображений в медицине. Эта область характеризуется получением информации из видеоданных для постановки медицинского диагноза пациентам. В большинстве случаев видеоданные получают с помощью микроскопии, рентгенографии, ангиографии, ультразвуковых исследований и томографии. Примером информации, которая может быть получена из таких видеоданных, является обнаружение опухолей, атеросклероза или других злокачественных изменений.

Другой прикладной областью компьютерного зрения является промышленность. Здесь информацию получают для целей поддержки производственного процесса. Примером может служить контроль качества, когда детали или конечный продукт автоматически проверяются на наличие дефектов. Другим примером является измерение положения и ориентации деталей, поднимаемых рукой робота.

Эволюционные алгоритмы

Эволюционные алгоритмы — направление в искусственном интеллекте (раздел эволюционного моделирования), которое использует и моделирует процессы естественного отбора.

Все они моделируют базовые положения в теории биологической эволюции — процессы отбора, мутации и воспроизводства. Поведение агентов определяется окружающей средой. Множество агентов принято называть популяцией. Такая популяция эволюционирует в соответствии с правилами отбора в соответствии с целевой функцией, задаваемой окружающей средой. Каждому агенту (индивидууму) популяции назначается значение его пригодности в окружающей среде. Размножаются только наиболее пригодные виды. Рекомбинация и мутация позволяют изменяться агентам и приспособляться к среде. Такие алгоритмы относятся к адаптивным поисковым механизмам.

Эволюционные алгоритмы успешно использовались для задач типа функциональной оптимизации и могут легко быть описаны на математическом языке.

Виды алгоритмов

генетические алгоритмы — эвристический алгоритм поиска, используемый для решения задач оптимизации и моделирования путём случайного подбора, комбинирования и вариации искомых параметров;

генетическое программирование — автоматическое создание или изменение программ с помощью генетических алгоритмов;

эволюционное программирование — аналогично генетическому программированию, но структура программы постоянна, изменяются только числовые значения;

программирование экспрессии генов

эволюционные стратегии — похожи на генетические алгоритмы, но в следующее поколение передаются только положительные мутации;

дифференциальная эволюция

нейроэволюция — аналогично генетическому программированию, но геномы представляют собой искусственные нейронные сети, в которых происходит эволюция весов при заданной топологии сети, или помимо эволюции весов также производится эволюция топологии;

системы классификаторов;

Генетический алгоритм

Генетический алгоритм — это в первую очередь эволюционный алгоритм, другими словами, основная идея алгоритма — скрещивание (комбинирование). Как несложно догадаться, идея алгоритма взята у природы. Так вот, путем перебора и самое главное отбора получается правильная «комбинация». Алгоритм делится на три этапа:

1. Скрещивание
2. Селекция (отбор)
3. Формирования нового поколения

Если результат нас не устраивает, эти шаги повторяются до тех пор, пока результат нас не начнет удовлетворять или произойдет одно из нижеперечисленных условий:

- Количество поколений (циклов) достигнет заранее выбранного максимума
- Исчерпано время на мутацию

Рекомендательные системы (recommendation-engine)

Используя передовые алгоритмы, такие как машинное обучение и ИИ, система рекомендаций может помочь клиентам получить соответствующие продукты, которые они хотят или нуждаются.

Механизм рекомендаций продуктов - это, по сути, решение, которое позволяет маркетологам предлагать своим клиентам соответствующие рекомендации по продуктам в режиме реального времени. В качестве мощных инструментов фильтрации данных рекомендательные системы используют алгоритмы и методы анализа данных, чтобы рекомендовать наиболее релевантный продукт / элементы конкретному пользователю.

Основной целью любого рекомендательного движка является стимулирование спроса и активное вовлечение пользователей. В первую очередь компонент стратегии персонализации электронной коммерции рекомендательные движки динамически заполняют различные продукты на веб-сайтах, приложениях или электронных письмах, тем самым улучшая качество обслуживания клиентов. Эти виды разнообразных рекомендаций делаются на основе нескольких точек данных, таких как предпочтения клиентов, история прошлых транзакций, атрибуты или ситуационный контекст.

Управление жестами (gesture control)

Жестовый интерфейс — подмножество системы ввода для графического пользовательского интерфейса для устройств, оснащённых специальными либо устройствами ввода (отличными от клавиатуры), либо сенсорными

экранами, и позволяющая эмулировать клавиатурные команды (либо сочетания клавиш) при помощи жестов (росчерков, англ. *gesture*). Основной мотивацией разработки таких интерфейсов является улучшение эргономичности управления, с отказом от привычного для компьютерных программ меню приложения.

Подобный интерфейс может быть реализован как при помощи устройств координатного ввода с возможностью считывания координаты одной точки касания (мышь либо графический планшет — см. «жесты мышью»), так и таких, в которых имеется возможность считывания координат более чем одной точки (т. н. мультикасание, *multitouch*) — сенсорные экраны и панели. Последние стали широко применяться в интерфейсах множества современных смартфонов с сенсорным экраном (напр. iPhone) и ноутбуков (как с тачпадом, так с сенсорным экраном) и прочих мобильных устройств.

Контекстно-зависимые вычисления (*Context-aware computing*)

Контекстно-зависимые системы относят к разряду «повсеместных вычислений» (англ. *Ubiquitous computing* или *Pervasive computing*). Основными источниками информации для контекстно-зависимых систем являются местоположение, социальное и физическое окружение. Несмотря на то, что определение местоположения широко используется в настоящее время, оно не всегда учитывает изменяющиеся интересы пользователя. Контекстная зависимость в широком понимании включает в себя находящиеся рядом людей, устройства, доступ к интернету, уровень освещенности, уровень шума, а также взаимодействие людей в бытовых ситуациях. Например, находитесь ли вы сейчас с семьей или со своим школьным другом.

Интеллектуальное видеонаблюдение (*Video Content Analysis*)

Видеоаналитика — технология, использующая методы компьютерного зрения для автоматизированного получения различных данных на основании анализа последовательности изображений, поступающих с видеокамер в режиме реального времени или из архивных записей. Видеоаналитика представляет собой программное обеспечение (ПО) для работы с видеоконтентом. В основе программного обеспечения лежит комплекс алгоритмов машинного зрения, позволяющих вести видеомониторинг и производить анализ данных без прямого участия человека. Алгоритмы видеоаналитики могут быть интегрированы в различные бизнес-системы, чаще всего используются в видеонаблюдении и других сферах безопасности.

Определения

Видеоаналитика (VCA, Video Content Analysis) – компьютеризированная обработка и автоматический анализ видеоконтента, который поступает на видеосервер от видеокамер, носимых устройств и устройств Интернета вещей IoT, оснащённых веб-камерами.

Видеоаналитика - это технология, использующая методы компьютерного зрения для автоматизированного получения различных данных на основании анализа последовательности изображений, поступающих с видеокамер в режиме реального времени или из архивных записей.

Видеоаналитика представляет собой программное обеспечение (ПО) для работы с видеоконтентом. В основе программного обеспечения лежит комплекс алгоритмов машинного зрения, позволяющих вести видеомониторинг и производить анализ данных без прямого участия человека.

Традиционное решение, включающее в себя функции какой-либо видеоаналитики строится по схеме: камера + back-end аналитика. Т.е. камера просто гонит поток видео на сервер, а специальное ПО на сервере уж делает весь видеоанализ.

Видеоаналитика – это частные приложения компьютерного зрения, которые извлекают информацию и знания из видеоконтента, то есть дают ответы на вопросы:

Кто: распознавание и идентификация людей;

Что: объекты, действия, события, поведение, взаимоотношения;

Где: геолокация, пространственная (3D) и планарная (2D) локация;

Когда: маркировка даты и времени, сезона.

Три основных типа приложений видеоаналитики

1. **Ретроспектива:** что уже случилось, т.е. управление архивами видеозаписей, поиск, сортировка, получение юридических доказательств;
2. **Настоящий момент:** что происходит сейчас, т.е. контроль ситуации, получение предупреждений в реальном времени, кодирование, компрессия видеопотока;
3. **Взгляд в будущее:** что может или скорее всего произойдёт, т.е. предсказания на основе событий прошлого и настоящего, прогнозирование событий или активности, детектирование намечающихся аномалий.

Функциональные возможности



Во время видеонаблюдения в промышленности, городском и жилищном хозяйстве, а также в различных социальных медиа, генерируется огромное количество видеоданных, для которых требуется системы хранения данных (СХД) с высокой ёмкостью. Разрешающая способность видеоизображений всё время возрастает, и количество хранимого контента растёт экспоненциально.

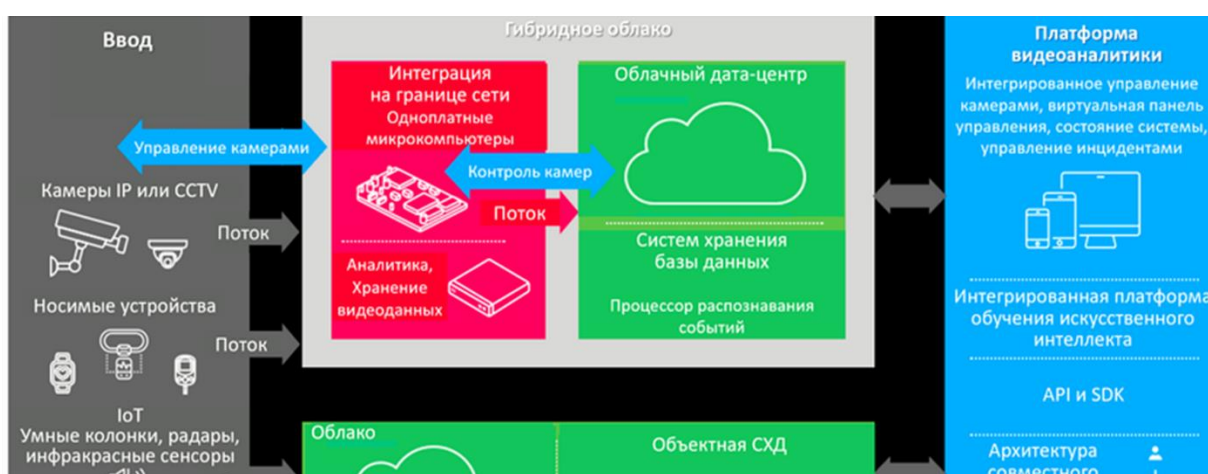
Видеоаналитика в последние годы набирает всё большую популярность по многим причинам. Она позволяет гибко управлять видеопотоками при анализе их контента «на лету», при автоматизации аналитических функций. Это позволяет персоналу концентрироваться на определённых инцидентах на видеозаписи, а не тратить время на просмотр длинных однообразных видеопотоков, что позволяет сократить затраты и численность персонала. Интеллектуальные системы безопасности с видеоаналитикой могут начинать запись, например, только при начале какого-то движения в зоне обзора камеры. При этом снижается нагрузка на сеть и экономится пространство в системе хранения.

При помощи систем видеоаналитики, можно получить ценную информацию о качестве работы персонала предприятия (например, продавцов-

консультантов в торговом зале), таким образом, можно сделать более адекватные оценки его работы.

Системы видеоаналитики не требуют чрезмерно громоздкой инфраструктуры и даже небольшие предприятия, магазины и пр. вполне могут себе позволить её использование. Интенсивность использования функций видеоаналитики можно гибко регулировать по мере потребностей бизнеса, выбирая именно те функции, которые нужны в конкретном случае. Это позволяет создавать кастомизированные решения.

Типовая системная архитектура видеоаналитики (VCA)



Стандарты

2020: В России разработан стандарт ИИ для ситуационной видеоаналитики
20 июля 2020 года стало известно о создании первого в России национально-го стандарта в области искусственного интеллекта для ситуационной видеоаналитики. Документ, подготовленный ООО «Видеоинтеллект» (развивает системы компьютерного зрения для использования в сложных условиях, общественных местах с большим скоплением людей, на объектах промышленности), представил технический комитет по стандартизации ТК 164 «Искусственный интеллект», созданный на базе РВК.

Виртуальный цифровой помощник

Виртуальный цифровой помощник (Virtual Digital Assistant, сокращенно VDA) — веб-сервис и (или) приложение для смартфонов и ПК, фактически исполняющий роль личного секретаря при пользователе. Решает задачи планирования графика, организации и выполнения повседневных дел и контекстного поиска информации для нужд конкретного человека.

Виртуальный цифровой помощник может создавать напоминания, подбирать места отдыха и развлечений, облегчить поиск и онлайн-бронирование

билетов и столиков, заказ такси[2]. Он способен самообучаться в ходе выполнения заданий, анализируя поведение и интересы пользователя.

Эксперты прогнозируют, что к 2025 году около половины работников, занятых в сфере обслуживания, начнут использовать виртуальных помощников при решении ежедневных задач. Отчасти это связано с коронавирусной пандемией, на фоне которой многие компании занялись разработкой систем искусственного интеллекта, включая чат-боты и цифровые ассистенты, чтобы оказывать поддержку онлайн-покупателям круглосуточно и без выходных.

Развитие ИИ способствует все большему использованию виртуальных помощников, поскольку клиенты довольны повышением качества обслуживания. Кроме того, автоматизация рутинных задач, для которых обычно нужны кол-центры, позволяет компаниям сосредоточиться на более сложных задачах.

В будущем виртуальные ассистенты почти ничем не будут отличаться от живых онлайн-консультантов. Уже сейчас появляются чат-боты, все больше похожие на человека и способные анализировать желания пользователей. Огромную роль в этом играет машинное обучение, когда искусственный интеллект тренируется выполнять определенные задачи на базе множества готовых решений.

Помимо экономии времени, которые люди тратят на выбор покупок, получение услуг и планирование ежедневного графика, виртуальные помощники могут усовершенствовать и процесс образования. Например, чат-боты могут круглосуточно консультировать студентов, которые опасаются того, что преподаватель оценит содержание их вопросов. Что более важно, виртуальные помощники смогут настраивать индивидуальную программу обучения, а также помогать учащимся сосредоточиться на тех областях знаний, которые пригодятся им в будущем.

Компании, включая Content Technologies и Carnegie Learning, разрабатывают онлайн-платформы для персонализированного обучения с помощью ИИ. Такие виртуальные помощники способны адаптировать способ взаимодействия к каждому учащемуся. Они смогут определять слабые стороны студента и использовать эту информацию, чтобы предложить примеры решения задач, вызывающих затруднение. Чат-боты и виртуальные помощники могут выполнять роль помощника преподавателя, но справляться с задачами намного быстрее и эффективнее.

Нейроинтерфейсы

Уже сейчас ведутся работы по созданию нейроинтерфейсов, соединяющих мозг с компьютером. Считается, что подобные разработки помогут парализо-

ванным людям взаимодействовать с окружающим миром и даже передвигаться, мысленно управляя экзоскелетом.

В 2019 году нейротехнологическая компания Илона Маска Neuralink представила прототип чипа, который должен вживляться в мозг, чтобы обеспечить связь между центральной нервной системой и компьютером. Чип представляет собой пучок из тысячи электродов, которые будут внедряться в зоны мозга, отвечающие за движение тела и обработку поступающей сенсорной информации. Чип был протестирован на животных, после чего Neuralink получила разрешение начать испытания с участием людей.

Микроэлектроды могут применяться и для нейростимуляции — технологии, позволяющей улучшить работу мозга с помощью электромагнитных импульсов, посылаемых в кору больших полушарий. Нейростимуляция может изменить к лучшему жизнь парализованных людей, пациентов с поражением органов чувств и страдающих хроническими болями, например, при невропатии.

Ожидается, что в будущем нейростимуляция и нейропротезирование будут прочно связаны с ИИ, который научится распознавать и дешифровать нервные сигналы. Это позволит человеку значительно расширить свои способности, поскольку он фактически обзаведется мозговой надстройкой — кибернетической.

Сама возможность того, что ИИ, превосходящий человека, станет дополнением биологического мозга, позволяет представить будущее, где грань между людьми и компьютерами станет более размытой. Ученые уверены: бояться такого будущего не стоит. Эти перспективы окажут значительное влияние на развитие нейробиологии и медицины, что в итоге лишь улучшит качество жизни.

Лекция 4. Области применения искусственного интеллекта

В зависимости от области и обширности сферы применения, выделяют два вида ИИ – Weak AI, называемый еще «слабым», и Strong AI, «сильный». В первом случае перед системой ставят узкоспециализированные задачи – диагностика в медицине, управление роботами, работа на базе электронных торговых платформ. Во втором же подразумевается решение глобальных задач.

Перечислить разом все области, в которых задействован искусственный интеллект, практически нереально. На данный момент он затрагивает все больше самых разных сфер. И причин на то немало – та же автоматизация

производственных процессов, стремительный рост информационного оборота и инвестиций в эту сферу, даже социальное давление.

Коммерция

Одна из наиболее популярных сфер применения ИИ – это Big Data в коммерции. Крупные торговые площадки используют подобные технологии для исследования потребительского поведения. Компания «Яндекс» вообще создает с их помощью музыку. В некоторые мобильные приложения встроены голосовые помощники вроде Siri, Алисы или Cortana. Они упрощают процесс навигации и совершения покупок в сервисе. И не стоит забывать про программы с нейросетями, обрабатывающими фото и видео.

Финансы

Финансовые учреждения давно используют нейронные сети для выявления подозрительных событий и действий. Использование ИИ в банковской сфере началось ещё в 1987 году, когда Security Pacific National Bank в США создал целевую группу по противодействию мошенничеству и несанкционированному использованию дебетовых карт.

Алгоритмическая торговля

Алгоритмическая торговля предполагает использование сложных систем искусственного интеллекта для принятия торговых решений со скоростью, превышающую скорость, на которую способен человеческий организм. Это позволяет делать миллионы сделок в день без какого-либо вмешательства человека. Автоматизированные торговые системы обычно используются крупными институциональными инвесторами.

Исследования рынка и интеллектуальный анализ данных

Несколько крупных финансовых учреждений вложили средства в развитие ИИ, чтобы использовать его в их инвестиционной практике. Широкий спектр функциональных возможностей таких систем включает обработку естественного языка для чтения текста, такого как новости, отчёты брокеров и каналы социальных сетей. Затем система оценивает настроения в упомянутых компаниях и присваивает им оценку.

Банки используют систему ИИ под названием Sqreem, которая может обрабатывать данные для разработки профилей потребителей и сопоставлять их с продуктами, которые они, скорее всего, захотят.

Управление финансовым портфелем

Автоматизированные помощники-советчики становятся все более широко используемыми в отрасли управления инвестициями. Автоматизированные системы предоставляют финансовые консультации и советы в управлении финансовым портфелем с минимальным вмешательством человека. Этот класс финансовых консультантов работает на основе алгоритмов, созданных для автоматического развития финансового портфеля в соответствии с инвестиционными целями и склонностью к риску клиентов. Он может корректировать изменения в реальном времени на рынке и калибровать портфель в соответствии с пожеланиями клиента.

Управление личными финансами

Существуют продукты, которые используют ИИ для помощи людям в управлении их личными финансами. Например, Digit — это приложение, основанное на искусственном интеллекте, которое автоматически помогает потребителям оптимизировать свои расходы и сбережения, основываясь на своих личных привычках и целях. Приложение может анализировать такие факторы, как ежемесячный доход, текущий баланс и привычки к расходам, затем принимать собственные решения и переводить деньги на отдельный сберегательный счёт. Wallet.AI, развивающийся в Сан-Франциско стартап, создаёт агентов, которые анализируют данные, которые генерирует потребитель, при взаимодействии со смартфонами и социальными сетями, чтобы проинформировать потребителя о своих расходах.

Военное дело

Применение ИИ является важным трендом в создании перспективных систем управления полем боя и вооружением.

С помощью ИИ возможно обеспечить оптимальный и адаптивный к угрозам выбор комбинации сенсоров и средств поражения, скоординировать их совместное функционирование, обнаруживать и идентифицировать угрозы, оценивать намерения противника. Существенную роль ИИ играет в реализации тактических систем дополненной реальности. Например, ИИ позволяет обеспечить классификацию и семантическую сегментацию изображений, локализацию и идентификацию мобильных объектов для эффективного целеуказания

Тяжелая промышленность

Роботы стали распространены во многих отраслях промышленности и часто занимаются работой, которая считается опасной для людей. Роботы оказа-

лись эффективными на рабочих местах, связанных с повторяющимися рутинными заданиями, которые могут привести к ошибкам или несчастным случаям из-за снижения концентрации с течением времени. Также широкое применение роботы получили в работе, которую люди могут найти унизительной.

Медицина

- Медицинская диагностика
- Компьютерная интерпретация медицинских изображений. Такие системы помогают сканировать цифровые изображения, например от компьютерной томографии, для типичных проявлений и для выделения заметных отклонений, таких как возможные заболевания. Типичным применением является обнаружение опухолей.
- Анализ сердечного ритма
- Проект Watson — это ещё одно использование ИИ в этой области, программа вопросов/ответов, которая создана для помощи врачам-онкологам
- Роботы-помощники для ухода за престарелыми
- Обработка медицинских записей для предоставления более полезной информации
- Создание планов лечения
- Выявление повышенного риска заболеваний
- Помощь в повторяющихся заданиях, включая управление приёмом медикаментов
- Предоставление консультаций
- Создание лекарств
- Использование человекоподобных манекенов вместо пациентов для клинического обучения

В настоящее время в отрасли здравоохранения работает более 90 стартапов, основанных на применении ИИ.

Иные области применения

- Управление человеческими ресурсами и рекрутинг
- Музыка
- Новости, издательство и писательство
- Новости, издательство и писательство
- Онлайн-овые и телефонные службы поддержки клиентов

- Техническое обслуживание телекоммуникаций
- Развлечение и игры
- Транспорт
- Спецслужбы
- Политика

Примеры использования ИИ

Автомобили битком набиты системами УИИ, от компьютеров, которые определяют, когда должна заработать антиблокировочная тормозная система, до компьютера, который определяет параметры системы впрыска топлива. Самоуправляемые автомобили Google, которые в настоящее время проходят испытания, будут содержать надежные системы УИИ, которые будут воспринимать и реагировать на мир вокруг себя.

Ваш телефон — маленькая фабрика УИИ. Когда вы используете приложение карт, получаете рекомендации по скачиванию приложений или музыки, проверяете погоду на завтра, говорите с Siri или делаете что-либо еще — вы используете УИИ.

Спам-фильтр вашей электронной почты — классический тип УИИ. Он начинает с выяснения того, как отделить спам от пригодных писем, а затем обучается в процессе обработки ваших писем и предпочтений.

Google Translate — еще одна классическая система УИИ, впечатляюще хороша в определенных вещах. Распознавание голоса — тоже. Когда ваш самолет садится, терминал для него определяет не человек. Цену билета — тоже. Лучшие в мире шашки, шахматы, нарды, балды и прочие игры сегодня представлены узконаправленными искусственными интеллектами.

Поиск Google — это один гигантский УИИ, который использует невероятно хитроумные методы для ранжирования страниц и определения результатов поисковой выдачи.

Основные проблемы ИИ

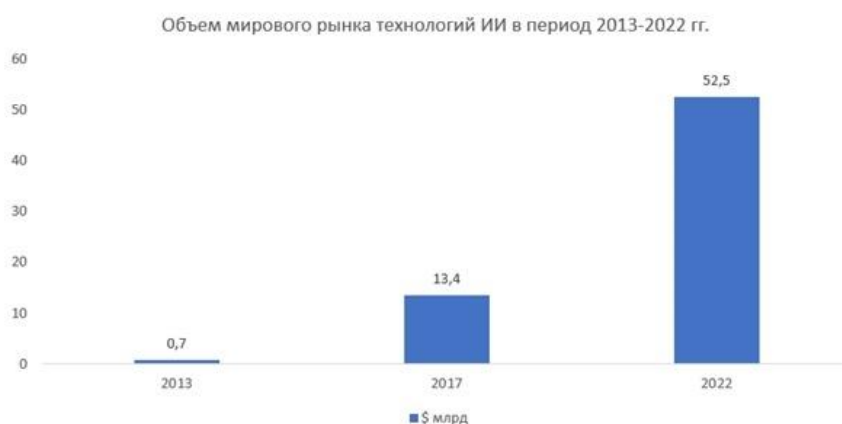
Обучение машин возможно только на основе массива данных. Это означает, что любые неточности в информации сильно сказываются на конечном результате.

Интеллектуальные системы ограничены конкретным видом деятельности. То есть умная система, настроенная на выявление мошенничества в сфере налогообложения, не сможет выявлять махинации в банковской сфере. Мы имеем дело с узкоспециализированными программами, которым ещё далеко до многозадачности человека.

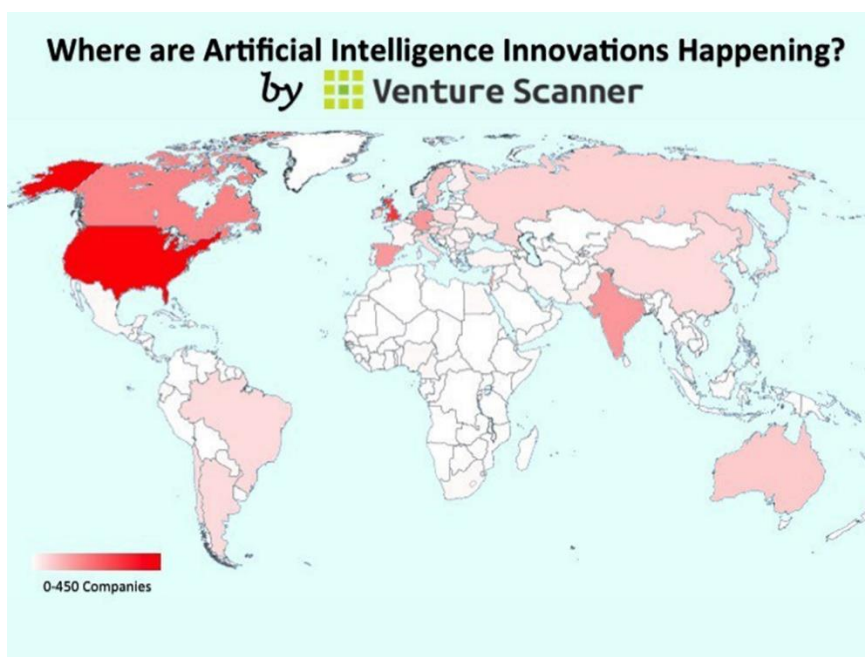
Интеллектуальные машины не являются автономными. Для обеспечения их «жизнедеятельности» необходима целая команда специалистов, а также большие ресурсы.

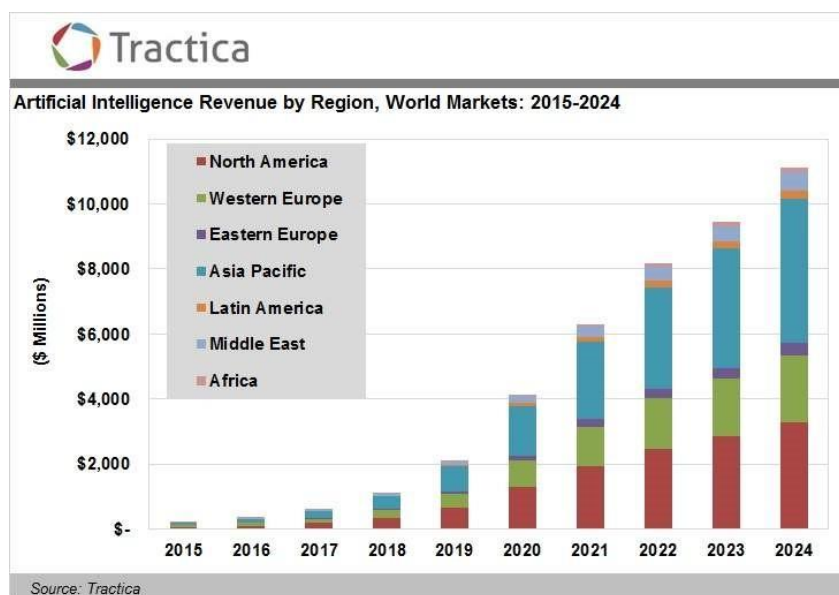
Обзор рынка

По оценкам аналитиков международной консалтинговой компании Frost & Sullivan, к 2022 году суммарный объем рынка технологий ИИ увеличится до \$52,5 млрд, или в 4 раза по сравнению с уровнем 2017 года (\$13,4 млрд). Ежегодный темп роста (CAGR) в прогнозируемый период будет сохраняться на уровне 31%. Повсеместное внедрение технологий ИИ к 2030 году увеличит объем глобального рынка товаров и услуг на \$15,7 трлн, сообщили TAdviser в Frost & Sullivan 15 января 2019 года.



Уровень развития ИИ в разных странах





Перспективы развития искусственного интеллекта

Современные компьютеры приобретают все больше знаний и «умений». Скептики же утверждают, что все возможности ИИ – не более чем компьютерная программа, а не пример самообучения. Однако это не мешает технологии широко распространяться в самых различных сферах и открывать невиданные ранее потенциалы для развития. Со временем компьютеры будут становиться все мощнее, а ИИ еще быстрее совершенствоваться в своем развитии.

А что с рынком труда? Прогнозируется, что роботы смогут заменить людей, работающих в банках, в магазинах, под прицел так же попадают: юристы, курьеры, таксисты, аналитики, журналисты... Все профессии, требующие выполнения монотонных действий, должны исчезнуть. Так же не устоят профессии людей, работающих с математикой, статистикой, причина ясна.

Однако, Элон Маск не считает, что людям негде будет работать. По его словам, со смертью одних профессий придёт много других. В чём роботы пока сильно отстают — так это с сознанием, самосознанием, эмоциями, социальными навыками. И неизвестно, появится ли у них сознание. Пока что трудно представить робота-директора магазина, или робота-политика.

Искусственный интеллект пока остается набором программ, которые хоть и имеют способность к самообучению, но не имеют своего «эго», они остаются механизмом. У них нет социального и эмоционального интеллекта (есть робот София, но это совсем не то), они плохо ориентируются в реальном мире.

Сможет ли механизм догнать создателя? Или AI станет чем-то вроде нового универсального инструмента? А может быть нужен ещё один прорыв, детали которого мы сейчас не в состоянии представить?

ПРАКТИЧЕСКИЕ (СЕМИНАРСКИЕ) ЗАНЯТИЯ

Номер	Наименование практического (семинарского) занятия
1	<p>Тест Тьюринга и интуитивный подход</p> <p><i>Методические рекомендации</i></p> <p>Кроме обсуждения исторического аспекта создания текста и особенностях его прохождения, необходимо рассмотреть вопрос о возможностях, которыми должна обладать система, чтобы пройти этот тест.</p> <p>Провести имитационная игру согласно описанию Тьюринга в статье «Вычислительные машины и разум». Игрок С путём задавания серии вопросов пытается определить, кто из двух других игроков — мужчина, а кто — женщина. Игрок А, мужчина, пытается запутать игрока С, а игрок В пытается помочь С.</p> <p>Вопросы:</p> <ol style="list-style-type: none">1. Согласны ли вы, что тест Тьюринга позволяет дать ответ на вопрос «является ли искусственная система разумной или нет?»2. Какие вопросы задали бы вы в этом тесте?3. Оказал ли тест Тьюринга влияние на процесс разработки искусственных интеллектуальных систем?4. Какие недостатки видите вы в этом тесте?5. Какой тест предложила бы вам интуиция сегодня? <p>Источники:</p> <ul style="list-style-type: none">• Интуитивный подход и тест Тьюринга Другая фаза Яндекс Дзен (yandex.ru)• https://ru.wikipedia.org/wiki/Тест_Тьюринга
2	<p>Биологическое моделирование искусственного интеллекта</p> <p><i>Методические рекомендации</i></p> <p>Сделать упор на интересном направлении биокомпьютинге — биологическом направлении в искусственном интеллекте, сосредоточенным на разработке и использовании компьютеров, которые функционируют как живые организмы или содержат биологические компоненты, так называемые биокомпьютеры.</p> <p>Вопросы:</p> <ol style="list-style-type: none">1. Что лежит в основе квазибиологической парадигмы?2. Приведите примеры использования нейронных сетей.3. Приведите примеры использования генетического подхода.4. Что можно отнести к понятию интеллектуальный агент? Компьютерные вирусы, боты, поисковые роботы.5. Что понимают под моделированием сознания?6. Можем ли мы сегодня сказать, что знаем как работает мозг?7. Каковы основные преимущества нейрокомпьютеров?

	<p>Источники:</p> <ul style="list-style-type: none"> • https://ru.wikipedia.org/wiki/Биокомпьютинг
3	<p>Представление и использование знаний</p> <p><i>Методические рекомендации</i></p> <p>Составить для одной предметной области разные модели знаний, представленные продукционными правилами, семантическими сетями, фреймами. Провести анализ полученных моделей.</p> <p>Рассмотреть подробнее технологию семантической паутины, в которой основанные на XML языки представления знаний используются для увеличения доступности компьютерным системам информации, хранящейся в сети.</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. Какой экспертной системой вы хотели бы воспользоваться? Кого вы видите в качестве эксперта? 2. Как происходит вывод в экспертной системе? 3. Какие методы ИИ положены в основу data mining? 4. Где находят применение семантические сети? 5. Какие знания считаются релевантными? 6. Какие языки программирования ориентированы на представление знаний? 7. Приведите пример логической модели представления знаний. <p>Источники:</p> <ul style="list-style-type: none"> • https://neerc.ifmo.ru/wiki/index.php?title=Представление_знаний • https://ru.wikipedia.org/wiki/Инженерия_знаний • https://ru.wikipedia.org/wiki/Представление_знаний
4	<p>Работа с естественными языками</p> <p><i>Методические рекомендации</i></p> <p>Активно использовать различные программные средства для демонстрации наглядных успехов в этой области ИИ.</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. Как вы лично оцениваете успехи ИИ в этой области? Часто ли приходится пользоваться? Какое из направлений вам более знакомо? 2. Как используется NLP в чат-ботах? 3. Почему речь компьютера немного отличается от человеческой? 4. Как вы считаете, методы анализа текстов сильно зависят от языка? 5. Как используются регулярные выражения для обработки текста? <p>Источники:</p> <ul style="list-style-type: none"> • https://ru.wikipedia.org/wiki/Обработка_естественного_языка • https://habr.com/ru/company/Voximplant/blog/446738/
5	<p>Символьное моделирование мыслительных процессов</p>

	<p>Методические рекомендации</p> <p>Сформировать общее представление по всем направлениям моделирования рассуждений - доказательство теорем, принятие решений и теория игр, планирование и диспетчеризация, прогнозирование.</p> <p>Привести как можно больше примеров, иллюстрирующих правила логических выводов.</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. В чем отличие синтаксического и семантических методов доказательства в логике? 2. Перечислите известные вам методы решения задач. 3. Какие логические виды и методы использует традиционная логика при получении новых знаний? (анализ, синтез, дедукция, индукция, сравнение, аналогия, ...) 4. Достаточно ли дедукции для моделирования аспектов человеческих рассуждений? 5. Что входит в состав простого суждения? 6. Дедуктивные и индуктивные модели формализации знаний. В чем отличие? <p>Источники:</p> <ul style="list-style-type: none"> • Моделирование рассуждений (raai.org) • https://fil.wikireading.ru/92133 • http://csaa.ru/predstavlenie-znanij-v-sistemah-iskusstvennogo/
6	<p>Робототехника</p> <p>Методические рекомендации</p> <p>Сформировать представление о роботе как об интеллектуальной системе, для построения которой необходимы знания из механики, электроники, кибернетики, мехатроники, информатики, ...</p> <p>Рассмотреть на примере робота-пылесоса его компоненты, систему управления, интерфейс, ...</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. На знаниях из каких областей строится современная робототехника? 2. Грозит ли нам восстание машин под предводительством ИИ? 3. Не начнется ли деградация людей, если мы научим компьютер думать за нас? 4. Каковы угрозы, которых мы не ожидали от искусственного интеллекта? 5. Какого робота вы бы хотели иметь дома? Как он должен выглядеть? Какими функциями обладать? 6. Где, по вашему мнению, скоро робот заменит человека? 7. Какие специальности скоро могут исчезнуть из-за роботизации? <p>Источники:</p>

	<ul style="list-style-type: none"> • https://ru.wikipedia.org/wiki/Робототехника
7	<p>Машинное обучение (МО)</p> <p><i>Методические рекомендации</i></p> <p>Основной акцент сделать на общей постановке задачи обучения по прецедентам и способам МО.</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. Назовите области использования МО. 2. Чем обусловлен необычайный успех machine learning? 3. В чем отличие дедуктивного и индуктивного обучения? 4. Какие классические задачи решаются с помощью МО? 5. Какие виды входных данных используются при МО? 6. Как определить качество обучения? 7. Что такое глубокое обучение? <p>Источники:</p> <ul style="list-style-type: none"> • https://ru.wikipedia.org/wiki/Машинное_обучение
8	<p>Машинное творчество</p> <p><i>Методические рекомендации</i></p> <p>Привести примеры написания компьютером музыки, литературных произведений, художественное творчество.</p> <p>Рассмотреть вопросы технического творчества (ТРИЗ).</p> <p>Вопросы:</p> <ol style="list-style-type: none"> 1. Что есть творчество? 2. Какие задачи, которые решаете вы или ваши друзья, можно считать творческими? 3. Как именно искусственный интеллект выстраивает творческий процесс? 4. Какие технологии искусственного интеллекта использует машинное творчество? 5. Можно ли считать произведение, созданное машиной, настоящим искусством? <p>Источники:</p> <ul style="list-style-type: none"> • https://www.theguardian.com/artanddesign/2016/apr/05/new-rembrandt-to-be-unveiled-in-amsterdam • https://www.kv.by/post/1049607-mashinnoe-tvorchestvo • https://www.orange-business.com/ru/blogs/iskusstvennii-intellekt-i-tvorchestvo

Перечень дополнительных вопросов для собеседования

1. Что такое технологии искусственного интеллекта?
2. В каких сферах ИИ применяется уже сейчас?
3. Каковы социальные последствия массового внедрения технологий ИИ?
4. Каковы правовые аспекты внедрения ИИ? Регулируется ли это законодательством?
5. Кто должен нести ответственность за действия искусственного интеллекта?
6. Как выглядит Россия в мировом рейтинге по ИИ?
7. Какие стандарты создания и применения искусственного интеллекта (ИИ) существуют в мире и России?
8. Приведите пример логической модели представления знаний
9. Что такое «Алгоритмическая торговля»? Где и когда она находит применение?
10. Как используется ИИ для управления личными финансами?
11. Дайте понятия сильного и слабого искусственного интеллекта
12. Назовите требования к созданию сильного искусственного интеллекта
13. Назовите основные направления развития ИИ.
14. Что понимают под моделированием рассуждений? Что входит в это направление?
15. В чем суть агентно-ориентированного подхода в ИИ?
16. Какова главная особенность символьных вычислений?
17. Что такое оптические нейронные сети?
18. Приведите примеры применения искусственного интеллекта в области финансов
19. Приведите примеры применения искусственного интеллекта в военном деле
20. Приведите примеры применения искусственного интеллекта в области медицины
21. Приведите примеры применения искусственного интеллекта в области тяжелой промышленности
22. Приведите примеры применения искусственного интеллекта в области транспорта

САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩИХСЯ

Виды самостоятельной работы распределяются в течение семестра. Подготовка к промежуточной аттестации ведется в установленные календарным учебным графиком сроки.

Темы для самостоятельной работы

1. Методология создания онтологий, основанная на системах представления декларативных знаний
2. Экспертная система MYCIN для диагностирования небольшого набора заболеваний
3. Генетические алгоритмы
4. Гибридные интеллектуальные системы
5. Универсальный решатель задач Ньюэлла и Саймона
6. Виртуальный личный помощник
7. Рекомендательные системы
8. Контекстно-зависимые вычисления
9. Представление знаний и вывод на знаниях
10. Молекулярный компьютер
11. Наноробот
12. Проект по компьютерному моделированию головного мозга человека Blue Brain Project
13. Перцептрон - математическая или компьютерная модель восприятия информации мозгом
14. Интеллектуальный анализ данных Data mining
15. Модели представления знаний

ЗАЧЕТ

Вопросы к зачету

1. Определения искусственного интеллекта.
2. Происхождение и понимание термина «искусственный интеллект».
3. Философские предпосылки к возникновению науки.
4. Технологические предпосылки к возникновению науки.
5. История развития искусственного интеллекта в СССР и России.
6. Национальная стратегия развития искусственного интеллекта.
7. Нейрокибернетика и кибернетика «чёрного ящика».
8. Эволюционный подход. Может ли машина мыслить. Тест Тьюринга.
9. Символьный подход.
10. Логический подход.
11. Подход, основанный на использовании интеллектуальных агентов.
12. Сильный и слабый искусственный интеллект. Усиление интеллекта.
13. Моделирование рассуждений.
14. Обработка естественного языка.
15. Экспертные системы.
16. Машинное обучение.
17. Нейронные сети.
18. Интеллектуальная робототехника.
19. Известные ИИ-системы. Примеры эффективного применения систем искусственного интеллекта.
20. Финансы. Медицина. Военное дело. Промышленность. Развлечение и игры. Связь с другими науками и явлениями культуры.

ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ И РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ «ИНТЕРНЕТ»

Рекомендуемый перечень учебной литературы:

1. Искусственный интеллект: современный подход, 4-е издание. Том 1. Решение проблем: знания и рассуждения | Рассел Стюарт, Норвиг Питер, изд-во Диалектика-Вильямс, 2020
2. Девятков В. В. Системы искусственного интеллекта / Гл. ред. И. Б. Фёдоров. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2001. — 352 с. — (Информатика в техническом университете)
3. Нильсон Н. Искусственный интеллект. — М.: Мир, 1973
4. Бруссард Мередит. Искусственный интеллект. Пределы возможного. Изд-во Альпина нон-фикшн, 2020
5. Компьютер учится и рассуждает (ч. 1) // Компьютер обретает разум = Artificial Intelligence Computer Images / под ред. В. Л. Стефанюка. — Москва: Мир, 1990
6. Методы бизнес-анализа / Пол Тернер, Джеймс Кадл, Дебра Пол
7. Ключевые показатели эффективности. 75 показателей, которые должен знать каждый менеджер / Бернард Марр.

Ресурсы информационно-телекоммуникационной сети «Интернет», полезные для самостоятельной работы.

1. История искусственного интеллекта https://ru.wikipedia.org/wiki/История_искусственного_интеллекта
2. Искусственный разум: от философии до нейрона. <http://neural.narod.ru/Main.htm>
3. Моделирование рассуждений Д.А. Поспелов <https://diary.ru/~Organon/p21769784.htm>
4. Общее строение искусственного разума <http://neural.narod.ru/Part2.htm>
5. Анатолий Гершман. Заблуждения искусственного интеллекта. <http://postnauka.ru/faq/80051>
6. Искусственные нейронные сети <http://bigor.bmstu.ru/?cnt/?doc=NN/base.cou>
7. Введение в нейронные сети: Курс Интернет-университета информационных технологий <http://www.intuit.ru/department/ds/intneuronets/>
8. Российский научно-исследовательский институт искусственного интеллекта (РосНИИ ИИ) <http://www.artint.ru>

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение
высшего образования

**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»**

А.А. ФИЛИППОВ

**ОСНОВЫ НАУЧНЫХ
ИССЛЕДОВАНИЙ**
Лабораторный практикум по дисциплине
«Наукометрия»

Ульяновск

УлГТУ

2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Филиппов, Алексей Александрович

Основы научных исследований и наукометрия : лабораторный практикум / А. А. Филиппов. – Ульяновск : УлГТУ, 2021. – 12 с.

Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Наукометрия». Предоставлены варианты заданий, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины. Предназначен для студентов, обучающихся по направлениям: 09.04.03 «Прикладная информатика (магистерская программа Информационно-аналитические системы)», 09.04.04 «Программная инженерия» (магистерская программа «Программные системы, сервисы и платформы») и других специальностей.

Работа подготовлена на кафедре «Информационные системы».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	5
Экзамен	6
Лабораторная работа №1	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №2	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №3	9
Контрольные вопросы к лабораторной работе	9
Лабораторная работа №4	10
Контрольные вопросы к лабораторной работе	10
Список используемой литературы	11

ВВЕДЕНИЕ

Целью освоения дисциплины «Основы научных исследований и наукометрия» является формирование у студентов профессиональных компетенций, связанных с использованием наукометрических показателей в международных базах научного цитирования, формированием профиля ученого в электронной научной библиотеке Elibrary, поддержкой исследований через научные фонды, изучением возможностей научных социальных сетей, использованием системы компьютерной верстки TeX при подготовке научных статей и магистерской диссертации. Особое внимание уделяется изучению наукометрических показателей, методам поиска литературы, особенностям верстки научных работ в системе TeX.

Задачами дисциплины являются:

- изучение основных наукометрических показателей;
- формирование навыков поиска литературы по выбранной тематике;
- рассмотрение основных возможностей научных социальных сетей;
- формирование понимания способов и механизмов поддержки исследований через научные фонды;
- рассмотрение структуры научных работ;
- формирование понимания требований к содержанию научных работ;
- приобретение теоретических знаний и практических навыков написания научных работ с применением системы компьютерной верстки TeX.

Кроме того, в результате изучения дисциплины «Основы научных исследований и наукометрия» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне их формирования.

Тематический план дисциплины:

1. Наукометрические показатели в международных базах научного цитирования

- 1.1. Определение библиометрии и наукометрии
- 1.2. Создание индекса цитирования
- 1.3. Расчет импакт-фактора и дополнительных индексов в Web of Science и Scopus
- 1.4. Сравнение различных областей знания в рейтингах Web of Science и Scopus, квартили, децили
- 1.5. Сроки прохождения публикаций в международных журналах

2. Формирование профиля ученого в электронной научной библиотеке Elibrary

- 2.1. Общее представление о Российском индексе научного цитирования (РИНЦ)
- 2.2. Список журналов рекомендуемых ВАК РФ
- 2.3. Выбор журналов для публикации научных результатов
- 2.4. Выбор конференции для публикации научных результатов

3. Поддержка исследований через научные фонды

- 3.1. Основные фонды поддержки исследований

- 3.2. Российский фонд фундаментальных исследований
- 3.3. Российский научный фонд
- 3.4. Фонд содействия инновациям
- 3.5. Совет по грантам президента РФ

4. Возможности научных социальных сетей

- 4.1. Научная социальная сеть ResearchGate
- 4.2. Научная социальная сеть Google Scholar
- 4.3. Научная социальная сеть Academia.edu

5. Структура и содержание научной работы

- 5.1. Основные разделы научной статьи
- 5.2. Основные разделы магистерской диссертации

6. Основы системы компьютерной верстки TeX

- 6.1. Настройка окружения для работы с TeX
- 6.2. Параграфы, списки и начертание символов
- 6.3. Математические символы и операторы
- 6.4. Изображения и таблицы
- 6.5. Ссылки и библиография
- 6.6. Работа с шаблонами
- 6.7. Создание сложных документов

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки уровня знаний, умений, владений, понимания студентом основных методов и методик проектирования, разработки и поддержки информационных систем в задачах автоматизации бизнес-процессов, умения применять на практике полученных знаний. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений. Шкала оценивания имеет вид (таблица П1)

Таблица П1

Шкала и критерии оценивания решения задач на лабораторных занятиях

Оценка	Критерии
Сдано	Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку.
Не сдано	Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.

Экзамен

Экзамен по дисциплине проводится в форме решения на компьютере (написания программы) практических задач по билетам, а также ответа на сопутствующие теоретические вопросы. Билет содержит практическое задание (задачу) для контроля освоения умений и навыков всех запланированных в ходе изучений дисциплины компетенций. Билет формируется таким образом, чтобы в него попали практические задания, контролирующие уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица ПЗ)

Таблица ПЗ

Шкала и критерии оценивания экзамена

Оценка	Критерии
Отлично	Выставляется обучающемуся, если студент выполнил в полном объеме практическое задание и способен обосновать свои решения
Хорошо	Выставляется обучающемуся, если студент выполнил практическое задание не в полном объеме (не менее $\frac{3}{4}$) либо в полном объеме, но с некоторыми погрешностями и ошибками
Удовлетворительно	Выставляется обучающемуся, если студент выполнил практическое задание не в полном объеме (не менее $\frac{1}{2}$) либо в полном объеме, но с существенными погрешностями и ошибками
Неудовлетворительно	Выставляется обучающемуся, если студент не справился с выполнением практического задания

Лабораторная работа №1

Необходимо сформулировать цели, задачи, новизны и положений, выносимых на защиту для Вашей научной работы.

Контрольные вопросы к лабораторной работе

1. Как формулируется цель исследования?
2. Как формулируются задачи исследования?
3. Как формируется объект исследования?
4. Как формируется предмет исследования?
5. Как формируются пункты научной новизны научного исследования?
6. Как формируются положения научного исследования, выносимые на защиту?
7. Какова структура магистерской диссертации? Дайте пояснение каждому элементу структуры.
8. Какова структура научной публикации? Дайте пояснение каждому элементу структуры.

Лабораторная работа №2

Необходимо составить научную статью в системе компьютерной верстки TeX согласно заданию преподавателя.

Контрольные вопросы к лабораторной работе

1. Основные команды TeX.
2. Основные операторы TeX.
3. Пакеты TeX.
4. Шаблон документа.
5. Набор формул.
6. Набор текста.
7. Вставка рисунков.

Лабораторная работа №3

Необходимо составить научную презентацию в системе компьютерной верстки TeX согласно заданию преподавателя.

Контрольные вопросы к лабораторной работе

1. Особенности составления презентаций в TeX.
2. Шаблоны презентаций.
3. Настройка внешнего вида презентации.

Лабораторная работа №4

Необходимо составить основу рукописи магистерской диссертации в системе компьютерной верстки TeX с использованием шаблона disser.

<https://github.com/AndreyAkinshin/Russian-Phd-LaTeX-Dissertation-Template>

Контрольные вопросы к лабораторной работе

1. Использование сторонних шаблонов документов.
2. Создание новых команд.
3. Модификация стандартных классов.

Список используемой литературы

1. Тронин, В.Г. Планирование и управление научными проектами с применением современных информационно-коммуникационных технологий : учебное пособие / В. Г. Тронин. – Ульяновск : УлГТУ, 2017. – 211 с.
<http://venec.ulstu.ru/lib/disk/2017/145.pdf>
2. Тронин, В.Г. Оценка результата в научно-исследовательской работе и наукометрия : учебное пособие / В. Г. Тронин, А. Р. Сафиуллин. – Ульяновск : УлГТУ, 2019. – 136 с.
<http://venec.ulstu.ru/lib/disk/2017/561.pdf>
3. Семушин И.В. Написание и презентация научной работы – Существенные навыки для студентов, магистрантов и аспирантов: Электронное учебное пособие / Составитель и разработчик макета И. В. Семушин. – Ульяновск: УлГТУ, 2013. – 1148 слайдов (312 фреймов).
<http://venec.ulstu.ru/lib/disk/2014/Semushin-root/Semushin.pdf>
4. Львовский, С.М. Работа в системе LaTeX : учебное пособие / С.М. Львовский. — 2-е изд. — Москва : ИНТУИТ, 2016. — 534 с.
<https://e.lanbook.com/book/100443>
5. Материалы сайта LaTeX - Викиучебник
<https://ru.wikibooks.org/wiki/LaTeX>
6. Overleaf - Online LaTeX Editor
<https://www.overleaf.com/>

**ОСНОВЫ НАУЧНЫХ
ИССЛЕДОВАНИЙ:**
практикум к выполнению
лабораторных работ по дисциплине
«Основы научных исследований и наукометрия»

Автор ФИЛИППОВ Алексей Александрович

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

А.А. ФИЛИППОВ
А.М. СКАЛКИН

СИСТЕМЫ И ТЕХНОЛОГИИ WEB-АНАЛИТИКИ
методические рекомендации по выполнению
лабораторных работ по дисциплине
«Системы и технологии Web-аналитики»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Филиппов, Алексей Александрович

Системы и технологии Web-аналитики : методические рекомендации по выполнению лабораторных работ / А. А. Филиппов, А.М. Скалкин. – Ульяновск : УлГТУ, 2021. – 12 с.

Методические рекомендации по выполнению лабораторных работ адресованы студентам для выполнения и оформления лабораторных работ по дисциплине «Системы и технологии Web-аналитики». В данных методических рекомендациях предоставлены варианты заданий, рекомендации и требования к лабораторным работам. Методические рекомендации разработаны в соответствии с рабочей программой дисциплины «Системы и технологии Web-аналитики». Методические рекомендации предназначены для студентов, обучающихся по направлению 09.04.03 «Прикладная информатика» профиль «Искусственный интеллект и бизнес-аналитика».

Работа подготовлена на кафедре «Информационные системы».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	4
Экзамен	5
Лабораторная работа №1	6
Контрольные вопросы к лабораторной работе	6
Лабораторная работа №2	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №3	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №4	9
Контрольные вопросы к лабораторной работе	9
Лабораторная работа №5	10
Контрольные вопросы к лабораторной работе	10
Список используемой литературы	11

ВВЕДЕНИЕ

Целью освоения дисциплины «Системы и технологии Web-аналитики» является формирование представления о методах и подходах построения систем Web-аналитики, а также средствах и технологиях, на которых основываются системы Web-аналитики.

Задачами дисциплины являются формирование компетенций для решения следующих профессиональных задач:

- анализ и исследование существующих методов работы с данными, положенных в основу современных систем Web-аналитики;

- анализ и исследование существующих методов и алгоритмов, положенных в основу современных систем Web-аналитики;

- исследование подходов и архитектурных решений для построения систем Web-аналитики.

Тематический план дисциплины:

Раздел 1. Методы анализа данных в системах Web-аналитики

1.1 Виды данных для Web-аналитики и их источники

1.2 Извлечение и предобработка данных в системах Web-аналитики

1.3 Анализ данных в системах Web-аналитики

Раздел 2. Методы Web-аналитики

2.1 Методы искусственного интеллекта для Web-аналитики

2.2 Инструментальные средства искусственного интеллекта для Web-аналитики

2.3 Алгоритмы рекомендательных систем

Раздел 3. Подходы к построению систем Web-аналитики

3.1 Архитектурные решения для построения систем Web-аналитики

3.1 Информационная безопасность в системах Web-аналитики

3.2 Средства, методы и подходы разработки систем Web-аналитики

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки освоения студентом всех запланированных по дисциплине «Системы и технологии Web-аналитики» компетенций. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений. Шкала оценивания имеет вид (таблица 1).

Таблица 1

Шкала оценивания	
Процент правильных ответов	Балл
Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает	Сдано

правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку.	
Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.	Не сдано

Экзамен

Экзамен по дисциплине проводится в устной форме по билетам с ответом на сопутствующие теоретические вопросы. Билет содержит теоретические вопросы для контроля освоения умений и навыков всех запланированных в ходе изучений дисциплины компетенций. Билет формируется таким образом, чтобы в него попали вопросы, контролируемые уровнем усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица 2)

Таблица 2

Шкала оценивания	
Критерии оценки уровня сформированности компетенций по дисциплине	Балл
выставляется обучающемуся, если он показал глубокие знания материала по поставленному вопросу, грамотно, логично и стройно его излагает	Отлично
выставляется обучающемуся, если он твердо знает материал, грамотно его излагает, но допускает несущественные неточности в ответе на вопрос	Хорошо
выставляется обучающемуся, если он показывает знания только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допускает отдельные неточности	Удовлетворительно
выставляется обучающемуся, если он допускает грубые ошибки в ответе на поставленный вопрос	Неудовлетворительно

Лабораторная работа №1

Необходимо провести подбор и извлечение данных для Web-аналитики.

Результатом лабораторной работы №1 является выбор возможных источников для получения данных для последующей Web-аналитики.

Необходимо:

1. Проанализировать возможные источники для получения данных.
2. Проанализировать существующие средства для извлечения данных с Web-ресурсов.
3. Составить отчет, содержащий описание полученных данных и способов их получения.

Контрольные вопросы к лабораторной работе

1. Что представляет собой Web-ресурс?
2. Какие существуют способы получения данных?
3. Какие полезные данные можно получить с Web-ресурса для проведения Web-аналитики?
4. Какие программные системы существуют для получения данных с ресурса?
5. Какие выводы можно сделать о Web-ресурсе на основе собранных данных?

Лабораторная работа №2

Необходимо провести анализ, выбор и разработку методов искусственного интеллекта для Web-аналитики.

Результатом лабораторной работы №2 является разработанный метод Web-аналитики с применением методов искусственного интеллекта.

Необходимо:

1. Проанализировать существующие научные работы по тематике интеллектуального анализа данных в области Web-аналитики.
2. Проанализировать существующие методы искусственного интеллекта подходящие для Web-аналитики.
3. Разработать метод искусственного интеллекта для Web-аналитики на основе проведенного анализа.
4. Составить отчет, содержащий сравнительный анализ методов искусственного интеллекта для Web-аналитики.

Контрольные вопросы к лабораторной работе

1. Какие методы искусственного интеллекта существуют?
2. Какие из них подходят для Web-аналитики?
3. Какие методы лучше подходят для каких данных?
4. Почему был выбран именно данный метод?
5. Применяется ли данный метод в других направлениях аналитики?

Лабораторная работа №3

Необходимо провести анализ, выбор и разработку инструментальных средств для построения системы Web-аналитики.

Результатом лабораторной работы №3 является разработанное инструментальное средство для Web-аналитики с использованием методов искусственного интеллекта.

Необходимо:

1. Проанализировать существующие инструментальные средства для Web-аналитики на основе методов искусственного интеллекта.
2. Выбрать наиболее подходящий метод интеллектуального анализа для собранных данных.
3. Разработать инструментальное средство анализа собранных данных.
4. Составить отчет, содержащий сравнительный анализ инструментальных средств искусственного интеллекта и описание разработанного инструментального средства Web-аналитики.

Контрольные вопросы к лабораторной работе

1. В чем состоит отличие методов от инструментов искусственного интеллекта?
2. Какие инструменты искусственного интеллекта существуют?
3. Существуют ли инструменты, направленные на выполнение задач Web-аналитики?
4. Обоснуйте выбор своего инструмента?
5. Какие результаты были получены?

Лабораторная работа №4

Необходимо провести анализ, выбор и разработку алгоритмов рекомендательных систем.

Результатом лабораторной работы №4 является реализация собственной рекомендательной системы.

Необходимо:

1. Проанализировать существующие алгоритмы рекомендательных систем.
2. Проанализировать использование популярными сервисами различных алгоритмов рекомендательных систем.
3. Выбрать рекомендательный алгоритм, подходящий под собранные данные.
4. Разработать рекомендательную систему на основе собранных данных и с использованием выбранного алгоритма рекомендации.
5. Составить отчет, содержащий сравнительный анализ архитектур рекомендательных систем и описание разработанной рекомендательной системы.

Контрольные вопросы к лабораторной работе

1. Что такое рекомендательная система?
2. Какие существуют алгоритмы рекомендательных систем?
3. Приведите пример использования различных алгоритмов рекомендательных систем?
4. Почему для работы был выбран данный алгоритм?
5. Какие алгоритмы подходят для каких данных?

Лабораторная работа №5

Необходимо разработать средства обеспечения информационной безопасности в системах Web-аналитики.

Результатом лабораторной работы №5 является интеграция средств обеспечения информационной безопасности в разработанную рекомендательную систему.

Необходимо:

1. Проанализировать существующие средства обеспечения информационной безопасности.
2. Выявить средства информационной безопасности, подходящие для проектов Web-аналитики с использованием методов искусственного интеллекта.
3. Реализовать выбранные средства информационной безопасности в разрабатываемом проекте.
4. Составить отчет, содержащий сравнительный анализ существующих средств обеспечения информационной безопасности и описание реализации методов обеспечения информационной безопасности в разрабатываемом проекте.

Контрольные вопросы к лабораторной работе

1. Что такое информационная безопасность?
2. Какие существуют способы обеспечения информационной безопасности в системах Web-аналитики?
3. Какие существуют особенности обеспечения информационной безопасности в системах искусственного интеллекта?
4. Какие виды обеспечения информационной безопасности использовались в работе?
5. Какое влияние оказывает использование средств информационной безопасности на проект в целом?

Список используемой литературы

1. Тюгашев, А. А. Интеллектуальные системы : учебное пособие / А. А. Тюгашев. — Самара : СамГУПС, 2020. — 151 с. — ISBN 978-5-98941-326-3. — Текст : электронный // Лань : электронно-библиотечная система.

URL: <https://e.lanbook.com/book/161308> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

2. Ясницкий, Л. Н. Интеллектуальные системы : учебник / Л. Н. Ясницкий. — 2-е изд. — Москва : Лаборатория знаний, 2020. — 224 с. — ISBN 978-5-00101-897-1. — Текст : электронный // Лань : электронно-библиотечная система.

URL: <https://e.lanbook.com/book/151510> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

3. Прокопенко, Н. Ю. Аналитические информационные системы поддержки принятия решений : учебное пособие / Н. Ю. Прокопенко. — Нижний Новгород : ННГАСУ, 2020. — 142 с. — ISBN 978-5-528-00395-5. — Текст : электронный // Лань : электронно-библиотечная система.

URL: <https://e.lanbook.com/book/164866> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

4. Баллод, Б. А. Методы и средства социологических исследований : учебное пособие / Б. А. Баллод. — Санкт-Петербург : Лань, 2019. — 184 с. — ISBN 978-5-8114-3778-8. — Текст : электронный // Лань : электронно-библиотечная система.

URL: <https://e.lanbook.com/book/122170> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

5. Краковский, Ю. М. Методы защиты информации : учебное пособие для вузов / Ю. М. Краковский. — 3-е изд., перераб. — Санкт-Петербург : Лань, 2021. — 236 с. — ISBN 978-5-8114-5632-1. — Текст : электронный // Лань : электронно-библиотечная система.

URL: <https://e.lanbook.com/book/156401> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

6. Национальный Открытый Университет «ИНТУИТ»

URL: <https://intuit.ru>

7. Википедия – свободная энциклопедия

URL: <https://ru.wikipedia.org>

8. Электронно-библиотечная система Лань

URL: <https://e.lanbook.com/>

9. Данные с Kaggle

URL: <https://www.kaggle.com/>

СИСТЕМЫ И ТЕХНОЛОГИИ WEB-АНАЛИТИКИ:

методические рекомендации

по выполнению лабораторных работ по дисциплине

«Системы и технологии Web-аналитики»

ФИЛИППОВ Алексей Александрович

СКАЛКИН Антон Михайлович

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»**

Г.Ю. Гуськов

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ
УПРАВЛЕНИЯ РЕСУРСАМИ ПРЕДПРИЯТИЯ**
лабораторный практикум по дисциплине
«Системы управления ресурсами предприятия»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Гуськов, Глеб Юрьевич

Проектирование и разработка информационных систем управления ресурсами предприятия: практикум / Г.Ю. Гуськов. – Ульяновск : УлГТУ, 2021. – 12 с.

Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Системы управления ресурсами предприятия». Предоставлены задания, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины.

Предназначен для студентов, обучающихся по направлениям: 09.04.03 «Прикладная информатика» (программа Искусственный интеллект и бизнес-аналитика)

Работа подготовлена на кафедре «Информационные системы».

© Гуськов Г.Ю., 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	5
Зачет	5
Лабораторная работа №1	6
Контрольные вопросы к лабораторной работе	6
Лабораторная работа №2	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №3	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №4	9
Контрольные вопросы к лабораторной работе	9
Лабораторная работа №5	10
Контрольные вопросы к лабораторной работе	10
Список используемой литературы	11

ВВЕДЕНИЕ

Целью освоения дисциплины «Системы управления ресурсами предприятия» является формирование у будущих выпускников компетенций в области создания и управления информационными ресурсами предприятия, а также интеллектуальных модулей в составе информационных систем.

Задачами дисциплины являются:

- исследование методологий разработки программного обеспечения;
- использование, эксплуатация и обслуживание систем, предназначенных для хранения, обработки, передачи структурированной информации и документов в разрезе различных бизнес-процессов;
- применение технологий и стандартов, используемых в системах электронного документооборота при разработке информационных систем.
- изучение основ управления проектной деятельностью, методов планирования и реализации проектов на основе стандарта РМВОК.
- приобретение навыков управления разработкой и реализацией проекта.

Задачами дисциплины являются:

- изучение основных положений стандарта РМВОК
- изучение различных аспектов проектного менеджмента
- изучение жизненного цикла проекта
- освоение методов управления разработкой проекта и методов управления реализацией проекта.

Кроме того, в результате изучения дисциплины «Системы управления ресурсами предприятия» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне их формирования.

Тематический план дисциплины

1. Раздел 1. Информационные системы

1.1. Информационные системы. Ценность и количество информации. Системы информационного обмена. Информационная сеть.

1.2. Состав и структура информационных систем. Задачи информационных систем. Цели использования профилей информационной системы.

1.3. Профиль информационной системы. Профиль пользователя информационной системы. Принципы формирования профилей. Структура профилей информационной системы.

1.4. Открытая информационная система. Принципы построения открытых информационных систем. Модели открытых информационных систем. Эталонная модель среда открытой системы.

Раздел 2. Корпоративные информационные системы

2.1 Классификация систем управления предприятием.

2.2. Метод управления MRP. Структура MRP-системы. Системы планирования производственных мощностей.

2.3. Метод управления MRP II. Структура MRP II-системы.

2.4. Scada-системы.

2.5. ERP-системы.

2.6. CRM-системы. Категории продуктов класса CRM.

2.7. PRM-системы.

Раздел 3. Информационные системы, основанные на знаниях

3.1. Информационные ресурсы, основанные на базах знаний. Базы знаний. Методы работы с базами знаний.

3.2. Проектные репозитории. Работа с проектными репозиториями. Системы контроля версий.

3.3. Интеграция баз знаний с проектными репозиториями.

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки освоения студентом всех запланированных по дисциплине «Системы управления ресурсами предприятия» компетенций. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений.

Шкала оценивания имеет вид (таблица П1)

Таблица П1

Оценка	Критерии
Сдано	Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку.
Не сдано	Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.

Зачет

Зачет по дисциплине проводится в форме ответа на теоретические вопросы билета. Билет содержит два теоретических вопроса, обеспечивающих контроль освоения умений и навыков всех, запланированных в ходе изучений дисциплины компетенций. Билет формируется таким образом, чтобы в него комплексные вопросы, контролирующие уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица П2)

Таблица П2

Оценка	Критерии
Зачтено	Выставляется обучающемуся, если студент в полном объеме ответил на оба теоретических вопроса и дал пояснения на ряд дополнительных вопросов
Не зачтено	Выставляется обучающемуся, если студент не смог ответить ни на один из предложенных вопросов

Лабораторная работа №1

EMC-Системы – это класс систем позволяющих управлять контентом. В ходе работы необходимо:

1. Выбрать одну из систем. Установить временную версию и изучить её особенности и возможности.
2. Сформировать краткое сообщение о возможностях системы подходах к интеграции с ней.
3. Представить сообщение на занятии и продемонстрировать основные возможности EMC-системы.

Контрольные вопросы к лабораторной работе

1. Что такое EMC системы?
2. Какие этапы проходит система при внедрении?
3. Как определяются бизнес-требования?
4. Как реализуется гибкость в ЕСМ системах под требования?
5. Приведите примеры инструментов моделирования бизнес-процессов.

Лабораторная работа №2

Проектирование подсистемы корпоративной информационной системы. Рассмотреть информационную систему, разрабатываемую в ходе диссертационного исследования с позиций корпоративных информационных систем.

Необходимо:

- 1) Сформировать диаграмму классов по интеграции информационную систему, разрабатываемую в ходе диссертационного исследования в корпоративные информационные системы.
- 2) Сформировать диаграмму прецедентов по интеграции информационную систему, разрабатываемую в ходе диссертационного исследования в корпоративные информационные системы.
- 3) Сформировать диаграмму последовательностей по интеграции информационную систему, разрабатываемую в ходе диссертационного исследования в корпоративные информационные системы.

Контрольные вопросы к лабораторной работе

1. Какие модели жизненного цикла вы знаете?
2. Из каких этапов состоит каскадная модель жизненного цикла?
3. Из каких процессов состоит этап проектирования?
4. Что является результатом этапа анализа предметной области?
5. Как распределяется время, отводимое на создание ИС между этапами?

Лабораторная работа №3

Методологии и модели жизненного цикла информационных систем.

Необходимо:

- 1) Выбрать одну из методологий разработки программного обеспечения и модель жизненного цикла.
- 2) Сформировать план диссертационного исследования в рамках выбранной методологии.

Контрольные вопросы к лабораторной работе

1. Какие требования к ИС необходимо сформулировать?
2. Что определяется требованиями к ИС?
3. Охарактеризуйте связь между требованиями и функциями ИС?
4. Требования к ИС формируются лишь однажды или итерационно?
5. Какие требования к ИС относятся к функциональным?
6. Какие требования к ИС относятся к нефункциональным?

Лабораторная работа №4

Протитипирование информационных систем.

Необходимо:

- 1) Разработать пользовательский интерфейс разрабатываемой информационной системы.
- 2) Сформировать переходы между экранами пользовательского интерфейса.
- 3) Разработать горизонтальный прототип разрабатываемой системы.

Контрольные вопросы к лабораторной работе

1. Из чего состоит диаграмма классов?
2. Какая из UML диаграмм как правило строиться первой?
3. Какая база данных называется реляционной?
4. Какие типы базы данных кроме реляционной вы знаете?
5. Как разбить программу на модули?
6. Возможно ли совместить микросервисную архитектуру программного обеспечения с объектно-ориентированной парадигмой?

Лабораторная работа №5

Планирование временных затрат на реализацию информационных систем.
Необходимо:

- 1) Разделить задачи на типы выполняемых работ.
- 2) Дополнить план работ графиком выполнения задач разбитым на итерации.

Контрольные вопросы к лабораторной работе

1. Какие задачи входят в анализ предметной области, и какие специалисты их выполняют?
2. Какие задачи входят в проектирование ИС и какие специалисты их выполняют?
3. Какие задачи входят в разработку бизнес логики ИС и какие специалисты их выполняют?
4. Какие задачи входят разработку пользовательского интерфейса, и какие специалисты их выполняют?
5. Какие задачи входят в тестирование ИС и какие специалисты их выполняют?
6. Какие задачи входят в этап внедрения ИС и какие специалисты их выполняют?

Список используемой литературы

1. Остроух, А. В. Проектирование информационных систем : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 164 с. — ISBN 978-5-8114-8377-8.

URL: <https://e.lanbook.com/book/175513> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

2. Андрианова, Е. Г. Информационные системы управления ресурсами предприятия : методические рекомендации / Е. Г. Андрианова. — Москва : РТУ МИРЭА, 2020. — 63 с.

URL: <https://e.lanbook.com/book/167615> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

Гуськов Глеб Юрьевич

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНЫХ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ
лабораторный практикум по дисциплине
«Системы управления ресурсами предприятия»**

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

В. В. Воронина

Методы искусственного интеллекта в предиктивной аналитике и бизнес-аналитике

Методические указания для проведения лабораторных занятий

для студентов направлений

09.04.04 «Программная инженерия»

профиль Искусственный интеллект и предиктивная аналитика

09.04.03 «Прикладная информатика»

профиль Искусственный интеллект и бизнес-аналитика

Ульяновск
УлГТУ
2021

УДК 004 (076)

ББК 32.973-018.1я7

В12

Рецензент

декан факультета информационных систем и технологий, канд. техн. наук, доцент
К.В. Святков.

Рекомендовано научно-методической комиссией факультета
информационных систем и технологий в качестве практикума.

Воронина, Валерия Вадимовна

- В12 Методы искусственного интеллекта в предиктивной и бизнес-аналитике : практикум для проведения лабораторных занятий для студентов направлений 09.04.04 «Программная инженерия» профиль Искусственный интеллект и предиктивная аналитика, 09.04.03 «Прикладная информатика» профиль Искусственный интеллект и бизнес-аналитика / В. В. Воронина. – Ульяновск : УлГТУ, 2021. – 26 с.

Составлен в соответствии с учебными планами направлений 09.04.04 «Программная инженерия» профиль Искусственный интеллект и предиктивная аналитика, 09.04.03 «Прикладная информатика» профиль Искусственный интеллект и бизнес-аналитика. Цель данного практикума – ориентировать студентов на содержание и порядок выполнения лабораторных работ во время прохождения ими курсов «Методы искусственного интеллекта в предиктивной аналитике» и «Методы искусственного интеллекта в бизнес-аналитике». Даются задания для лабораторных занятий, а также приводится обзор возможных данных для анализа.

Работа подготовлена на кафедре «Информационные системы».

УДК 004 (076)

ББК 32.973-018.1я7

© В. В. Воронина, 2021.

© Оформление. УлГТУ, 2021.

Оглавление

Общая информация	3
Задания на лабораторные работы по предмету «Методы искусственного интеллекта в бизнес-аналитике».....	3
Задания на лабораторные работы по предмету «Методы искусственного интеллекта в предиктивной аналитике».....	5
Предлагаемые массивы данных для анализа	9
Список используемых источников	26

Общая информация

В рамках лабораторного практикума вам предлагается проанализировать данные с платформы Kaggle. Вы можете взять один из предложенных массивов данных, или выбрать свой. Согласно [1]: «Kaggle – это платформа для исследователей разных уровней, где они могут опробовать свои модели анализа данных на серьезных и актуальных задачах. Суть такого ресурса – не только в возможности получить неплохой денежный приз в случае, если именно ваша модель окажется лучшей, но и в том (а, это, пожалуй, гораздо важнее), чтобы набраться опыта и стать специалистом в области анализа данных и машинного обучения. Ведь самый важный вопрос, зачастую стоящий перед такого рода специалистами – где найти реальные задачи? Здесь их достаточно».

Общее требование к лабораторным. Организовать следующую архитектуру системы: каждый метод анализа данных оформить в виде веб-сервиса и создать управляющую систему, вызывающую реализованные веб-сервисы.

Задания на лабораторные работы по предмету «Методы искусственного интеллекта в бизнес-аналитике»

Лабораторная работа 1. Работа с данными и проектирование общей системы

Исходя из общих требований, разработать архитектуру системы для выполнения всех лабораторных работ. Выбрать массив данных для анализа, определить, какие задачи классификации, кластеризации и прогнозирования вы сможете на нем решить (по две задачи каждого типа). Выбрать необходимые признаки, осуществить загрузку, очистку и визуализацию данных.

Вопросы для собеседования по лабораторной работе:

1. В каких бизнес-процессах может быть использован выбранный вами набор данных?

2. Из каких источников данных может быть получен выбранный вами набор данных?
3. Решение каких задач с использованием искусственного интеллекта будет эффективно для выбранного вами набора данных?
4. Каким образом вы подбирали признаки для ваших задач?
5. Какова будет архитектура вашей системы и какие функции?

Лабораторная работа 2. Решение задач прогнозирования

Для выбранного вами массива данных решите две различных задачи прогнозирования различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи прогнозирования?
3. В чем особенность нейросетевых методов для решения задачи прогнозирования?
4. В чем особенность нечетких методов для решения задачи предсказания?
5. Расскажите об особенностях TimeSeries DataMining в контексте решаемой задачи.

Лабораторная работа 3. Решение задач классификации

Для выбранного вами массива данных решите две различных задачи классификации различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи классификации?
3. В чем особенность нейросетевых методов для решения задачи классификации?
4. В чем особенность нечетких методов для решения задачи классификации?
5. В чем особенность деревьев решений для решения задачи классификации?

Лабораторная работа 4. Решение задач кластеризации

Для выбранного вами массива данных решите две различных задачи кластеризации различными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. В чем особенность задачи кластеризации в отличие от классификации?
2. Почему в задачах кластеризации так важен выбор критерия качества?
3. Для решения каких задач используется кластеризация, как метод обработки данных?
4. Какие вы знаете методы кластеризации данных?
5. Какие методы для решения задач вы выбрали и почему?

Лабораторная работа 5. Оценка качества и выбор лучшей модели

Для лабораторных работ 2-4 выполните решение каждой из задач не используемым в лабораторных методом. Оцените каждое из решений и выберите лучшую модель для каждой задачи.

Вопросы для собеседования по лабораторной работе:

1. Какими критериями можно оценить качество работы кластеризационной модели?
2. Какими критериями можно оценить качество работы предсказательной модели?
3. Как переобучение влияет на качество работы модели?
4. Какие есть методы борьбы с переобучением?
5. Расскажите подробнее про кросс-валидацию.

Задания на лабораторные работы по предмету «Методы искусственного интеллекта в предиктивной аналитике»

Лабораторная работа 1. Работа с данными и проектирование общей системы

Исходя из общих требований, разработать архитектуру системы для выполнения всех лабораторных работ. Выбрать массив данных для анализа. Выбрать перспективные признаки для решения задач прогнозирования, осуществить загрузку, очистку и визуализацию данных.

Вопросы для собеседования по лабораторной работе:

1. В каких бизнес-процессах может быть использован выбранный вами набор данных?
2. Из каких источников данных может быть получен выбранный вами набор данных?
3. Решение каких задач предсказания с использованием искусственного интеллекта будет эффективно для выбранного вами набора данных?

4. Какими критериями вы можете оценить эффективность внедрения искусственного интеллекта в анализ выбранного вами набора данных?
5. Какова будет архитектура вашей системы и какие функции?

Лабораторная работа 2. Решение задач прогнозирования регрессионными методами

Для выбранного вами массива данных решите две различных задачи прогнозирования регрессионными методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность регрессионных методов решения задачи прогнозирования?
3. Как подготовить текстовые данные для обработки линейной регрессией?
4. Что общего и чем отличаются модели Ridge и Lasso?
5. Для каких задач хорошо подходят линейные регрессионные модели?

Лабораторная работа 3. Решение задач прогнозирования нейросетевыми методами

Для выбранного вами массива данных решите две различных задачи прогнозирования нейросетевыми методами. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какие методы для решения задач вы выбрали и почему?
2. В чем особенность нейросетевых методов для решения задачи прогнозирования?
3. Как работа с нейронными сетями реализована в python?
4. Как необходимо готовить данные для обработки нейронными сетями?
5. Как вы выполнили визуализацию данных?

Лабораторная работа 4. Решение задач прогнозирования методами нечеткой логики

Для выбранного вами массива данных решите задачу прогнозирования методами нечеткой логики. Выполните визуализацию результатов.

Вопросы для собеседования по лабораторной работе:

1. Какую задачу вы выбрали для решения и почему?

2. В чем особенность нечетких методов для решения задачи предсказания?
3. Расскажите о TimeSeries DataMining.
4. Расскажите подробнее о нечеткой регрессии.
5. В чем преимущества и недостатки использования методов нечеткой логики?

Лабораторная работа 5. Оценка качества и выбор лучшей модели

Для лабораторных работ 2-3 выполните решение каждой из задач не используемым в лабораторных методом. Оцените каждое из решений и выберите лучшую модель для каждой задачи.

Вопросы для собеседования по лабораторной работе:

1. Какими критериями можно оценить качество работы предсказательной модели?
2. Как переобучение влияет на качество работы модели?
3. Какие есть методы борьбы с переобучением?
4. Расскажите подробнее про кросс-валидацию.
5. Какие методы решения задач вы выбрали и почему?

Задания на лабораторные работы по предмету «Методы глубокого обучения в бизнес-аналитике»

Лабораторная работа 1. Работа со сверточной нейронной сетью. Часть 1

Ознакомьтесь с пунктом «Сравнение эффективности моделей из библиотеки Keras» (ресурс [3], стр. 263). Проведите серию экспериментов на тестовых данных, интерпретируйте полученные результаты, ответьте на вопросы, сформулированные в конце пункта. Для одно из приведенных ниже наборов данных предложите задачу и соответствующее дополнение этого набора данных, чтобы применение сверточной сети для его анализа было целесообразным.

Вопросы для собеседования по лабораторной работе:

1. Как необходимо подготовить данные для работы со сверточной нейронной сетью?
2. Для решения каких задач лучше всего подходят сверточные нейронные сети?
3. В чем сложность работы с данным видом моделей глубокого обучения?
4. Что вы научились делать в ходе лабораторной работы?
5. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?

Лабораторная работа 2. Работа со сверточной нейронной сетью. Часть 2

Ознакомьтесь с пунктом «Работа с библиотекой OpenCV» (ресурс [3], стр. 265). Используя библиотеку OpenCV, сравните эффективность моделей, построенных в предыдущей лабораторной работе на массиве данных фотографий.

Вопросы для собеседования по лабораторной работе:

1. В чем преимущества и недостатки сверточной нейронной сети по сравнению с другими моделями?
2. Что вы научились делать в ходе лабораторной работы?
3. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?
4. Какой объем экспериментов вам пришлось проделать?
5. Что вы можете сказать о качестве исследуемых моделей?

Лабораторная работа 3. Работа с рекуррентной сетью. Часть 1

Ознакомьтесь с пунктом «Работа с рекуррентными сетями» (ресурс [3], стр. 274), выполните сформулированные задачи.

Вопросы для собеседования по лабораторной работе:

1. Как необходимо подготовить данные для работы с рекуррентной нейронной сетью?
2. Для решения каких задач лучше всего подходят рекуррентные нейронные сети?
3. В чем сложность работы с данным видом моделей глубокого обучения?
4. Что вы научились делать в ходе лабораторной работы?
5. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?

Лабораторная работа 4. Работа с рекуррентной сетью. Часть 2

Выберите один из приведенных ниже наборов данных, содержащий текст или самостоятельно найдите на Kaggle [2] подходящий набор. Исходя из контекста данных сформулируйте задачу обработки текста и решите ее с помощью рекуррентной нейронной сети.

Вопросы для собеседования по лабораторной работе:

1. В чем преимущества и недостатки рекуррентной нейронной сети по сравнению с другими моделями в плане решения задачи обработки текста?
2. Что вы научились делать в ходе лабораторной работы?
3. С какими сложностями вы столкнулись в ходе выполнения лабораторной работы?
4. Какой объем экспериментов вам пришлось проделать?

5. Что вы можете сказать о качестве вашего решения?

Лабораторная работа 5. Разработка коммерческого предложения

Подберите комплектацию самого дешевого, среднего и самого дорогого компьютера (и «железо» и программное обеспечение) для использования методов глубокого обучения в бизнес-аналитике. Рассмотрите варианты аренды серверов. Сделайте выводы, при каких задачах какой вариант предпочтительнее. Оформите лабораторную работу в виде коммерческого предложения для какого-либо бизнеса с представленным анализом различных вариантов и обоснованием основного.

Вопросы для собеседования по лабораторной работе:

1. Что такое CUDA-вычисления? Когда они необходимы и на каком «железе» выполняются?
2. Что может служить критерием целесообразности использования методов глубокого обучения в бизнес-аналитике?
3. В чем сложность внедрения глубокого обучения в бизнес-аналитику?
4. Из чего складывается стоимость внедрения глубокого обучения в бизнес-аналитику?
5. Что вы научились делать в ходе лабораторной работы?

Предлагаемые массивы данных для анализа

Рестораны Бангалора

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

Приведенное описание набора данных:

«The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant, establishment of different types of restaurant at different places, Bengaluru being one such city has more than 12,000 restaurants with restaurants serving dishes from all over the world. With each day new restaurants opening the industry has'nt been saturated yet and the demand is increasing day by day. Inspite of increasing demand it however has become difficult for new restaurants to compete with established restaurants. Most of them serving the same food. Bengaluru being an IT capital of India. Most of the people here are dependent mainly on the restaurant food as they don't have time to cook for themselves. With such an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality. Do the entire locality loves vegetarian food. If yes then is that locality populated by a particular sect of people for eg. Jain, Marwaris,

Gujaratis who are mostly vegetarian. These kind of analysis can be done using the data, by studying different factors.»

Перевод:

«Основная идея анализа набора данных Zomato состоит в том, чтобы получить четкое представление о факторах, влияющих на совокупный рейтинг каждого ресторана, создание различных типов ресторанов в разных местах, поскольку в одном из таких городов в Бангалоре насчитывается более 12 000 ресторанов, в которых подают блюда из по всему миру. С каждым днем новые рестораны открываются, индустрия еще не насыщена, и спрос растет день ото дня. Несмотря на растущий спрос, новым ресторанам стало трудно конкурировать с известными ресторанами. Большинство из них подают одинаковую еду. Бангалор, являющийся ИТ-столицей Индии. Большинство людей здесь зависят в основном от ресторанной еды, так как у них нет времени готовить для себя. Поэтому при таком огромном спросе на рестораны стало важно изучать демографию места. Какая еда более популярна в местности? У всего населенного пункта любит вегетарианскую еду. Если да, то является ли эта местность населенной определенной сектой людей, например. Джайны, марвари, гуджараты, которые в основном вегетарианцы. Этот вид анализа может быть сделан с использованием данных, изучая различные факторы».

- Пример строки данных (одной из 60 149):

url	https://www.zomato.com/bangalore/caf-eleven-banashankari-bangalore?context=eyJzZSI6eyJIJjpbIjE4MTQ4OTQxIiwMTg0NzA3NzEiLCIxODcxNjA2MSIsIjE4NjMyOTc3IiwxODIwMjk5NywiNTkwOTAiLCIxODU1OTIxMSIsIjE4MzIzNjM5IiwMTg0ODg0ODkiLCIxODUyOTA5MCIsljU2MTY2IiwNjA5MTkiLCIxODM5Njc2MCIsljE4NTkzNTE5IiwMTg1MDAwMjA5IiwMTg0ODg0NDI2MCIsljU3OTk0IiwNTUwOTAiLCI1NzQzNSJdLCJ0IjoiQ2FmXHUwMGU5cyBhbmQgRGVsaXMgaW4gQmFuYXNoYW5rYXJpIn19
address	111, Sapphire Toys Building, 100 Feet Ring Road, Banashankari, Bangalore
name	Caf-Eleven
online_order	No
book_table	No
rate	4.0/5
votes	424
phone	8 049 577 715
location	Banashankari
rest_type	Cafe
dish_liked	Sandwich, Omelette, Ice Tea, Virgin Mojito, Hot Chocolate, Pasta Arrabiata, Hazelnut Cappuccino
cuisines	Cafe, Continental
approx_cost(for two people)	450
reviews_list	[('Rated 2.0', "RATED\n This is a hookah cafe. It was not mentioned anywhere in the page that it was a hookah place.

There's also no separate seating for non smokers. Horrible ambience. Ppl including waiters smoking hookah everywhere. I'm never gonna visit this place again. Also, I found an eyelash in the hot chocolate shake. The hot and cold chocolate are below average.", ('Rated 4.0', "RATED\n Loved the ambience and the interior, very colourful and lively.\nThe seating is spacuous!\nThe food tasted really good\nThe chicken was evenly cooked and was soft\nWe ordered lemon chicken dragon chicken and jalapeno poppers\nThe poppers were oily but otherwise it was good!\nGreat music\nNice sheesha!\nOne disadvantage , the sheesha is only for one hour on the weekends ... We'd like to obviously sit for more time!\nOtherwise everything is great.\nStaff can be a little more well learnt or greetable"), ('Rated 4.0', "RATED\n It's actually a hookah cafe with awesome ambience and good food and a bit slow service , I ordered chicken burger which wasn't very good but the barbecue chicken wings were awesome Its a great place for hanging out with your friends."), ('Rated 3.5', "RATED\n Visited this place on a weekday and only 2 or 3 tables were filled. This place is located in the same building of sapphire toys in the main road of kamakya road.\n\nOrdered KitKat shake, ide to milkshake, onion rings and chilly chicken. Everything was pretty decent.\nChilly chicken was spicy and tasty. Onion rings could have been done more well. Both the milkshake was above average. Sheesha is the main attraction.\n\nAmbience is decent. They've used an old scooter and made it look attractive. Service is prompt and good."), ('Rated 4.0', "RATED\n A hookah lounge!\n\nBakasura makes it a point to mention such info before every review. As some are particular in such ambience.\n\nIt's bakasura Recommended!!\n\nThe food bakasura ordered were fries, nachos, pizza, blueberry Oreo and virgin mojito!\n\nWhen it comes to food. Every single thing was good. The nachos was crazily special. The blueberry Oreo milkshake is something that bakasura Recommends. Pizza was thick crust and very well stuffed.\n\nComing to the ambience. It's rooftop and it's crazily colourful. The yellow chairs and the Christmas tree that was put up. Everything is perfect.\nStaff needs a slight improvement.\nOverall it's great\n\nFollow Bakasura for more such reviews.\n\n#BakasuraForever"), ('Rated 1.0', "RATED\n For the price we pay here worst food I have tasted , steaks are bit ok but not that good. I love pasta to the core but red sauce pasta had lots and lots of basil in it and taste was also not gud .I wouldn't recommend this place."), ('Rated 3.0', "RATED\n Ambience: This cafe is located in banashankari 2nd stage near kamakya theatre. This is basically a cafe on the rooftop which is very busy all the time. The place has nostalgia feels because old scooters are present inside which attracts everyone. Has small LED lights, lamps, dream catchers and plants around which looks cute during evening.\n\nFood : 1) BBQ veg pizza was very big and was good. But wasn't yummy as expected.\n2) Arrabiata pasta,This is a must try. Was really tasty and yummy and creamy.\n3) Jalapeno cheese balls was very cheesy and hot hot

	served with mayo.\n4) Oreo milkshake was okay okay.\n\nService: was really bad."), ('Rated 5.0', 'RATED\n Great hangout spot to spend long time over conversations with friends with delicious food.\n\nMust try: cheese toasts, babycorn pepper dry\n\n#rajeshwrites'), ('Rated 3.0', "RATED\n Good place to hang out with friends. Food wise nothing great. Not advisable to go with family. Even if it's roof and sometimes it's suffocating"), ('Rated 4.0', "RATED\n First thing first, do book a table if you're going here on a weekend. We had to wait for about 20 minutes.\nTalking about the service, I think it has to get slightly better. 3/5\nThough people prefer this place for sheesha, I'd been there just to grab a few snacks. The cafe frappe and mushroom pepper were great! No complains. 4.5/5\nAmbience: 3/5 I wish to go there again, some time during the week."), ('Rated 2.0', 'RATED\n Before cafe eleven was good enough with the food stuffs and all but now staff has been changed and it is not a pocket friendly cafe too and service is too poor ? hookahs are good enough but food taste is up to the mark ? ambience is good enough ??'), ('Rated 3.0', "RATED\n It's a decent rooftop with pretty funky ambience like scooters, colourful seats and green pots on it. Food tastes good! I didn't liked brownie fraape and our order took a lot of time too, just look towards that once. It's pretty good place for hookah lovers. Reasonable price too! Overall one time visit is ok.")]
menu_item	[]
listed_in(type)	Cafes
listed_in(city)	Banashankari

Цены на топливо в Бразилии

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/matheusfreitag/gas-prices-in-brazil>

Приведенное описание набора данных:

«The Data

The National Agency of Petroleum, Natural Gas and Bio fuels (ANP in Portuguese) releases weekly reports of gas, diesel and other fuels prices used in transportation across the country. These datasets bring the mean value per liter, number of gas stations analyzed and other information grouped by regions and states across the country.

Source

As stated before, these datasets are provided by ANP, and are regularly updated with new dates and information - which can be retrieved here (in portuguese).

What can be done with this?

How different regions of Brazil saw their gas prices change?

Within a region, which states increased more their prices?

Which states are the cheapest (or most expensive) for different types of fuels?

Перевод:

«Данные

Национальное агентство нефти, природного газа и биотоплива (ANP на португальском языке) публикует еженедельные отчеты о ценах на газ, дизельное топливо и другие виды топлива, используемые при перевозках по всей стране. Эти наборы данных содержат среднее значение на литр, количество проанализированных заправочных станций и другую информацию, сгруппированную по регионам и штатам по всей стране.

Источник

Как указывалось ранее, эти наборы данных предоставляются ANP и регулярно пополняются новыми датами и информацией, которые можно найти здесь (на португальском языке).

Что можно сделать с этим? Можно ответить на вопросы:

Как в разных регионах Бразилии цены на газ изменились?

В пределах региона, какие государства увеличили свои цены?

Какие государства являются самыми дешевыми (или самыми дорогими) для различных видов топлива?»

- Пример строки данных (одной из 106 824):

DATA INICIAL	09.05.2004
DATA FINAL	15.05.2004
REGIÃO	CENTRO OESTE
ESTADO	DISTRITO FEDERAL
PRODUTO	ETANOL HIDRATADO
NÚMERO DE POSTOS PESQUISADOS	127
UNIDADE DE MEDIDA	R\$/l
PREÇO MÉDIO REVENDA	1,288
DESVIO PADRÃO REVENDA	0,016
PREÇO MÍNIMO REVENDA	1,19
PREÇO MÁXIMO REVENDA	1,35
MARGEM MÉDIA REVENDA	0,463
COEF DE VARIAÇÃO REVENDA	0,012
PREÇO MÉDIO DISTRIBUIÇÃO	0,825
DESVIO PADRÃO DISTRIBUIÇÃO	0,11
PREÇO MÍNIMO DISTRIBUIÇÃO	0,4201
PREÇO MÁXIMO DISTRIBUIÇÃO	0,9666
COEF DE VARIAÇÃO DISTRIBUIÇÃO	0,133
MÊS	5
ANO	2004

Преступления в Бостоне

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/AnalyzeBoston/crimes-in-boston#crime.csv>

Приведенное описание набора данных:

«Context

Crime incident reports are provided by Boston Police Department (BPD) to document the initial details surrounding an incident to which BPD officers respond. This is a dataset containing records from the new crime incident report system, which includes a reduced set of fields focused on capturing the type of incident as well as when and where it occurred.

Content

Records begin in June 14, 2015 and continue to September 3, 2018.

Acknowledgements

The data is provided by Analyze Boston. The most up-to-date version can be found here.

Inspiration

What types of crimes are most common? Where are different types of crimes most likely to occur? Does the frequency of crimes change over the day? Week? Year?»

Перевод:

«Бостонское полицейское управление (BPD) предоставляет отчеты об инцидентах с преступностью, чтобы задокументировать первоначальные подробности инцидента, на которые реагируют сотрудники BPD. Это набор данных, содержащий записи из новой системы отчетов о происшествиях с преступностью, которая включает в себя сокращенный набор полей, предназначенных для регистрации типа происшествия, а также того, когда и где он произошел.

содержание

Записи начинаются 14 июня 2015 года и продолжаются до 3 сентября 2018 года.

Подтверждения

Данные предоставлены Analyze Boston. Самую актуальную версию можно найти здесь.

возможные вопросы

Какие виды преступлений наиболее распространены? Где различные виды преступлений наиболее вероятны? Меняется ли частота преступлений за день? Неделю? Год?»

- Пример строки данных (одной из 106 824):

Incident_number	I182070945
Offense_code	619
Offense_code_group	Larceny
Offense_description	LARCENY ALL OTHERS
District	D14
Reporting_area	808
Shooting	

Occurred_on_date	02.09.2018 13:00:00
Year	2018
Month	9
Day_of_week	Sunday
Hour	13
Ucr_part	Part One
Street	LINCOLN ST
Lat	42.35779134
Long	-71.13937053
Location	"(42.35779134-71.13937053)"

Продажи женской обуви

- URL с данными и приведенным описанием [2]:
<https://www.kaggle.com/datafiniti/womens-shoes-prices/>

«About This Data

This is a list of 10,000 women's shoes and their product information provided by Datafiniti's Product Database.

The dataset includes shoe name, brand, price, and more. Each shoe will have an entry for each price found for it and some shoes may have multiple entries.

Note that this is a sample of a large dataset. The full dataset is available through Datafiniti.

What You Can Do with This Data

You can use this data to determine brand markups, pricing strategies, and trends for luxury shoes. E.g.:

What is the average price of each distinct brand listed?

Which brands have the highest prices?

Which ones have the widest distribution of prices?

Is there a typical price distribution (e.g., normal) across brands or within specific brands?

Further processing data would also let you:

Correlate specific product features with changes in price.

You can cross-reference this data with a sample of our Men's Shoe Prices to see if there are any differences between women's brands and men's brands».

Перевод:

«Об этих данных

Это список из 10000 женской обуви и информации об их продукции, предоставленной базой данных продуктов Datafiniti.

Набор данных включает имя обуви, марку, цену и многое другое. Каждая обувь будет иметь запись для каждой найденной цены, а некоторые ботинки могут иметь несколько записей.

Обратите внимание, что это образец большого набора данных. Полный набор данных доступен через Datafiniti.

Что вы можете сделать с этими данными

Вы можете использовать эти данные для определения наценок бренда, ценовых стратегий и тенденций для роскошной обуви. Например.:

Какова средняя цена каждого отдельного бренда в списке?

Какие бренды имеют самые высокие цены?

Какие из них имеют наибольшее распределение цен?

Существует ли типичное распределение цен (например, нормальное) по брендам или по конкретным брендам?

Дальнейшая обработка данных также позволит вам:

Соотнесите конкретные особенности продукта с изменениями в цене.

Вы можете сопоставить эти данные с образцом наших цен на мужскую обувь, чтобы увидеть, есть ли различия между женскими и мужскими брендами».

- Пример строки данных (одной из 33 802):

id	AVpe__eOilAPnD_xSt-H
asins	
brand	Novica
categories	Access.,Clothing,Shoes,Women's Clothing
colors	Purple
count	
dateAdded	2017-02-03T22:06:24Z
dateUpdated	2017-03-28T11:53:51Z
descriptions	[{"dateSeen":["2017-03-26T05:04:46.539Z","2017-02-03T22:06:24Z"],"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":"This handmade creation is offered in partnership with NOVICA, in association with National Geographic.Peru's Alfredo Falcon designs an attractive purple sweater with bell sleeves. Knitted of a fine alpaca wool blend, lace centers the scoop neck while the sleeves and sweater feature a rolled hem.Product Features:Worldstock Country: PeruColors: PurpleFit: ContemporarySleeve type: RolledNeckline: ScoopUnlinedClosure: PulloverMaterials: 45-percent acrylic/40-percent alpaca/15-percent woolCare instructions: Hand wash in cold water, lay flat to dryThis item has been made with natural fibers that are soft to the touchModel: 231516Story Behind the Art:Alfredo Falcon says, 'When I was but a small child I learned from my parents the best weaving techniques. Like me, they had learned from their own parents. This makes us a family entirely devoted to the world of textiles, and we have been for generations. Nowadays I combine knitting techniques, remaining true to Inca and colonial motifs.'What is Worldstock?The handcrafted touch of artisan skill creates variations in color, size and design. If buying two of the same item, slight differences should be expected. Note: Color discrepancies may occur between this product and your computer screen.ImportedPlease allow 10 business days for the product to leave our warehouse and to receive tracking information. You should expect to receive this item within 15 business days."}]
dimension	
ean	

features	[{"key":"Pattern","value":["Solid"]}, {"key":"Worldstock Country","value":["Peru"]}, {"key":"Neck Style","value":["Scoop Neck"]}, {"key":"Sleeve-length","value":["Long Sleeve"]}, {"key":"Front","value":["Flat Front"]}, {"key":"Care Instruction","value":["Hand Wash"]}, {"key":"Closure","value":["Pull Over"]}, {"key":"Dress Length","value":["Short"]}, {"key":"Size","value":["L, M, S, XL"]}, {"key":"Dimensions","value":["See Details"]}, {"key":"Model Number","value":["231516"]}, {"key":"material","value":["Acrylic, Alpaca Wool"]}]
flavors	
imageURLs	
isbn	
keys	handcraftedalpacablendpurplecharismasweaterperu/19238764,novica/231516,handcraftedalpacablendpurplecharismasweaterperubynovica/19238764
manufacturer	
manufacturerNumber	231516
merchants	
name	Handcrafted Alpaca Blend 'Purple Charisma' Sweater (Peru)
prices.amount Min	62.99
prices.amount Max	62.99
prices.availability	
prices.color	
prices.condition	
prices.count	
prices.currency	USD
prices.dateAdded	2017-03-28T11:53:51Z
prices.dateSeen	2017-03-26T05:04:46.539Z
prices.flavor	
prices.isSale	False
prices.merchant	Overstock.com
prices.offer	
prices.returnPolicy	
prices.shipping	
prices.size	L
prices.source	
prices.sourceURLs	https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html
prices.warranty	
quantities	[{"dateSeen":["2017-03-26T05:04:46.539Z","2017-02-08T07:38:48.141Z"],"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":3}]

reviews	[{"date":"2017-01-10T00:00:00.000Z","dateAdded":"2017-03-06T14:47:43Z","dateSeen":["2017-03-02T18:10:15.094Z"],"rating":5.0,"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/customer-reviews.html"],"text":"well made sweater fit true to size","title":"AS","username":"Antoinette M."}]
sizes	
skus	[{"sourceURLs":["https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html"],"value":"19238764"}]
sourceURLs	https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/product.html,https://www.overstock.com/Worldstock-Fair-Trade/Handcrafted-Alpaca-Blend-Purple-Charisma-Sweater-Peru/12420822/customer-reviews.html
upc	
vin	
websiteIDs	
weight	

Здравоохранение в Ганне

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/citizen-ds-ghana/health-facilities-gh>

«Context

This dataset is provided as part of Citizen Data Science project, to gather & provide fairly clean data (which is a challenge in these regions) to support the data science practice in Ghana and other regions at the beginning of their data science learning curve. So your support is welcome

This dataset provides a listing of healthcare facilities in Ghana, by exploring it we gain new understanding of the country's health infrastructure.

Content

This dataset contains information about health facilities in Ghana organised by Region and District. It also includes the type of health facility and the ownership as well as its geo-location.

Dataset Usecases (are you up to the task? try any of below)

1. Learning/familiarisation with cleaning data and resolving in challenging context of data acquisition.

2. Understanding Ghana's health infrastructure

3. Complex Joint of health facilities and tier data

The health facilities data and the tier data are from different sources but we like to join them because they refer to the same facility however this join may not be a simple join because the health facility name in both datasets are not exact string match.

4. Understanding the level of access to facilities

Combined with population data we want to understand whether some regions or areas are deprived?

Any other creative stuff you can do with this data»

Перевод:

«Контекст

Этот набор данных предоставляется в рамках проекта Citizen Data Science для сбора и предоставления достаточно чистых данных (что является проблемой в этих регионах) для поддержки практики обработки данных в Гане и других регионах в начале их кривой обучения науке о данных. Так что вы поддерживаете приветствуется

Этот набор данных содержит список медицинских учреждений в Гане, изучая его, мы получаем новое понимание инфраструктуры здравоохранения страны. содержание

Этот набор данных содержит информацию о медицинских учреждениях в Гане, организованных по регионам и районам. Он также включает в себя тип медицинского учреждения и владельца, а также его географическое положение.

Варианты использования набора данных (у вас есть задача? Попробуйте любой из нижеприведенных)

1. Изучение / ознакомление с очисткой данных и разрешением в сложных условиях сбора данных.

2. Понимание инфраструктуры здравоохранения Ганы

3. Комплексное соединение медицинских учреждений и уровня данных

Данные о медицинских учреждениях и данные об уровне взяты из разнотных источников, но мы хотели бы объединить их, поскольку они ссылаются на одно и то же учреждение, однако это объединение может быть не простым объединением, поскольку имя медицинского учреждения в обоих наборах данных не является точным совпадением строк.

4. Понимание уровня доступа к объектам

В сочетании с данными о населении мы хотим понять, лишены ли некоторые регионы или районы?

Любые другие творческие вещи, которые вы можете сделать с этими данными».

- Пример строки данных (одной из 3775):

Region	Ashanti
District	Offinso North
FacilityName	A.M.E Zion Clinic
Type	Clinic
Town	Afrancho
Ownership	CHAG
Latitude	7.408010000000001
Longitude	-1.96317

Пиццерии

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/datafiniti/pizza-restaurants-and-the-pizza-they-sell>
«About this Data

This is a list of over 3,500 pizzas from multiple restaurants provided by Datafiniti's Business Database. The dataset includes the category, name, address, city, state, menu information, price range, and more for each pizza restaurant.

Note that this is a sample of a large dataset. The full dataset is available through Datafiniti.

What You Can Do with this Data

You can use this data to discover how much you can expect to pay for pizza across the country. E.g.:

What are the least and most expensive cities for pizza?

What is the number of restaurants serving pizza per capita (100,000 residents) across the U.S.?

What is the median price of a large plain pizza across the U.S.?

Which cities have the most restaurants serving pizza per capita (100,000 residents)?»

Перевод:

«Об этих данных

Это список из более чем 3500 пицц из нескольких ресторанов, предоставленных бизнес-базой данных Datafiniti. Набор данных включает в себя категорию, имя, адрес, город, штат, информацию о меню, ценовой диапазон и многое другое для каждой пиццерии.

Обратите внимание, что это образец большого набора данных. Полный набор данных доступен через Datafiniti.

Что вы можете сделать с этими данными

Вы можете использовать эти данные, чтобы узнать, сколько вы можете заплатить за пиццу по всей стране. Например.:

Какие города наименее и самые дорогие для пиццы?

Каково количество ресторанов, где подают пиццу на душу населения (100 000 жителей) по всей территории США?

Какова средняя цена большой обычной пиццы в США?

В каких городах больше всего пиццы на душу населения (100 000 жителей)?».

- Пример строки данных (одной из 10001):

id	AVz3Y-7h3D1zeR_xDAqm
dateAdded	2017-06-30T05:05:40Z
dateUpdated	2019-05-01T15:43:09Z
address	4203 E Kiehl Ave
categories	Pizza,Restaurant,American restaurants,Pizza Place,Restaurants
primaryCategories	Accommodation & Food Services
city	Sherwood
country	US
keys	us/ar/sherwood/4203ekiehlave/-1051391616
latitude	34.8323
longitude	-92.1838

menuPageURL	http://www.citysearch.com/profile/menu/1550074?singlePlatformId=shotgun-dans-pizza
menus.amountMax	18.21
menus.amountMin	18.21
menus.currency	USD
menus.dateSeen	2018-05-01T04:25:37.197Z,2018-04-16T04:36:02.356Z,2018-02-15T19:58:01.612Z,2018-04-02T23:29:46.353Z,2018-06-28T11:37:25.942Z
menus.description	
menus.name	Cheese Pizza
name	Shotgun Dans Pizza
postalCode	72120
priceRangeCurrency	USD
priceRangeMin	0
priceRangeMax	25
province	AR

Вина

- URL с данными и приведенным описанием [2]:
<https://www.kaggle.com/zynicide/wine-reviews>

«Inspiration

I think that this dataset offers some great opportunities for sentiment analysis and other text related predictive models. My overall goal is to create a model that can identify the variety, winery, and location of a wine based on a description. If anyone has any ideas, breakthroughs, or other interesting insights/models please post them.»

Перевод:

«Я думаю, что этот набор данных предлагает большие возможности для анализа настроений и других прогнозирующих моделей, связанных с текстом. Моя общая цель - создать модель, которая могла бы определять сорт, винодельню и местоположение вина на основе описания».

- Пример строки данных (одной из 129927):

country	US
description	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.
designation	Reserve Late Harvest
points	87
price	14.0
province	Oregon
region_1	Willamette Valley
region_2	Willamette Valley
taster_name	Paul Gregutt

taster_twitter_handle	@paulgwineB
title	Rainstorm 2013 Pinot Gris (Willamette Valley)
variety	Pinot Gris
winery	Rainstorm

Аренда жилья в Берлине

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/brittabetendorf/berlin-airbnb-data>

«Context

Airbnb has successfully disrupted the traditional hospitality industry as more and more travelers decide to use Airbnb as their primary accommodation provider. Since its beginning in 2008, Airbnb has seen an enormous growth, with the number of rentals listed on its website growing exponentially each year. In Germany, no city is more popular than Berlin. That implies that Berlin is one of the hottest markets for Airbnb in Europe, with over 22,552 listings as of November 2018.

Content

The datasets were scraped on November 07th, 2018 and contain detailed listings data, review data and calendar data of current Airbnb listings in Berlin.

Acknowledgements

This data was created by Murray Cox and his Inside Airbnb project which can be found here.

Inspiration

Can you predict the price for each Berlin neighborhood using listing descriptions? What are the busiest times of the year to visit Berlin? By how much do prices spike? Can you uncover trends in reviews of Airbnb visitors to Berlin?»

Перевод:

«Контекст

Airbnb успешно разрушил традиционную индустрию гостеприимства, так как все больше и больше путешественников решают использовать Airbnb в качестве основного поставщика жилья. С начала своего существования в 2008 году Airbnb пережил огромный рост, причем количество арендных плат, перечисленных на его веб-сайте, растет в геометрической прогрессии с каждым годом. В Германии нет более популярного города, чем Берлин. Это означает, что Берлин является одним из самых популярных рынков для Airbnb в Европе, с 22,552 списками на ноябрь 2018 года.

содержание

Наборы данных были очищены 7 ноября 2018 года и содержат подробные данные о списках, обзорные данные и календарные данные о текущих списках Airbnb в Берлине.

Подтверждения

Эти данные были созданы Мюрреем Коксом и его проектом Inside Airbnb, которую можно найти здесь.

возможные вопросы

Можете ли вы предсказать цену для каждого района Берлина, используя описания списков?

Какое время года для посещения Берлина? На сколько цены растут?

Можете ли вы раскрыть тенденции в обзорах посетителей Airbnb в Берлине?».

- Пример строки данных (одной из 401964):

listing_id	2015
id	69544350
date	11.04.2016
reviewer_id	7178145
reviewer_name	Rahel
comments	Mein Freund und ich hatten gute gemГјtliche vier NГјchte in Jans Studio. Es ist unfassbar ruhig und hat sogar ein gedГјmpftes Klavier fГјr Leute die gerne spielen. Die Lage ist hervorragend und es gibt in der Strasse einen tollen alimentari und eine гјnstige weinerei. Das Bad ist sehr alt aber funktionstГјchtig und der insgesamte Standard ist ok, ein wenig chaotisch aber dafГјr hat man das GefГјhl man ist bei jemandem zuhause ! Jan ist ein aufgeschlossener lustiger Mann und auch um frische HandГјcher und sonstiges bemГјht! Wir danken.

Японский фаст-фуд

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/residentmario/ramen-ratings>

«Context

The Ramen Rater is a product review website for the hardcore ramen enthusiast (or "ramenphile"), with over 2500 reviews to date. This dataset is an export of "The Big List" (of reviews), converted to a CSV format.

Content

Each record in the dataset is a single ramen product review. Review numbers are contiguous: more recently reviewed ramen varieties have higher numbers. Brand, Variety (the product name), Country, and Style (Cup? Bowl? Tray?) are pretty self-explanatory. Stars indicate the ramen quality, as assessed by the reviewer, on a 5-point scale; this is the most important column in the dataset!

Note that this dataset does *not* include the text of the reviews themselves. For that, you should browse through <https://www.theramenrater.com/> instead!

Acknowledgements

This dataset is republished as-is from the original BIG LIST on <https://www.theramenrater.com/>.

Inspiration

What ingredients or flavors are most commonly advertised on ramen package labels? How do ramen ratings compare against ratings for other food products (like, say, wine)?

How is ramen manufacturing internationally distributed?»

Перевод:

«Контекст

Ramen Rater - это веб-сайт с обзором продуктов для хардкорных энтузиастов рамен (или «ramenphile»), на сегодняшний день более 2500 обзоров. Этот набор данных является экспортом «Большого списка» (обзоров), преобразованного в формат CSV.

Содержание:

Каждая запись в наборе данных представляет собой отдельный обзор продукта рамэн. Числа обзора являются смежными: более поздние исследованные сорта рамэна имеют более высокие числа. Бренд, Разнообразие (название продукта), Страна и Стил (Кубок? Чаша? Поднос?) Говорят сами за себя. Звезды показывают качество рамэна, оцениваемое рецензентом, по 5-балльной шкале; это самый важный столбец в наборе данных!

Обратите внимание, что этот набор данных не включает в себя текст самих обзоров. Для этого вы должны просмотреть <https://www.theramenrater.com/> вместо этого!

Подтверждения

Этот набор данных переиздан как есть из оригинального БОЛЬШОГО СПИСКА на <https://www.theramenrater.com/>.

Возможные вопросы:

Какие ингредиенты или ароматизаторы чаще всего рекламируются на этикетках упаковки рамэн?

Как рейтинги рамэн сравниваются с оценками других продуктов питания (например, вина)?

Как производство рамэна распределяется на международном уровне?»

- Пример строки данных (одной из 2581):

Review #	1350
Brand	Mamee
Variety	Chef Curry Laksa Flavour
Style	Pack
Country	Malaysia
Stars	5
Top Ten	2014 #7

Игроки FIFA-18

- URL с данными и приведенным описанием [2]:

<https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset>

«Possible Explorations

Make your dream team

Analyse which Club or National Team has the best-rated players

Assess the strength of a team at a particular position

Analyse the team with the best dribbling speed

Co-relate between Age and Overall rating

Co-relate between Age and Nationality

Co-relate between Age and Potential»

Перевод:

«Возможные исследования

Сделай команду своей мечты

Проанализируйте, в каком клубе или сборной лучшие игроки

Оценить силу команды на определенной позиции

Проанализируйте команду с лучшей скоростью дриблинга

Сопоставить возраст и общий рейтинг

Взаимосвязь между возрастом и национальностью

Взаимосвязь между возрастом и потенциалом»

- Пример строки данных (одной из 17 982):

Name	Cristiano Ronaldo
Age	32
Photo	https://cdn.sofifa.org/48/18/players/20801.png
Nationality	Portugal
Flag	https://cdn.sofifa.org/flags/38.png
Overall	94
Potential	94
Club	Real Madrid CF
Club Logo	https://cdn.sofifa.org/24/18/teams/243.png
Value	В, -95.5M
Wage	В, -565K
Special	2228
Acceleration	89
Aggression	63
Agility	89
Balance	63
Ball control	93
Composure	95
Crossing	85
Curve	81
Dribbling	91
Finishing	94
Free kick accuracy	76
GK diving	7
GK handling	11
GK kicking	15
GK positioning	14
GK reflexes	11
Heading accuracy	88
Interceptions	29
Jumping	95
Long passing	77
Long shots	92
Marking	22
Penalties	85
Positioning	95
Reactions	96
Short passing	83
Shot power	94

Sliding tackle		23
Sprint speed		91
Stamina		92
Standing tackle		31
Strength		80
Vision		85
Volleys		88
CAM	89.0	
CB	53.0	
CDM	62.0	
CF	91.0	
CM	82.0	
ID		20801
LAM	89.0	
LB	61.0	
LCB	53.0	
LCM	82.0	
LDM	62.0	
LF	91.0	
LM	89.0	
LS	92.0	
LW	91.0	
LWB	66.0	
Preferred Positions	ST LW	
RAM	89.0	
RB	61.0	
RCB	53.0	
RCM	82.0	
RDM	62.0	
RF	91.0	
RM	89.0	
RS	92.0	
RW	91.0	
RWB	66.0	
ST	92.0	

Список используемых источников

1. Хабр. <https://habr.com/ru/post/248395/>
2. Kaggle. <https://www.kaggle.com>
3. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. – Ульяновск : УлГТУ, 2017. – 290 с.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»**

А.А. Филиппов
И.А. Андреев

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНЫХ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ**
лабораторный практикум по дисциплине
«Интеллектуальные информационно-аналитические системы»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Филиппов, Алексей Александрович
Андреев, Илья Алексеевич

Разработка интеллектуальных информационно-аналитических систем : лабораторный практикум / А. А. Филиппов, И.А. Андреев. – Ульяновск : УлГТУ, 2021. – 11 с.

Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Интеллектуальные информационно-аналитические системы». Предоставлены задания, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины.

Предназначен для студентов, обучающихся по направлениям: 09.04.03 «Прикладная информатика» (программа «Искусственный интеллект и бизнес-аналитика»).

Работа подготовлена на кафедре «Информационные системы».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	4
Экзамен	5
Лабораторная работа №1	6
Контрольные вопросы к лабораторной работе	6
Лабораторная работа №2	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №3	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №4	9
Контрольные вопросы к лабораторной работе	9
Список используемой литературы	10

ВВЕДЕНИЕ

Целью освоения дисциплины «Интеллектуальные информационно-аналитические системы» является формирование представления об этапах жизненного цикла интеллектуальной информационно-аналитической системы, а также о способах и методах управления данным жизненным циклом в целом и на каждом этапе в частности. В ходе изучения дисциплины предполагается приобретение обучающимися прочных знаний методов и алгоритмов машинного обучения, а также способов их внедрения для решения задач систем искусственных информационно-аналитических систем.

Задачами дисциплины являются:

- изучение принципов, заложенных в разработку интеллектуальных информационно-аналитических систем
- формирование навыков интеграции различных методов искусственного интеллекта в программное обеспечение
- приобретение теоретических знаний и практических навыков разработки информационной системы с нуля, а также настройка созданного ПО для работы в выбранной среде.

Кроме того, в результате изучения дисциплины «Интеллектуальные информационно-аналитические системы» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне их формирования.

Тематический план дисциплины

1. Технологии сбора и хранения данных в ИАС
 - 1.1 Технологии извлечения, преобразования и загрузки данных
 - 1.2 Концепции организации хранения данных
2. Архитектура ИАС
 - 2.1 OLAP-системы
 - 2.2 Задачи Data mining
 - 2.3 Модели данных информационного хранилища
3. Применение методов искусственного интеллекта в ИАС
 - 2.1 Нечеткая логика и нечеткие множества в задачах ИАС
 - 2.2 Применение генетических алгоритмов
 - 2.3 Интеграция экспертных систем

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки освоения студентом всех запланированных по дисциплине «Интеллектуальные информационно-аналитические системы» компетенций. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений.

Шкала оценивания имеет вид (таблица П1)

Таблица П1

Количество правильных ответов /Процент правильных ответов	Балл
Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку.	Сдано
Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.	Не сдано

Экзамен

Экзамен по дисциплине проводится в форме ответа на теоретические вопросы билета. Билет содержит два теоретических вопроса, обеспечивающих контроль освоения умений и навыков всех запланированных в ходе изучения дисциплины компетенций. Билет формируется таким образом, чтобы в него комплексные вопросы, контролируемые уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица П2)

Таблица П2

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Выставляется обучающемуся, если студент полностью ответил на оба вопроса билета и способен обосновать свой ответ	Отлично
Выставляется обучающемуся, если студент ответил на оба вопроса, но с некоторыми погрешностями и ошибками или неспособностью обосновать свой ответ	Хорошо
Выставляется обучающемуся, если студент ответил полностью хотя бы на один вопрос, а на второй не смог ответить или ответил с сильными погрешностями и ошибками	Удовлетворительно
Выставляется обучающемуся, если студент не смог ответить ни на один вопрос	Неудовлетворительно

Лабораторная работа №1

Планирование временных затрат на реализацию ИАС.

Необходимо:

1. Оценить все основные виды жизненных циклов и определить какой жизненный цикл информационной системы больше всего подходит для разработки интеллектуальной информационно-аналитической системы.
2. Адаптировать выбранный жизненный цикл под текущую задачу и построить его видоизмененную модель.

Контрольные вопросы к лабораторной работе

1. Что такое жизненный цикл информационной системы?
2. Перечислите основные виды жизненных циклов информационных систем.
3. Назовите основные составляющие архитектуры информационно-аналитической системы.
4. Перечислите основные преимущества и недостатки разработки кроссплатформенных информационно-аналитических систем.

Лабораторная работа №2

Прототипирование системы хранения данных ИАС.

Необходимо:

- 1) Оценить и выбрать СУБД, которая будет использоваться для работы интеллектуальной информационно-аналитической системы.
- 2) Написать скрипт (может быть реализован в виде подпрограммы - системы инициализации разрабатываемого ПО). Скрипт должен в автоматизированном режиме скачивать, устанавливать и настраивать выбранную СУБД, а также создавать базы данных и таблицы для дальнейшей работы ИАС. Предусмотреть возможность работы СУБД внутри локальной вычислительной сети.
- 3) Иметь возможность подключиться к установленной СУБД клиентским приложением в локальной сети, отличным от ИАС для просмотра данных и управления СУБД.

Контрольные вопросы к лабораторной работе

1. Перечислите виды СУБД по применяемым моделям данных.
2. Расскажите об архитектуре ANSI-SPARC.
3. Что такое локальная вычислительная сеть?
4. Перечислите основные особенности настройки и работы СУБД в вычислительной локальной сети.

Лабораторная работа №3

Прототипирование подсистемы извлечения данных, используемых ИАС.

Необходимо:

- 1) Реализовать извлечение данных, которые будут использованы для анализа из любой социальной сети, используя API данной соц. сети.
- 2) Реализовать ПО, организующее хранение данных в таблицах БД, установленной и настроенной в лабораторной работе №2. Предусмотреть окно настроек ПО, которое будет содержать в себе интерфейс конфигурирования подключения к выбранной БД в локальной вычислительной сети.

Контрольные вопросы к лабораторной работе

1. Что такое API?
2. Назовите основные виды API, их назначение и различия.
3. Что такое RPC?
4. Что такое SOAP?
5. Что такое REST?

Лабораторная работа №4

Прототипирование ИАС с использованием методов искусственного интеллекта
Необходимо:

- 1) Разработать ПО, реализующее обработку накопленных данных при помощи одного из методов искусственного интеллекта. ПО должно содержать в себе гибкий графический интерфейс для использования конечным пользователем. Предусмотреть возможность работы с некоторой выборкой данных (фильтрация данных).
- 2) Подготовить сценарий (пакет) установки ПО с нуля, интегрировав в него результаты лабораторных работ №2 и №3.

Результатом лабораторной работы является полностью рабочая система, способная установиться в автоматизированном режиме на любой персональный компьютер, соответствующий требованиям.

Контрольные вопросы к лабораторной работе

1. Дайте определение интеллектуальной информационно-аналитической системы.
2. Что такое сценарий установки программного обеспечения?
3. Перечислите основные методы искусственного интеллекта.
4. Расскажите о способах хранения нечетких данных в СУБД.
5. Расскажите о способах хранения результатов работы генетических алгоритмов в СУБД.
6. Расскажите о способах обработки данных, хранящихся в СУБД, при помощи Data Mining.

Список используемой литературы

1. Прокопенко, Н. Ю. Аналитические информационные системы поддержки принятия решений [Электронный ресурс] : учебное пособие / Н. Ю. Прокопенко. — Нижний Новгород : ННГАСУ, 2020. — 142 с. — ISBN 978-5-528-00395-5.
<https://e.lanbook.com/book/164866>
2. Боровская, Е. В. Основы искусственного интеллекта [Электронный ресурс] : учебное пособие / Е. В. Боровская, Н. А. Давыдова. — 4-е изд. — Москва : Лаборатория знаний, 2020. — 130 с. — ISBN 978-5-00101-908-4.
<https://e.lanbook.com/book/151502>
3. Рочев, К. В. Информационные технологии. Анализ и проектирование информационных систем [Электронный ресурс] : учебное пособие / К. В. Рочев. — 2-е изд., испр. — Санкт-Петербург : Лань, 2019. — 128 с. — ISBN 978-5-8114-3801-3.
<https://e.lanbook.com/book/122181>
4. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных [Электронный ресурс] / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : УлГТУ, 2020. — 220 с. — ISBN 978-5-9795-2088-9.
<https://e.lanbook.com/book/170653>
5. Сергеев, А. Н. Основы локальных компьютерных сетей [Электронный ресурс] : учебное пособие для вузов / А. Н. Сергеев. — 3-е изд., стер. — Санкт-Петербург : Лань, 2021. — 184 с. — ISBN 978-5-8114-6855-3.
<https://e.lanbook.com/book/152651>
6. Остроух, А. В. Системы искусственного интеллекта [Электронный ресурс] : монография / А. В. Остроух, Н. Е. Суркова. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 228 с. — ISBN 978-5-8114-8519-2.
<https://e.lanbook.com/book/176662>
7. Материалы сайта API ВКонтакте | Разработчикам
https://vk.com/dev.php?method=first_guide
8. Материалы сайта Facebook API Graph
<https://developers.facebook.com/docs/graph-api/overview>
9. Материалы сайта Введение в JSON
<https://www.json.org/json-ru.html>
10. Материалы сайта OpenNet
<http://www.opennet.ru/>

**РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНЫХ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИХ СИСТЕМ
лабораторный практикум по выполнению
лабораторных работ по дисциплине
«Интеллектуальные информационно-аналитические системы»**

Авторы

Филиппов Алексей Александрович

Андреев Илья Алексеевич

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»**

Г.Ю. Гуськов

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ
АНАЛИЗА БОЛЬШИХ ДАННЫХ**
лабораторный практикум по дисциплине
«Обработка больших данных в бизнес-аналитике»

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Гуськов, Глеб Юрьевич

Проектирование и разработка информационных систем анализа больших данных: лабораторный практикум / Г.Ю. Гуськов. – Ульяновск : УлГТУ, 2021. – 12 с.

Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Обработка больших данных в бизнес-аналитике». Предоставлены задания, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины.

Предназначен для студентов, обучающихся по направлениям: 09.04.03 «Прикладная информатика» (программа «Искусственный интеллект и бизнес-аналитика»)

Работа подготовлена на кафедре «Информационные системы».

© Гуськов Г.Ю., 2021

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	5
Экзамен	5
Лабораторная работа №1	6
Контрольные вопросы к лабораторной работе	6
Лабораторная работа №2	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №3	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №4	9
Контрольные вопросы к лабораторной работе	9
Лабораторная работа №5	10
Контрольные вопросы к лабораторной работе	10
Список используемой литературы	11

ВВЕДЕНИЕ

Целью освоения дисциплины «Обработка больших данных в бизнес-аналитике» является формирование у будущих выпускников компетенций, связанных с использованием методов, алгоритмов, программных и технических средств реализации и использования прикладных интеллектуальных технологий обработки и анализа больших данных для бизнес-аналитики.

Основные навыки и теоретические знания, приобретаемыми в ходе освоения дисциплины: исследование данных генерируемых в ходе процессов проходящих в системе, использование специальных подходов к обработке и анализу больших объемов данных, исследование прикладных информационных процессов и данных в условиях развития цифровой экономики для получения дескриптивной и предикативной информации с помощью методов интеллектуального анализа, использование и разработка информационно-аналитических методов анализа и оценки эффективности прикладных информационных процессов на основе интеллектуального анализа данных и процессов, использование и разработка программных реализаций информационно-аналитических методов для исследования прикладных и информационных процессов цифровой экономики.

Кроме того, в результате изучения дисциплины «Обработка больших данных в бизнес-аналитике» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне их формирования.

Тематический план дисциплины

Раздел 1. Экосистема цифровой экономики (ЦЭ).

Основные понятия и задачи анализа данных в цифровой экономике. Экосистема цифровой экономики (ЦЭ). Направления развития цифровой экономики и цифровой трансформации в России и за рубежом. Основные понятия, виды и задачи аналитики в цифровой экономике. Методы и подходы решения информационно-аналитических задач для цифровой экономики.

Раздел 2. Методы предиктивной аналитики ЦЭ

Методы классификации данных. Примеры и программные средства систем анализа данных. Стандарты в области разработки систем интеллектуального анализа данных. Методы прогнозирования временных рядов в статистическом подходе. Методы прогнозирования временных рядов в нечетком подходе. Прогнозирование на основе искусственных нейронных сетей.

Раздел 3. Основные понятия анализа бизнес-процессов

Основные понятия анализа бизнес-процессов. Виды зависимостей в показателях бизнес-процессов. Представление о временных рядах. Подходы к прогнозированию временных рядов.

Раздел 4. Методы обработки и анализа больших данных.

Объём, скорость накопления и разнообразие данных. Ценность данных. Особенности хранения, индексирования и анализа больших данных.

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки уровня знаний, умений, владений, понимания студентом основных методов и методик проектирования, разработки и поддержки информационных систем в задачах анализа бизнес-процессов, умения применять на практике полученных знаний. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений. Шкала оценивания имеет вид (таблица П1)

Таблица П1

Количество правильных ответов /Процент правильных ответов	Балл
Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, в конце занятия студент выдает законченную и полностью функционирующую разработку.	Сдано
Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.	Не сдано

Экзамен

Экзамен по дисциплине проводится в форме ответа на теоретические вопросы билета. Билет содержит два теоретических вопроса, обеспечивающих контроль освоения умений и навыков всех запланированных в ходе изучения дисциплины компетенций. Билет формируется таким образом, чтобы в него комплексные вопросы, контролируемые уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица П2)

Таблица П2

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Выставляется обучающемуся, если студент полностью ответил на оба вопроса билета и способен обосновать свой ответ	Отлично
Выставляется обучающемуся, если студент ответил на оба вопроса, но с некоторыми погрешностями и ошибками или неспособностью обосновать свой ответ	Хорошо
Выставляется обучающемуся, если студент ответил полностью хотя бы на один вопрос, а на второй не смог ответить или ответил с сильными погрешностями и ошибками	Удовлетворительно
Выставляется обучающемуся, если студент не смог ответить ни на один вопрос	Неудовлетворительно

Лабораторная работа №1

Изучение методов разработки и применения методов интеллектуального анализа данных на примере научных работ зарубежных исследователей. По предметной области необходимо произвести процесс моделирования объекта:

1. Выбрать объект моделирования в предметной области.
2. Сформировать модель.
3. Построить гипотезы и повести анализ модели.

Контрольные вопросы к лабораторной работе

1. Приведите определение модели, моделирования и охарактеризуйте этапы моделирования. Приведите примеры.
2. Приведите графическую схему процесса моделирования применительно к задачам дескриптивного анализа. Приведите примеры.
3. Приведите графическую схему процесса моделирования применительно к задачам предиктивного анализа. Приведите примеры.
4. Охарактеризуйте задачу построения оптимизационной модели. Приведите примеры.
5. Что понимают под объектом и контекстом моделирования. Приведите примеры.
6. Опишите выбранный объект и контекст моделирования.

Лабораторная работа №2

Выбор объекта, вида и метода его аналитики. Разработка и демонстрация программной системы, реализующей метод интеллектуального анализа данных выбранного объекта.

Необходимо:

- 1) Описать объект в виде модели несколькими способами.
- 2) Произвести интеллектуальный анализ используя методы разобранные в ЛР1.
- 3) Построить результаты анализа в виде диаграмм.

Контрольные вопросы к лабораторной работе

1. Опишите общую методику и задачи построения моделей.
2. Приведите виды и классы моделей. Приведите примеры.
3. Приведите выражения для оценки качества моделей.
4. Какие предположения и гипотезы использованы для выбранного объекта моделирования?
5. Приведите обоснование выбранного метода моделирования для выбранного объекта моделирования.
6. Охарактеризуйте статистические методы моделирования. Приведите примеры.
7. Охарактеризуйте нечеткие методы моделирования. Приведите примеры.

Лабораторная работа №3

Модификация и демонстрация разработанной программной системы для получения лингвистического резюмирования результатов анализа выбранного объекта.

Необходимо:

- 1) Разработать информационную систему, реализующую моделирование выбранного объекта.
- 2) Провести тестирование разработанной информационной системы на основе реальных данных поведения объекта.

Контрольные вопросы к лабораторной работе

1. Опишите ход построения и структуру полученной модели для выбранного объекта.
2. Приведите проблемы моделирования, которые требовалось решить.
3. Охарактеризуйте методы и модели распознавания в обработке данных. Приведите примеры.
4. Опишите сущность и основные понятия методов обработки сигналов. Приведите примеры.

Лабораторная работа №4

Изучение методов предиктивной аналитики на основе временных рядов.

Необходимо:

- 1) Выделить показатели объекта, которые можно представить в виде временного ряда.
- 2) Реализовать один из методов прогнозирования вручную.
- 3) Использовать три метода прогнозирования временных рядов используя подключаемую библиотеку.

Контрольные вопросы к лабораторной работе

1. Приведите общую методику оценки качества моделей.
2. Опишите процесс тестирования построенной модели (моделей).
3. Приведите используемые критерии качества для оценки построенной моделей.
4. Какие предположения и гипотезы подтвердились в процессе моделирования для выбранного объекта моделирования?
5. Сформулируйте выводы по результатам моделирования.
6. Охарактеризуйте методы оптимизации.
7. Приведите используемые методы оптимизации при моделировании выбранного объекта

Лабораторная работа №5

Проведение анализа данных: поиск скрытых зависимостей в данных.

Необходимо:

- 1) Найти в открытых источниках data set по предметной области по варианту.
- 2) Исследовать data set и предобработать данные для анализа.
- 3) Провести анализ данных и выделить зависимости в данных.

Контрольные вопросы к лабораторной работе

1. Что такое коэффициент корреляции Пирсона?
2. Какие функции характеризующие выборку по отдельно взятому показателю вы знаете?
3. Какие скрытые зависимости удалось найти в выборке? Как вы можете их объяснить?
4. Как можно применить ИНС в анализе больших данных?
5. Зачем производить предобработку данных перед анализом?

Список используемой литературы

1. Советов, Б. Я. Информационные технологии: теоретические основы : учебное пособие / Б. Я. Советов, В. В. Цехановский. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 444 с. — ISBN 978-5-8114-1912-8.

URL: <https://e.lanbook.com/book/167404> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

2. Прасолов, А. В. Математические методы экономической динамики : учебное пособие / А. В. Прасолов. — 2-е изд., испр. — Санкт-Петербург : Лань, 2021. — 352 с. — ISBN 978-5-8114-0797-2.

URL: <https://e.lanbook.com/book/168869> (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ
АНАЛИЗА БОЛЬШИХ ДАННЫХ
лабораторный практикум по дисциплине
«Обработка больших данных в бизнес-аналитике»
Гуськов Глеб Юрьевич
УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.**

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Интеллектуальные информационные системы

Профиль подготовки
09.04.03 Прикладная информатика
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

Раздел 1. Лабораторные работы

Лабораторная работа №1 Основы языка Пролог. Организация вычислений

1. Основные сведения о языке

Структура программы

Программа на Прологе состоит из следующих разделов.

Constants	Объявление констант
Global Domains	Объявление глобальных типов.
Domains	Объявление типов.
Global DataBase	Объявление глобальных предикатов для работы с базой данных.
DataBase	Объявление предикатов для работы с базой данных.
Global Predicates	Объявление глобальных предикатов.
Predicates	Объявление предикатов
Goal	Цель программы
Clauses	Предложения (факты или правила)

Наличие всех разделов в программе необязательно. В частности, программа может состоять только из одного раздела. Обычно разделы включаются в программу по мере необходимости и могут повторяться. Но при этом должны выполняться следующие ограничения:

- Раздел программы должен быть представлен соответствующим ключевым словом (**Domains, DataBase, Predicates, Clauses** или **Goal**).
- Раздел **Goal** может встретиться только один раз
- Все предложения, которые описывают один и тот же самый предикат должны следовать одно за другим.
- Раздел **Global Predicates** может встретиться только однажды до объявления раздела (разделов) **Predicates**.
- Разделы, содержащие предикаты для работы с базой данных должны быть объявлены перед всеми глобальными и обычными объявлениями предикатов.

Комментарии

В Прологе имеется два способа написания комментариев.

Первый – однострочный. При этом строка должна начинаться со знака процента - **%**. При таком способе комментарием считается часть строки от знака процента до окончания строки.

Второй – многострочный. В этом случае комментарии начинаются с символов **/*** и заканчиваются символами ***/**. Между этими символами может быть произвольное количество строк текста.

Стандартные типы данных

Имеется шесть стандартных типов:

1. Тип **char** представляет собой одиночный символ, определяемый стандартным 8-и битовым ASCII-кодом. При записи в программе переменная или константа этого типа заключается в апострофы (одиночные кавычки). Символы можно задавать также значением ASCII-кода, предваряемыми слэш-символом (****). Для некоторых символов предусмотрена специальная нотация. Например, **\n** и **\t** соответствуют соответственно символам перехода на новую строку и символу табуляции. Символ **** в сочетании с любым другим символом производит сам символ.

2. Тип **integer** принадлежит к типу целых чисел и представляет собой все целые числа в диапазоне от -32768 до +32767.

3. Тип **real** относится к числам вещественного типа и включает в себя числа в диапазоне

от +/-1e-307 до +/-1e+308. Числа могут записываться в обычной и экспоненциальной форме. Во втором случае число представлено в виде дробной мантиссы и экспоненты в соответствующей степени (все без включения пробелов). При необходимости Пролог автоматически преобразует целочисленный тип в вещественный формат.

4. Тип **string** принадлежит строковому типу (это любая последовательность символов заключенных в двойные кавычки). Строка может содержать символы, произведенные ESC-последовательностью так, как это упомянуто выше в случае типа **'char'**.

5. Тип **symbol** принадлежит к символическому типу (его имя должно начинаться с символов нижнего регистра). Строки этого типа воспринимаются также как последовательности символов, но при этом символы хранятся в специальной таблице поиска для более быстрой идентификации. Такая таблица занимает очень мало памяти. Кроме того, время для работы с ней - минимально.

6. Тип **symbolic filename** принадлежит к файловому типу (его имя должно начинаться или с символа нижнего регистра и появляться справа от объявления файлового типа или с одного из предопределенных символических файловых имен: - **'printer', 'screen', 'keyboard'** или **'com1'**).

Имена переменных

Для обозначения переменной ей присваивается специальная символьная метка, называемая именем (идентификатором) переменной. В Прологе требования к идентификаторам практически аналогичны требованиям в других языках программирования:

а) именем переменной может быть любая комбинация символов латинского алфавита, символов подчеркивания и цифр, при следующих ограничениях:

б) имя переменной не должно начинаться с цифры;

в) имя переменной должно начинаться с символа верхнего регистра или символа подчеркивания.

Анонимные переменные

При формировании запроса к системе или оформлении программного текста возможны ситуации, когда нас не интересует значение какой-либо переменной, но вследствие синтаксиса предиката эту переменную указать необходимо. В таких случаях используются анонимные переменные. Для их обозначения применяется знак символа подчеркивания.

Предикаты ввода/вывода

Использование предикатов ввода/вывода позволяет вести активный диалог с программой.

Для вывода информации Пролог располагает практически единственным предикатом - **Write**. Предикат может вызываться с произвольным числом параметров:

Write(Arg1,Arg2,Arg3,)

Выводимые параметры могут быть или константами (стандартного типа) или переменными.

Предикат **Write** обычно используется в связке со стандартным предикатом **nl**, который обеспечивает перевод курсора на новую строку.

В Прологе имеются стандартные предикаты для чтения простых переменных различного типа. Самостоятельно эти предикаты не могут быть использованы для чтения составных объектов или списков. Организация ввода подобных объектов - обязанность программиста. Ниже приводится перечень предикатов, которые можно использовать для ввода.

1. Readln(Line)

Предикат позволяет считывать с текущего входного устройства строки длиной до 150 символов. Вводимая строка должна оканчиваться символом возврата каретки. Тип переменной **Line** должен быть либо **String** либо **Symbol**. Переменная должна быть свободной до вызова.

2. ReadInt(X)

Предикат считывает символы с текущего входного устройства до нажатия Enter. После чтения строки она будет преобразована в число целого типа. Переменная **X** должна быть целого типа и свободной до вызова. Если строка не соответствует синтаксису целого типа, то предикат

терпит неудачу и включается механизм перебора с возвратом.

3. ReadReal(X)

Предикат считывает символы с текущего входного устройства до нажатия Enter. По окончании ввода полученная строка обращается в вещественное число. Если строка не соответствует синтаксису **Real**, то предикат терпит неудачу. Переменная **X** должна быть вещественного типа и свободной до вызова.

4. File_Str(Filename, X)

Предикат считывает содержимое файла **Filename** и преобразует его в переменную **X**. Считывание производится до появления признака конца файла (обычно **Ctrl+Z**). Переменные **Filename** и **X** должны быть типа **Symbol** или **String**. Переменная **X** не должна быть связана до вызова. Длина считываемого файла может быть до 64кб.

5. ReadChar(CharParam)

Предикат читает одиночный символ из текущего устройства ввода и тут же завершает свою работу. В результате считанный символ связывается с переменной **CharParam**. Эта переменная должна быть соответствующего типа и свободной до вызова.

Математические возможности Пролога

В Прологе имеется полный набор арифметических функций, стандартных предикатов преобразования, тригонометрических функций и семейство предикатов, обеспечивающих по-рядные логические операции.

Арифметические операции

В Прологе может выполнять все основные арифметические операции (сложение, вычитание, умножение и деление) над целыми и вещественными числами. Тип получаемого при этом результата определяется согласно следующей таблице.

Операнд1	Операция	Операнд2	Тип результата
integer	+, -, *	integer	integer
integer	+, -, *	real	real
real	+, -, *	integer	real
real	+, -, *	real	real
integer или real	/	integer или real	real

Математические выражения также вычисляются стандартным образом - слева направо, с учетом приоритета операции и наличия круглых скобок. Приоритет операций приведен в таблице.

Приоритет операций

Операция	Обозначение	Приоритет
Унарный плюс	+	1
Унарный минус	-	1
Целочисленное деление	Div	2
Нахождение остатка	Mod	2
Умножение	*	3
Деление	/	3
Сложение	+	4
Вычитание	-	4

Значение выражений может быть вычислено только в том случае, если все переменные являются связанными с каким-то конкретным значением во время вычисления.

Допускается использование шестнадцатеричных чисел. Для представления числа в этом формате используется предшествующий знак доллара (без пробела).

Операции отношения

Имеются следующие операции отношения

< меньше;
 <= меньше или равно;
 = равно;
 > больше;
 >= больше или равно
 <> или > < не равно

В Прологе допустима привычная формулировка вычисления отношений:

X+4<9-Y

Кроме числовых выражений, возможно также сравнение одиночных символов и строк.

Например:

'a' < 'b'

peter > sally

"Antony" < "antonia"

Сравнение символов и строк производится посимвольно слева направо. При этом сравниваются соответствующие ASCII-коды символов. Если сравниваемые символы равны, то происходит переход к следующей паре символов, и т.д. Если окажется, что в какой-то паре символы не равны, то одна строка признается меньшей (или большей).

В приведенных выше примерах второе сравнение приведет к неудаче, так как ASCII-коды уже первых символов строк `peter` и `sally` имеют обратное соотношение.

Следует иметь в виду, что при операциях над вещественными числами может происходить потеря точности и кажущиеся очевидными равенства могут не выполняться.

Например:

X= Sqrt(25), X=5.

Поэтому в подобных случаях лучше проверять полученные результаты на вхождение в диапазон.

Например:

X = Sqrt(25), X>4,999999999, X<5.000000001

Математические функции

Имеющийся набор встроенных математических функций представлен в следующей таблице.

Предикат	Описание
BitAnd(X,Y,Z) BitNot(X,Z) BitOr(X,Y,Z) BitXor(X,Y,Z)	Если X и Y связаны с целыми числами, то Z связывается с целым, являющимся результатом соответствующей побитовой логической операции (AND, OR, NOT, XOR) над X и Y, выраженными в виде 16-и разрядного двоичного числа.
BitLeft(X,N,Y) BitRight(X,N,Y)	Здесь X и N целые числа. Y связывается со значением, которое получается в результате соответствующего сдвига числа X, представленного в виде двоичного числа, на N позиций влево или вправо
X mod Y	Вычисление остатка при целочисленном делении
X div Y	Целочисленное деление
Abs(X)	Вычисление абсолютной величины X
Cos(X) Sin(X) Tan(X) ArcTan(X)	Тригонометрические функции. Угол выражается в радианах.
Exp(X)	Экспонента
Ln(X)	Натуральный логарифм
Log(X)	Десятичный логарифм
Sqrt(X)	Квадратный корень

Для вычисления степеней можно использовать следующую формулу:

$$y = \text{Exp}(a \cdot \text{Ln}(x)),$$

где a - основание степени;

x – показатель степени.

Например.

- для вычисления $y = 4^3$ записывается выражение: $y = \text{Exp}(3 \cdot \text{Ln}(4))$;

- для вычисления $y = \sqrt[5]{243}$ записывается выражение $y = \text{Exp}(1/5 \cdot \text{Ln}(243))$.

Для вычисления логарифмов с произвольным основанием можно использовать формулу:

$$Y = \text{Log}_n X = \text{Ln}(X) / \text{Ln}(n)$$

Например.

- для вычисления $Y = \text{Log}_5 25$ записывается выражение $Y = \text{Ln}(25) / \text{Ln}(5)$.

2. Пример

Пусть необходимо реализовать вычисление значения следующей функции:

$$y = \begin{cases} \sin \sqrt[5]{|x+2|} + e^{2+x}, & \text{если } x \leq -1 \\ \text{tg} \sqrt[3]{(x-1)^2} + e^{3-3x}, & \text{если } -1 < x < 3 \\ \log_4(2 + \sqrt{x}) + \text{tg} \sqrt[3]{|4-x|}, & \text{если } x \geq 3 \end{cases}$$

Работа производится в среде Visual Prolog 5.2. Данная среда позволяет создавать полноценные многооконные Windows-приложения. Однако, поскольку целью данной работы является получение навыков логического программирования, то работа по созданию программ будет вестись в режиме интерпретатора.

После запуска системы появится ее рабочее окно – рис. 1.

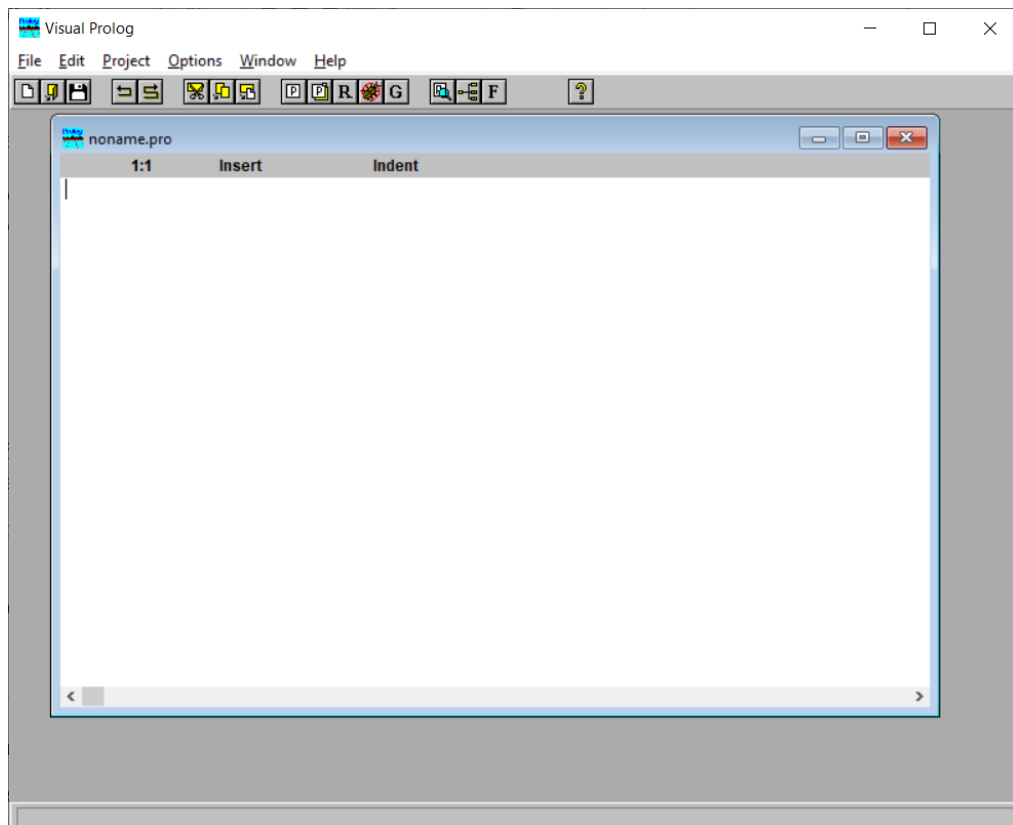


Рисунок 1 – Рабочее окно Visual Prolog 5.2

В главном меню выбирается пункт **File > New**.
 В появившемся модуле набирается текст программы.
 Для данного примера программа имеет вид:

Predicates

nondeterm F(real,real)

Goal

Write("x="), ReadReal(X), F(X,Y), Write("y=",Y),nl.

Clauses

*F(X, Y):- X<=-1,Y=sin(exp(1/5*ln(abs(X+2))))+exp(2+X).*

*F(X, Y):- X>-1,X<3,Y=tan(exp(2/3*ln(X-1)))+exp(3-3*X).*

*F(X, Y):- X>=3,Y=ln(2+sqrt(X))/ln(4)+tan(exp(1/3*abs(4-X))).*

Для запуска программы нажимает кнопка **G (goal test)** главного меню или комбинация клавиш **Ctrl/G**.

3. Варианты заданий

Номер варианта соответствует номеру студента по классному журналу.

№	Система
1.	$y = \begin{cases} \log_2(1+x^2) - \operatorname{tg}^5\sqrt{ 1-x }, & \text{если } x \leq 1 \\ e^2 - e^x + \sin^3\sqrt{2-x} + 1, & \text{если } 1 < x \leq 2 \\ 12 + \sin^3(3-x), & \text{если } x > 2 \end{cases}$
2.	$y = \begin{cases} \cos(x) - \cos(1) + 4e^{1-x}, & \text{если } x \leq 1 \\ 12 + \sin^3(2-x), & \text{если } 1 < x < 3 \\ 3^{ x-3 } - \sin(3) + \sin(x), & \text{если } x \geq 3 \end{cases}$
3.	$y = \begin{cases} 3^{ x-1 } - \sin 2 + \sin(x), & \text{если } x \leq 2 \\ \cos^2(3-x) + \log_2(1+x), & \text{если } 2 < x \leq 3 \\ 2\lg(x+6), & \text{если } x > 3 \end{cases}$
4.	$y = \begin{cases} \operatorname{tg}^5\sqrt{ 1-x ^3} + e^{0,5-0,5x}, & \text{если } x < 1 \\ \cos^2(2-x) + \log_3(1+x), & \text{если } 1 \leq x < 5 \\ \sin\left(1 - \frac{x}{3}\right) + 2\lg(1+3x), & \text{если } x \geq 5 \end{cases}$
5.	$y = \begin{cases} e^2 - e^{2x} + \sin^3\sqrt{ x-1 } + 1, & \text{если } x \leq 1 \\ 2\lg(6+2x), & \text{если } 1 < x \leq 2 \\ 12 + \sin^3(3-x), & \text{если } x > 2 \end{cases}$
6.	$y = \begin{cases} \sin(8+2x) + e^{1+\frac{x}{4}}, & \text{если } x = -4 \\ \lg(2+2x) + \operatorname{tg}(2x-8), & \text{если } x = 4 \\ 6 + 3\log_{10}(1-2x), & \text{в остальных случаях} \end{cases}$

7.	$y = \begin{cases} \cos(3+x) - 2e^{x+3}, & \text{если } x \leq -3 \\ \sin\left(1 - \frac{x}{3}\right) + 2\lg(1+3x), & \text{если } -3 < x \leq 3 \\ 3 - \lg(10 + 8 - 2x) - e^{4-x}, & \text{если } x > 3 \end{cases}$
8.	$y = \begin{cases} e^{2+x} + \sin\sqrt[3]{ 1+x }, & \text{если } x \leq -1 \\ \log_3(2+x^3) + \operatorname{tg}\left(\frac{1}{x} - 1\right), & \text{если } -1 < x \leq 4 \\ \log_5(1+x^2) - \operatorname{tg}\sqrt[5]{ 2-x }, & \text{если } x > 4 \end{cases}$
9.	$y = \begin{cases} \sqrt[3]{1+ x-1 } + 2\cos(x-1), & \text{если } x = 1 \\ 6 + 3\log_{10}(8+2x), & \text{в остальных случаях} \end{cases}$
10.	$y = \begin{cases} \sqrt[3]{1+ 2x/10+1 } + 2\cos(x-5), & \text{если } x < -5 \\ 6 + 3\log_{10}(8+2x), & \text{если } -5 \leq x < 5 \\ \operatorname{tg}\sqrt[3]{(x-5)^2} + e^{\sqrt{10-2x}}, & \text{если } x \geq 5 \end{cases}$
11.	$y = \begin{cases} \sqrt[3]{1+ 2x/10+1 } + 2\cos(x+5), & \text{если } \cdot x \leq -5 \\ 6 + 3\log_{10}(8+2x), & \text{если } \cdot -5 < x < 5 \\ \operatorname{tg}\sqrt[3]{(x-5)^2} + e^{\sqrt{10-2x}}, & \text{если } \cdot x \geq 5 \end{cases}$
12.	$y = \begin{cases} \log_3(2+x^2) + \operatorname{tg}\sqrt[3]{ 1-x }, & \text{если } \cdot x < 3 \\ \cos(4-x) + e^{8-2x}, & \text{если } \cdot 3 \leq x \leq 6 \\ \sin\sqrt[5]{x-10} + e^{\frac{1-x}{10}}, & \text{если } \cdot x > 6 \end{cases}$
13.	$y = \begin{cases} 3 - \lg(10 + 2x+4) - e^{x+2}, & \text{если } \cdot x < -1 \\ 5 - \cos^2(2\pi x), & \text{если } \cdot -1 \leq x \leq 1 \\ 2\lg_5(5 + 4 - x^2), & \text{если } \cdot x > 1 \end{cases}$
14.	$y = \begin{cases} \cos^2(x) + \log_2(2+ x), & \text{если } \cdot x \leq 0 \\ \cos(1-x) + e^{2-2x}, & \text{если } \cdot 0 < x \leq 4 \\ 2\lg(10 + 25 - x^2), & \text{если } \cdot x > 4 \end{cases}$
15.	$y = \begin{cases} 2\cos^2(x+2), & \text{если } \cdot x < 0 \\ 2\lg(10 + 1 - x^2), & \text{если } \cdot 0 \leq x \leq 5 \\ 2^{\frac{10}{x}-1} - \sin(x-10), & \text{если } \cdot x > 5 \end{cases}$

4. Вычисление логических выражений

В качестве примера рассмотрим следующую задачу.

Дано логическое уравнение:

$$Y = X1 \vee \overline{X2} \wedge \overline{\overline{X3} \vee X2} \wedge X3$$

Необходимо составить программу для вычисления его значений. Значения входных переменных вводятся с клавиатуры.

Программная реализация:

Predicates

nondeterm F(integer,integer,integer,integer)
nondeterm OrP(integer,integer,integer)
nondeterm AndP(integer,integer,integer)
nondeterm NotP(integer,integer)

Goal

Write("x1="), ReadInt(X1), nl,
Write("x2="), ReadInt(X2), nl,
Write("x3="), ReadInt(X3), nl,
F(X1, X2, X3, Y), Write("y=", Y).

Clauses

F(X1, X2, X3, Y):- NotP(X3, Y1), AndP(X2, X3, Y2),
OrP(Y1,Y2,Y3),NotP(Y3,Y4),
NotP(X2, Y5), AndP(Y4, Y5, Y6),
OrP(X1, Y6, Y).

OrP(0,0,0).

OrP(0,1,1).

OrP(1,0,1).

OrP(1,1,1).

AndP(0,0,0).

AndP(0,1,0).

AndP(1,0,0).

AndP(1,1,1).

NotP(0,1).

NotP(1,0).

Обратите внимание на созданные в программе предикаты для реализации стандартных логических операций И, ИЛИ и НЕ (**OrP**, **AndP** и **NotP**).

5. Варианты заданий

Составить программу для вычисления логического уравнения. Значения входных переменных вводятся с клавиатуры.

Номер варианта соответствует номеру студента по классному журналу.

№	Уравнение
1.	$Y = X1 \vee X2 \wedge (\overline{X1} \vee X2) \vee X1 \wedge \overline{X3}$
2.	$Y = (X1 \wedge \overline{X2}) \vee X3 \vee (\overline{X1} \vee X2) \wedge \overline{X3}$
3.	$Y = \overline{X1} \vee X2 \wedge \overline{X3} \vee X1 \wedge (\overline{X2} \vee X3)$
4.	$Y = (X1 \vee X2) \wedge (\overline{X1} \vee \overline{X2}) \vee X1 \wedge X2$
5.	$Y = X1 \wedge X2 \wedge \overline{X3} \vee X1 \wedge \overline{X2} \wedge X3$
6.	$Y = \overline{X1} \vee X2 \vee (\overline{X1} \wedge X2) \vee (\overline{X1} \vee X2)$
7.	$Y = \overline{X1} \vee \overline{X1} \vee X2 \vee (X2 \wedge (\overline{X1} \vee X2))$

8.	$Y = X1 \wedge X2 \vee (X3 \vee X1) \wedge \overline{X2}$
9.	$Y = (X1 \vee X2 \vee X3) \wedge ((\overline{X2} \wedge \overline{X3}) \vee X1)$
10.	$Y = (X1 \vee X2 \vee X3) \wedge (X1 \vee X2) \wedge (X2 \vee X3)$
11.	$Y = X1 \vee \overline{X2} \wedge (X1 \vee \overline{X3})$
12.	$Y = \overline{X1} \wedge X2 \vee \overline{\overline{X1}} \wedge X3$
13.	$Y = (\overline{X1} \wedge X2) \vee (X1 \wedge \overline{X3})$
14.	$Y = \overline{X1} \vee X2 \wedge X1 \vee \overline{X3}$
15.	$Y = (\overline{X1} \vee X2) \wedge (\overline{X1} \vee \overline{X3})$

Лабораторная работа №2 Основы языка Prolog. Поиск решений.

1. Работа со списками.

Список - широко используемая структура данных, которую удобно применять при рекурсивной обработке информации, состав и количество которой изменяется в ходе процесса обработки.

Список - это упорядоченная последовательность элементов, которая может иметь произвольную длину. Некоторым аналогом списков являются данные типа массив в процедурных алгоритмических языках

Элементами списка могут быть любые термы:

- константы
- переменные
- структуры, которые включают и другие списки.

Списки широко используются для представления деревьев синтаксического разбора, грамматик, карт городов, программ на ЭВМ и математических объектов (графы, формулы, функции).

Для записи списков в Прологе предусмотрена скобочная форма записи, при которой элементы списка заключены в квадратные скобки и разделены запятой.

Например:

- [] - пустой список;
- [a] - список из одного компонента;
- [a, b, c] - список из трех компонентов.

Кроме того, в Прологе используется представление списков в виде структуры "голова - хвост". В этом представлении список - это или любой пустой список, не содержащий ни одного элемента, или структура, имеющая два компонента: "голову" и "хвост" списка.

При такой форме записи "головой" списка является первый его элемент, "хвост" - это список, состоящий из всех элементов исходного списка, за исключением первого элемента.

Для этого введена специальная форма представления списка с "головой" - X и "хвостом" - Y, где для разделения X и Y используется вертикальная черта.

В соответствии с этой формой список [a, b, c] можно записать как

[a |[b, c]].

Для вывода на экран элементов списка можно воспользоваться стандартным предикатом **Write**. При этом вывод происходит в синтаксисе Пролога. Поэтому чаще для более «приличного» вывода приходится конструировать специальный предикат распечатки списка. Примеры приведены в последующих программах.

/ Распечатка элементов списка */*

Domains

*List=integer**


```

Predicates
Print(list)
Goal
L=[12,23,34,45,56],
Print(L).
Clauses
Print([]).
Print([X|L]):- Print(L), Write(X, " ").

/* Распечатка элементов двойного списка */
Domains
List=integer*
LList=list*
Predicates
Print1(list)
Print2(lList)
Goal
L1=[1,2,3,4],
L2=[2,3,4,5],
L3=[3,4,5,6],
L4=[4,5,6,7],
L=[L1,L2,L3,L4],
Print2(L).
Clauses
Print1([]).
Print1([X|L]):- Write(X, " "),Print1(L).

Print2([]).
Print2([X|L]):-Print1(X),nl,Print2(L).

```

Для ввода списков в программу можно использовать практически все методы, используемые в алгоритмических языках:

- прямое их определение в тексте программы;
- ввод с клавиатуры;
- генерация с помощью встроенного генератора случайных чисел.

В качестве примеров приводятся следующие программы.

```

/* Ввод элементов списка с клавиатуры */
Domains
List=integer*
Predicates
Vvod(list,list)
Print(list)
Add(string,list,list)
Goal
Vvod([],L),
Print(L).
Clauses
Vvod(L,L1):- Readln(S), Add(S,L,L1).

Add("",L,L).
Add(S,L,L1):- Str_Int(S,X),L2=[X|L], Vvod(L2,L1).

```

```
Add(_,L,L1):- Write("Ошибка ввода."),nl,  
Write("Для прекращения ввода - просто 'Enter'\n"),  
Vvod(L,L1).
```

```
Print([]).  
Print([X/L]):- Write(X, " "), Print(L).
```

В качестве простого примера работы со списками приводится текст программы, которая распечатывает элемента списка с указанным номером.

```
/* Выбор элемента списка по его номеру */  
Domains  
List=integer*  
Predicates  
Print(list)  
Vibor(integer, list)  
Goal  
L=[12,23,34,45,56],  
Print(L),nl,  
Write("Укажите номер элемента:"),  
ReadInt(N),  
Vibor(N,L).  
Clauses  
Print([]).  
Print([X/L]):- Write(X, " "), Print(L).  
  
Vibor(_,[]):- Write("Указанный номер больше размерности списка").  
Vibor(1,[X/_]):- Write(X),!.  
Vibor(N,[_L]):- N1=N-1, Vibor(N1,L).
```

2. Решение задач на поиск решений

В качестве примера рассмотрим задачу о волке, козе и капусте, которая формулируется следующим образом:

Однажды крестьянину понадобилось перевезти через реку волка, козу и капусту. У крестьянина есть лодка, в которой может поместиться, кроме самого крестьянина, только один объект — или волк, или коза, или капуста. Если крестьянин оставит без присмотра волка с козой, то волк съест козу; если крестьянин оставит без присмотра козу с капустой, коза съест капусту. Как крестьянину перевезти на другой берег всё своё имущество в целостности и сохранности?

Основная трудность при решении подобных задач заключается в их формализации, т. е. в переводе условия задачи на язык структур данных и алгоритмов.

В данном случае можно предложить следующее:

- в задаче 4 объекта;
- их состояние в начале решения можно определить списком [left, left, left, left], т. е. все объекты на левом берегу;
- из этого состояния возможны четыре хода:
Крестьянин плывет в одиночку;
Крестьянин плывет с волком;
Крестьянин плывет с козой;

Крестьянин плывет с капустой.

- в результате каждого хода получается новое состояние объектов (например, в результате первого хода получится состояние [right, left, left, left]);

- полученное состояние проверяется на соответствие условиям задачи. Например, в случае первого хода новое состояние попадает под то ограничение, при котором объекты, оставшиеся на другом берегу без присмотра, съедят друг друга;

- если полученное состояние разрешено, то оно добавляется в цепочку переходов.

- при каждом разрешенном переходе производится проверка – на достижение конечного состояния. Если оно достигнуто, то решение найдено.

Таким образом алгоритм решения этой и всех подобных задач можно представить следующим образом:

Основная программа

Ввести S /* S - исходное состояние объектов

Ввести S0 /* S0 - конечное состояние объектов

Solution(S) /* Вызов процедуры решения

КЦ программы

Процедура Solution(S)

/* проверка – не достигнуто ли нужное состояние

Если S = S0

Вывод списка переходов

КЦ процедуры

КЦ Если

Цикл i = 1..n /* n – количество возможных ходов

Maybe(i) /* проверка – возможен ли i-ый ход

Move(i, S, S1) /* S1 – новое состояние в результате хода

P1=Zapret(S1) /* проверка – не является S1 запрещенным

P2=Povtor(S1) /* проверка – не повторялось ли состояние S1 в цепочке переходов

Если Maybe(i) And Not P1 And Not P2 then Solution(S1)

КЦ цикла

КЦ процедуры

В зависимости от условий задачи и специфики используемой среды программирование алгоритм может немного меняться. При решении на Прологе программная реализация может выглядеть следующим образом:

/ Волк, коза и капуста */*

Domains

State = Locate(symbol, symbol, symbol, symbol)

Way = State*

Predicates

nondeterm Solve(state, way, way)

nondeterm Move(integer, state, state)

nondeterm Zapret(state)

nondeterm Povtor(state, way)

nondeterm Print(way)

nondeterm Info(integer)

Goal

Write(" Волк, коза и капуста ">,

X=locate(left, left, left, left),

Solve(X,[X],L),

Print(L), nl, Fail.
Clauses
Solve(locate(right, right, right, right),L,L).
Solve(X,L,L1):- Move(_,X,X1),Not(Zapret(X1)),Not(Povtor(X1,L)),Solve(X1,[X1|L],L1).

Move(1,locate(left,X1,X2,X3), locate(right,X1,X2,X3)).
Move(2,locate(right,X1,X2,X3), locate(left,X1,X2,X3)).
Move(3,locate(left,left,X2,X3), locate(right,right,X2,X3)).
Move(4,locate(right,right,X2,X3), locate(left,left,X2,X3)).
Move(5,locate(left,X1,left,X3), locate(right,X1,right,X3)).
Move(6,locate(right,X1,right,X3),locate(left,X1,left,X3)).
Move(7,locate(left,X1,X2,left), locate(right,X1,X2,right)).
Move(8,locate(right,X1,X2,right), locate(left,X1,X2,left)).

Zapret(locate(left, right, right, _)).
Zapret(locate(right, left, left, _)).
Zapret(locate(left, _,right, right)).
Zapret(locate(right, _, left, left)).
Povtor(X,[X|_]).
Povtor(X,[_L]):- Povtor(X,L).
Print([_]).
Print([X,Y|L]):- Print([Y|L], Move(N,Y,X), Info(N), nl).

Info(1):- Write("Переплыл хозяин").
Info(2):- Write("Хозяин переплыл обратно ").
Info(3):- Write("Переплыли хозяин- и волк").
Info(4):- Write("Хозяин и волк переплыли обратно").
Info(5):- Write("Переплыли хозяин- и коза").
Info(6):- Write("Хозяин и коза переплыли обратно ").
Info(7):- Write("Переплыли хозяин и капуста").
Info(8):- Write("Хозяин и капуста переплыли обратно").

3. Варианты заданий

Номер варианта соответствует номеру студента по классному журналу

1. Ученая блоха

Блоха находится на поле размером $n \times n$. В клетках поля находятся числа, которые показывают, на сколько клеток влево, вправо, вверх или вниз может прыгнуть блоха с этой клетки. Заданы начальные координаты поля, в котором находится блоха, и конечные, куда она должна попасть. Определить может ли попасть блоха с начального поля в конечное и вывести ее путь. Если нет, то вывести сообщение об этом.

2. Обезьяна и банан

Обезьяна находится в комнате размером $N \times N$ клеток.

В какой-то клетке лежит ящик. В другой какой-то клетке под потолком висит банан. Обезьяна может ходить по комнате брать с собой ящик и ходить с ним или бросать его или залазить на него.

Конечная цель: обезьяна вместе с ящиком находится в клетке под бананом, залазит на ящик и хватает банан.

3. Перестановка мебели

Задана исходная расстановка мебели:

Диван	Шкаф	Стол
Комод		Кровать

Произвести ее перестановку следующим образом:

Шкаф	Стол	Диван
Комод		Кровать

Двигать мебель можно только на свободное рядом место.

Если достичь результата невозможно, то вывести сообщение об этом.

4. Переливашка

Дано три сосуда заданной емкости и исходное количество воды в них. Необходимо путем переливаний получить заданное количество воды. Метки на сосудах отсутствуют. Если решения нет, то вывести сообщение об этом.

Например:

даны сосуды емкостью 3, 5 и 7 литров. Во третьем сосуде находится 7 литров воды.

Необходимо получить 1 литр воды (в любом сосуде).

Тогда возможна следующая цепочка переливаний:

0 0 7

0 5 2

2 5 0

2 0 5

3 0 4

3 4 0

3 1 3

5. Кенгуру

На узенькой дорожке встречаются две семьи кенгуру (по N кенгуру в каждой семье). Между ними одно пустое место. Кенгуру могут либо передвинуться на пустое перед собой место, либо перепрыгнуть через «чужое» кенгуру на пустое место сзади него. Прыгать через своих нельзя. Необходимо найти последовательность прыжков, в результате которых семьи продолжат свой путь.

Например:

начальная позиция: >>> <<<

конечная позиция: <<< >>>

6. Пираты

Четыре пирата нашли сундук с сокровищами. Они хотят переправиться через реку. Пираты весят 45, 50, 60 и 65 кг, сундук — 100 кг, их лодка выдерживает груз не более 200 кг. Сундук можно погрузить в лодку или вытащить из нее только вчетвером. Как им всё-таки всем переправиться, не оставив и сундук?

7. Лисица и волчица

К реке подошли Волчица с двумя волчатами и Лисица с двумя лисятами. У берега привязана лодка, которая вмещает только двух зверей. Ситуация осложняется тем, что Волчица с Лисицей не доверяют друг другу и не оставят своих детей в своё отсутствие с другой мамой ни на берегу, ни в лодке. Грести умеют только Лисица и один из лисят. Как им переправиться?

8. Переход через мост

К мосту в полной темноте подошли четыре человека. Время перехода через мост для них составляет соответственно 1, 2, 4 и 10 минут.

За какое наименьшее время они могут перейти через мост при следующих условиях:

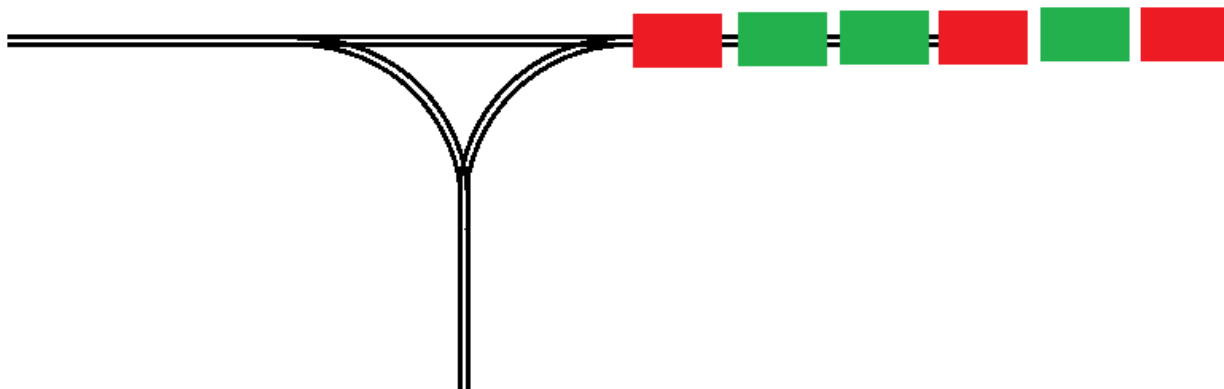
- переходить через мост можно только с фонариком;

- фонарик только один и передавать его можно только из рук в руки;

- одновременно по мосту могут двигаться максимум два человека. При этом время их перехода равно времени перехода самого медленного человека из пары.

9. Железнодорожный узел

Имеется железнодорожный узел следующего вида:



В правой ветке узла находится состав, состоящий из N вагонов одного вида и M вагонов другого. Необходимо, используя тупик, разделить состав на два, состоящих из вагонов одинакового вида. Предполагается, что в тупик могут поместиться все $N+M$ вагонов

9. Охотники

К реке подошли четверо охотников, каждый из которых весит 60, 65, 70 и 70 кг. На берегу находится лодка грузоподъемностью 130 кг. Найти последовательность переправ, позволяющую переправиться на другой берег.

10. Волки и зайцы.

К реке одновременно подошли три зайца и три волка. Всем необходимо было переправиться на противоположный берег. У берега стояла лодка, которая могла вместить только двух зверей. Зайцы боязливо поглядывали на волков, так как знали, что во время переправы может всякое случиться. Если во время переправы на берегу число волков и зайцев будет одинаковым, то волки не тронут зайцев; если же число волков превысит число зайцев хотя бы на одного, то волки съедят зайцев.

Найти последовательность переправ, позволяющую всем зверям, переправиться на другой берег.

11. Ореховый клад.

Лесная тропинка разделена на N клеток. Бельчонок закопал в клетке номер M клад с орехами. Сам бельчонок находится в клетке номер 1. Для каждой клетки известно, расстояние (в клетках) прыжка вперед или назад. Необходимо найти последовательность прыжков, приводящих к кладу (или указать, что такой последовательности нет).

12. Задача о волке, козе и капусте 2.

В отличие от классической задачи здесь ограничения выглядят следующим образом:

- нельзя оставлять наедине волка с козой и козу с капустой;
- в лодку кроме крестьянина может помещаться еще два объекта;
- никакой объект нельзя оставлять на берегу в одиночестве (украдут или убежит).

13. Фишки

Дана полоска, состоящая из N клеток. На ней в случайном порядке находится $(N-2)$ пронумерованных фишек. Необходимо упорядочить фишки по возрастанию номеров.

Например, из положения:

	2	8	10	3		4	7	5	6	9	1
--	---	---	----	---	--	---	---	---	---	---	---

Получить:

1	2	3	4	5	6	7	8	9	10		
---	---	---	---	---	---	---	---	---	----	--	--

При этом фишки могут быть перемещены налево или направо на свободное место или их можно перемещать на свободное место через одну фишку.

14. Переправа

На одной стороне реки находятся три человека и три льва. Все они должны оказаться на

другом берегу реки. Есть лишь одна лодка, в которой могут поместиться лишь два живых существа одновременно (человека или льва). Вы не можете оставлять на том или другом берегу реки больше львов, чем людей, так как в этом случае животные съедят людей, оставшихся в меньшинстве. Как вы переправите всех через реку?

15. Люди и обезьяны

Три человека, одна большая и две маленькие обезьяны должны переправиться через реку. Есть одна лодка, в которой может поместиться не больше двоих. Только люди и большая обезьяна умеют грести. Нельзя, чтобы оставались вместе больше обезьян, чем людей, иначе обезьяны убьют людей. Обезьяны могут выпрыгивать на берег, когда лодка причаливает.

Как им переправиться через реку?

16. Семья

Отец, мать и двое детей – сын и дочь, должны переправиться через реку. Поблизости случился рыбак, который мог бы одолжить им свою лодку. Однако, в лодке могут поместиться только один взрослый или двое детей.

Как семье переправиться через реку и вернуть рыбаку его лодку?

Лабораторная работа № 3 Использование статистических методов. Методы прогнозирования при регрессионном анализе

1. Общая схема регрессионного анализа

Согласно [1] регрессионный анализ состоит из следующих этапов.

1.1. Выбор аппроксимирующей функции

В тех случаях, когда математическая модель процесса неизвестна, для аппроксимации имеющихся данных используются полиномиальные зависимости. Для одномерного случая такой полином имеет вид:

$$y = \sum_{k=0}^n a_k x^k, \quad (1.1)$$

где

y – выходной параметр;

x – фактор (входной параметр);

a – коэффициенты уравнения;

n – степень полинома.

Например:

полином первой степени имеет вид: $y = a_0 + a_1 x$ (1.2)

полином второй степени: $y = a_0 + a_1 x + a_2 x^2$ (1.3)

полином третьей степени: $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ (1.4)

и т.д.

Полиномы более высоких степеней использовать не рекомендуется, поскольку даже если при последующем анализе окажутся статистически значимыми коэффициенты при высоких степенях, то их физический смысл будет очень трудно объяснить.

1.2. Расчет коэффициентов полинома

Вычисление коэффициентов удобнее всего производить используя аппарат матричной алгебры.

Основное уравнение для вычисления коэффициентов в матричном виде выглядит следующим образом:

$$A = (X^T X)^{-1} X^T Y, \quad (1.5)$$

где X – матрица входных параметров;
 X^T - транспонированная матрица входных параметров;
 $(X^T X)^{-1}$ - обратная матрица произведения $X^T X$;
 Y - вектор-столбец значений выходного параметра.

1.3. Проверка уравнения на адекватность

Суть проверки заключается в сравнении дисперсии выходного параметра, обусловленной применяемым уравнением с дисперсией ошибки измерений.

В случае технических экспериментов ошибка измерений оценивается по результатам повторных измерений при одних и тех же значениях входных параметров.

Для экономических данных повторные измерения как правило невозможны. Поэтому в таких случаях ошибка измерений оценивается как разность между общей дисперсией выходного параметра и дисперсией, которая обусловлена применением уравнения, т.е.:

$$SS_{ошибки} = SS_0 - SS_{уравнения} \quad (1.6)$$

где SS_0 – сумма квадратов отклонений значений выходного параметра от его среднего значения;

$SS_{ошибки}$ - сумма квадратов отклонений значений выходного параметра от рассчитанных значений;

$SS_{уравнения}$ - сумма квадратов отклонений, вызванная применением уравнения.

Далее находятся средние дисперсии для ошибки измерений и уравнения. Вычисления производятся по следующим формулам:

для ошибки:
$$s_{ошибки}^2 = \frac{SS_{ошибки}}{f_{ошибкиj}}; \quad (1.7)$$

для уравнения:
$$s_{уравнения}^2 = \frac{SS_{уравнения}}{f_{уравнения}}, \quad (1.8)$$

где f – число степеней свободы. Этот параметр вычисляется по формулам:

для уравнения: $f_{уравнения} = k + 1; \quad (1.9)$

для ошибки: $f_{ошибки} = N - k - 1, \quad (1.10)$

где N - общее число измерений;

k - количество коэффициентов в уравнении.

Для проверки адекватности уравнения вычисляется критерий Фишера:

$$F = \frac{s_{уравнения}^2}{s_{ошибки}^2}, \quad (1.11)$$

Который затем сравнивается с критическим значением критерия при выбранном уровне значимости (стандартным является уровень значимости, равный 0,05) и числе степеней свободы для числителя и знаменателя.

Если рассчитанное значение критерия больше критического, то уравнения признается адекватным. В противном случае уравнение является не адекватным.

1.4. Проверка коэффициентов на значимость

Проверка производится следующим образом.

Сначала вычисляется дисперсия каждого коэффициента уравнения:

$$s_{A_i}^2 = (X^T X)^{-1}_i s_{ошибки}^2, \quad (1.12)$$

где $(X^T X)_i^{-1}$ - значение i -ого диагонального элемента матрицы $(X^T X)^{-1}$.

Далее вычисляется стандартное отклонение каждого коэффициента:

$$s_{A_i} = \sqrt{s_{A_i}^2} \quad (1.13)$$

Затем для каждого коэффициента вычисляется доверительный интервал:

$$\Delta A_i = \pm t s_{A_i}, \quad (1.14)$$

где t - критерий Стьюдента при выбранном уровне значимости и числе степеней, равном $f_{ошибки}$.

Если значение коэффициента больше его доверительного интервала, то данный коэффициент признается статистически значимым. В противном случае коэффициент является не значимым.

Далее незначимые коэффициенты удаляются из исходного уравнения и для упрощенного таким образом уравнения производятся повторные расчеты согласно этапам 1 – 4.

5. Прогнозирование и оценка ошибки прогноза

Вычисление выходного параметра в предсказываемой точке факторного пространства производится по формуле:

$$\hat{y}_p = X_p^T A, \quad (1.15)$$

где \hat{y}_p - значение выходного параметра в точке предсказания;

X_p^T - транспонированный вектор входного параметра в точке предсказания;

A - вектор коэффициентов.

Определение дисперсии выходного параметра в точке предсказания производится по формуле:

$$s_{\hat{y}_p}^2 = X_p^T (X^T X)^{-1} X_p s_y^2 \quad (1.16)$$

Доверительный интервал выходного параметра в точке предсказания:

$$\Delta y_p = \pm t \sqrt{s_{\hat{y}_p}^2}, \quad (1.17)$$

где t - критерий Стьюдента при выбранном уровне значимости и числе степеней свободы, равном $N-k-1$.

2. Пример

В приложении *Транспорт и связь* приведены некоторые параметры по транспорту и связи для Чувашской республики.

Используем для обработки данные о наличии квартирных телефонных аппаратов телефонной сети общего пользования или имеющих на нее выход на 1000 чел. городского населения.

995	997	998	999	000	001	002	003
33,6	49,9	57,3	69,8	80,1	91,8	05,4	26,6

Количество телефонных аппаратов будет являться выходным параметром. Время будет выступать в качестве входного параметра.

Для расчетов будем использовать данные за 1995 – 2003 г. Данные за 2004 г. будут использованы для сопоставления с результатами прогноза.

Выбор уравнение регрессии

В качестве уравнения регрессии выберем полином третьей степени:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 \quad (1.18)$$

Формирование матрицы переменных

Для реализации расчетов разместим данные следующим образом:

	В	С	Д
4			
5		Y	X
6		133,6	1995
7		149,9	1997
8		157,3	1998
9		169,8	1999
10		180,1	2000
11		191,8	2001
12		205,4	2002
13		226,6	2003

До начала расчетов входную переменную необходимо нормировать и центрировать. Нормирование и центрирование производится по формуле:

$$\tilde{x}_i = 2 \frac{(x_i - x_c)}{(x_{\max} - x_{\min})}, \quad (1.19)$$

где \tilde{x}_i - кодированное значение входного параметра;

x_i - натуральное значение входного параметра;

x_{\max}, x_{\min} - максимальное и минимальное значение входного параметра;

x_c - центральное значение входного параметра.

Центральное значение параметра определяется по формуле:

$$x_c = \frac{x_{\min} + x_{\max}}{2} \quad (1.20)$$

Для реализации расчетов по формулам (1.19)-(1.20):

в ячейку D14 вводится формула =МИН(D6:D13);

в ячейку D15 вводится формула =МАКС(D6:D13);

в ячейку D16 вводится формула =(D14+D15)/2);

	В	С	Д	Е	Ф	Г	Н
3							
4			Натуральные значения	Кодированные значения			
5		Y	X	x0	x1	x2	x3

6		133,6	1995	1	-1	1	-1
7		149,9	1997	1	-0,5	0,25	-0,125
8		157,3	1998	1	-0,25	0,0625	-0,01563
9		169,8	1999	1	0	0	0
10		180,1	2000	1	0,25	0,0625	0,015625
11		191,8	2001	1	0,5	0,25	0,125
12		205,4	2002	1	0,75	0,5625	0,421875
13		226,6	2003	1	1	1	1
14		Мин=	1995				
15		Макс=	2003				
16		Центр=	1999				

- ячейки E6:E13 соответствуют фиктивной переменной - коэффициенту a_0 уравнения (18) и заполняются единицами:

- в ячейку F6 вводится формула $= (D6 - \$D\$16) / (\$D\$15 - \$D\$14) * 2$;

- в ячейку G6 вводится формула $= F6^2$;

- в ячейку H6 вводится формула $= F6^3$.

Все введенные формулы копируются до 13 строки.

В результате пересчета получены кодированные значения входных переменных с диапазоном изменения $-1..+1$.

Для удобства дальнейшей работы диапазону C6:C13 присвоим имя Y, а диапазону E6:H13 имя X.

Расчет коэффициентов уравнения

Расчет коэффициентов производится согласно уравнению (1.5).

Выполним поэтапное вычисление по указанной формуле.

2.1. Получим транспонированную матрицу X.

Для этого перейдем на ячейку E19 и вызовем функцию ТРАНСП. В окне параметров функции в качестве «массив» указать X,

При этом в ячейке E19 появится сообщение «#ЗНАЧ!». Выделим диапазон E19:L22, затем нажимаем F2 и выполним тройное нажатие Ctrl+Shift+Enter.

*Размер выделяемого диапазона (количество строк и столбцов) должен соответствовать размерности матрицы X - т.е. количество выделенных строк должно быть равно количеству коэффициентов в уравнении регрессии, а количество столбцов количеству данных. Чтобы не ошибиться при выделении рекомендуется заранее пронумеровать нужные строки и столбцы.

	D	E	F	G	H	I	J	K	L	
18										
19		1	1	1	1	1	1	1	1	
20	Xt=	-1	-0,5	-0,25	0	0,25	0,5	0,75	1	
21		1	0,25	0,0625	0	0,0625	0,25	0,5625	1	
22		-1	-0,125	-0,0156	0	0,015625	0,125	0,421875	1	

Присвоим этому диапазону имя Xt.

Выполним умножение матриц X^T и X

Для этого перейдем на ячейку E25 и вызовем функцию МУМНОЖ. В окне параметров функции в качестве «массив 1» указать Xt, а в качестве «массив 2» указать X.

При этом в ячейке E25 появится число 8. Выделим диапазон E25:H28*, затем нажимаем F2 и выполним тройное нажатие Ctrl+Shift+Enter.

*Размер выделяемого диапазона (количество строк и столбцов) должен соответствовать количеству коэффициентов в уравнении регрессии.

В результате в диапазоне E25:H28 появится результат перемножения:

		D	E	F	G	H	I
24							
25			8	0,75	3,1875	0,421875	
26			0,75	3,1875	0,421875	2,449219	
27		XtX=	3,1875	0,421875	2,449219	0,237305	
28			0,421875	2,449219	0,237305	2,209717	
29							

Присвоим этому диапазону имя XtX.

Выполним обращение этой матрицы.

Для этого перейдем на ячейку E31 и вызовем функцию МОБР. В качестве аргумента укажем XtX.

При этом в ячейке E31 появится число 0,260879.

Выделим диапазон E31:H34, затем нажимаем F2 и выполним тройное нажатие Ctrl+Shift+Enter.

В указанных ячейках должен появиться результат обращения матрицы:

	C	D	E	F	G	H	I
30							
31			0,260879	-0,04314	-0,33538	0,034024	
32		XtXobr=	-0,04314	2,169717	-0,08628	-2,38738	
33			-0,33538	-0,08628	0,853042	0,068049	
34			0,034024	-2,38738	0,068049	3,084884	
35							

Присвоим этому диапазону имя XtXobr.

Выполним умножение матриц Xt и Y.

Для этого переходим на ячейку E37 и вызовем функцию МУМНОЖ. В окне параметров функции в качестве «массив 1» указать Xt, а в качестве «массив 2» указать Y.

При этом в ячейке E37 появится число 1414,5.

Выделим диапазон E37:E40, затем нажимаем F2 и выполним тройное нажатие Ctrl+Shift+Enter.

*Количество выделенных строк соответствует количеству коэффициентов в уравнении регрессии.

	C	D	E	F
36				
37			1414,5	
38		XtY=	273,7	
39			582,25	
40			185,2469	
41				

Диапазону E37:E40 присвоим имя XtY.

И наконец последнее вычисление по формуле (1.5), т.е. перемножим матрицы XtXobr и XtY.

Для этого переходим на ячейку E43 и вызовем функцию МУМНОЖ. В окне параметров функции в качестве «массив 1» указать XtXobr, а в качестве «массив 2» указать XtY.

При этом в ячейке E43 появится число 168,2327.

Выделим диапазон E43:E46*, затем нажимаем F2 и выполним тройное нажатие Ctrl+Shift+Enter.

*Количество выделенных строк соответствует количеству коэффициентов в аппроксимирующем полиноме.

Должен получиться следующий результат:

	C	D	E
42			
43			168,2327
44		A=	40,34315
45			11,27501
46			5,78756
47			

Таким образом у нас получилось уравнение:

$$y = 168,23 + 40,343x + 11,275x^2 + 5,788x^3 \quad (1.21)$$

2.2. Проверка адекватности уравнения

Проверка адекватности производится на основе формул (1.6) - (1.11).

Реализация расчетов в Excel выглядит следующим образом:

1. Сначала вычисляются компоненты формулы (1.6) (эти компоненты называются также остаточными суммами квадратов).

Вычисление общей остаточной суммы квадратов производится по формуле:

$$SS_0^2 = \sum_{i=1}^n (y_i - \bar{y})^2, \quad (1.22)$$

где n - число измерений;

y_i - i -ое значение выходного параметра;

\bar{y} - среднее значение выходного параметра.

Остаточная сумма квадратов, обусловленная ошибкой измерений после применения модели вычисляется по внешне аналогичной формуле:

$$SS_e^2 = \sum_{i=1}^n (y_i - \bar{y}_i)^2, \quad (1.23)$$

где \bar{y}_i - значение выходного параметра, рассчитанное с помощью уравнения регрессии в i -ой точке.

Остаточная сумма квадратов, обусловленная моделью вычисляется как разность первых двух уже определенных компонентов.

Все вычисления удобнее свести в следующую таблицу:

	I	J	K	L
24				
25		Степень полинома	0	3
26		Остаточная сумма квадратов	6552,989	10,13743
27		Сумма квадратов, приходящаяся на уравнение		6542,851
28		Число степеней свободы ошибки		3
29		Число степеней свободы уравнения		4
30		Дисперсия ошибки		3,379144

31		Дисперсия уравнения		1635,713
32		Критерий Фишера		484,0613

Для вычислений по формуле (1.23) предварительно рассчитаем значения выходного параметра по полученному уравнению:

в ячейку I6 вводится формула:

$$=МУМНОЖ(Е6:Н6;А),$$

которая копируется до 13 строки столбца I.

Далее выполним расчеты в следующих ячейках:

В ячейки K25, L25 вводятся степень используемого уравнения.

В ячейку K26 вводится формула:

$$=ДИСП(Y)*7,$$

а в ячейку L26 вводится формула:

$$=СУММКВРАЗН(Y;I6:I13).$$

В ячейку L27 вводится формула:

$$=K26-L26.$$

В ячейках L28 рассчитывается число степеней свободы для ошибки. Расчет производится по формуле:

$$f=N-1-k,$$

где N - общее число измерений;

k - количество коэффициентов в уравнении.

В ячейке L30 рассчитывается дисперсия ошибки:

$$=L26/L28.$$

В ячейке L31 рассчитывается дисперсия уравнения:

$$=L27/L29.$$

В ячейке M129 рассчитывается критерий Фишера:

$$=L31/L30.$$

Рассчитанное значение критерия сравнивается с критическим при заданном уровне значимости (0,05) и числе степеней свободы для числителя, равном k, знаменателя, равном N-1-k.

Критическое значение можно определить с помощью функции Excel =ФРАСПОБР(0,05;4;3). В данном случае оно будет равно 9,12.

Рассчитанное значение критерия Фишера больше критического. Поэтому делается вывод о том, что полученное уравнение является адекватным.

2.3. Проверка коэффициентов модели на статистическую значимость

Проверка значимости коэффициентов производится на основе вычислений по формулам (4.12)-(4.14).

Сами вычисления удобно представить в виде таблицы:

	D	E	F	G	H	I
		Значение коэффициентов	Значение диагональных элементов матрицы $(X^T X)^{-1}$	Дисперсии коэффициентов	Стандартное отклонение коэффициентов	Доверительные интервалы коэффициентов
42		168,2327	0,260879	0,881549	0,938909	2,220166
43		40,34315	2,169717	7,331787	2,707727	6,402758
44		11,27501	0,853042	2,882551	1,697808	4,014677

45		5,78756	3,084884	10,42427	3,228663	7,634576
46						

В ячейки F42:F45 копируются диагональные элементы матрицы вводится $(X^T X)^{-1}$;

в ячейку G42 вводится формула (4.12), т.е. =F43*\$M\$30;

в ячейку H42 вводится формула (4.13), т.е. =G43^0,5;

в ячейку I42 вводится формула (4.14), т.е. =H43*\$H\$40

При этом в ячейку H40 введена формула, которая вычисляет критическое значение критерия Стьюдента:

$$=\text{СТЮДЕНТ.ОБР.2X}(0,05;7).$$

Все указанные формулы копируются до 46 строки.

Определение статистической значимости коэффициентов определяется путем сравнения значений каждого коэффициента с соответствующим доверительным интервалом:

- если значение коэффициента по модулю больше величины доверительного интервала, то данный коэффициент признается статистически значимым. В противном случае коэффициент считается незначимым.

Прямое сравнение столбцов E и I в вышеприведенной таблице показывает, в уравнении (4.18) статистически значимыми являются коэффициенты a_0, a_1, a_2 .

Основным следствием этого результата является следующее:

- в качестве уравнения регрессии необходимо взять уравнение:

$$y = a_0 + a_1 x + a_2 x^2 \quad (1.24)$$

и произвести перерасчет его коэффициентов по описанной выше схеме.

2.4. Результаты пересчета по уравнению (1.24)

Пересчет удобнее произвести на новом листе Excel.

Далее без подробных комментариев приводятся только основные результаты такого пересчета.

Матрица X:

x0	x1	x2
1	-1	1
1	-0,5	0,25
1	-0,25	0,0625
1	0	0
1	0,25	0,0625
1	0,5	0,25
1	0,75	0,5625
1	1	1

Транспонированная матрица X:

	1	1	1	1	1	1	1	1
Xt=	-1	-0,5	-0,25	0	0,25	0,5	0,75	1
	1	0,25	0,0625	0	0,0625	0,25	0,5625	1

Матрица XtX

	8	0,75	3,1875
XtX=	0,75	3,1875	0,421875
	3,1875	0,421875	2,449219

Обратная матрица $(X^T X)^{-1}$:

	0,260504	-0,01681	-0,33613
$X^T X_{obr}^{-1}$	-0,01681	0,322129	-0,03361
	-0,33613	-0,03361	0,851541

Матрица $X^T Y$:

	1414,5
$X^T Y$	273,7
	582,25

Матрица A:

	168,1689
A	44,82213
	11,14734

Результат проверки уравнения на адекватность:

Степень полинома	0	2
Остаточная сумма квадратов	6552,989	20,99549
Сумма квадратов, приходящаяся на уравнение		6531,993
Число степеней свободы ошибки		4
Число степеней свободы уравнения		3
Дисперсия ошибки		5,248873
Дисперсия уравнения		2177,331
Критерий Фишера		414,8188

Уравнение адекватно.

Проверка коэффициентов уравнения на значимость:

Значение коэффициентов	Значение диагональных элементов матрицы $(X^T X)^{-1}$	Дисперсии коэффициентов	Стандартное отклонение коэффициентов	Доверительные интервалы коэффициентов
168,1689	0,260504	1,367353	1,169339	2,765047
44,82213	0,322129	1,690813	1,300313	3,074751
11,14734	0,851541	4,469628	2,11415	4,999169

Все коэффициенты статистически значимы.

Таким образом уравнение описывающее наши данные имеет вид:

$$y = 168,17 + 44,82x + 11,15x^2 \quad (1.25)$$

2.5. Прогнозирование и оценка ошибки прогноза

Согласно постановке задачи нам необходимо спрогнозировать количество телефонных аппаратов в 2004 году.

Сначала рассчитаем вектор входного параметра в точке прогноза. Использование формул

(1.18)-(1.19) дает следующие значения:

	D	E	F	G
50	Xp=	1	1,25	1,5625

При этом в ячейку F5 введена формула $=(2004-D16)/(D15-D14)*2$, а в ячейку G5 - формула $=F50^2$

Прогнозное значение рассчитывается по формуле (1.15).

	D	E
52	Yp=	241,6143

При этом в ячейку E52 введена формула:

$=\text{МУМНОЖ}(E50:G50;Q37:Q39)$

(при условии того, что рассчитанные коэффициенты уравнения размещены в ячейках Q37:Q39).

Ошибка прогноза рассчитывается по формулам (1.16) - (1.17).

Если данные размещены следующим образом:

- вектор переменных в точке предсказания:

	K	L
48		Xp=
49		1
50		1,25
51		1,5625

- транспонированный вектор переменных в точке предсказания:

	D	E	F	G
50	Xp=	1	1,25	1,5625

- обратная матрица:

	P	Q	R	S
27		0,260504	-0,01681	-0,33613
28	XtXobr=	-0,01681	0,322129	-0,03361
29		-0,33613	-0,03361	0,851541

то в ячейку P50 введем формулу:

$=\text{МУМНОЖ}(Q27:S29;L49:L51)$,

которая после выделения ячеек P50:P52, нажатий F2 и Ctrl+Shift+Enter даст значения:

	P
49	
50	-0,28571
51	0,333333
52	0,952381

в ячейку P55 введем формулу =МУМНОЖ(E50:G50;P50:P52)*L30 (в L30 размещено значение дисперсии ошибки), которая даст значение: 8,498175.

Стандартное отклонение будет равно $s = \sqrt{8,498} = 2,915$

Доверительный интервал будет равен $\Delta y = \pm 2,776 \cdot 2,915 = \pm 8,09$ (где 2,776 критическое значение критерия Стьюдента при уровне значимости 0,05 и числе степеней свободы, равном 4 – определяется вызовом функции =СТЮДЕНТ.ОБР.2Х(0,05;4)).

Прогнозируемое значение будет равно:

$$y = 241,61 \pm 8,09 = 249,7...233,52$$

Реальное значение показателя в 2004 г. составило 249,9.

Лабораторная работа № 4 **Аппроксимация функций с использованием нейронных сетей** **в пакете «STATISTICA Neural Networks»**

Аппроксимация функций является одним из самых общих использований искусственных нейронных сетей. Нейронные сети — могут аппроксимировать непрерывные функции.

Рассмотрим применение нейронных сетей для аппроксимации функций на следующем примере.

Создадим, используя нейронную сеть, аппроксимирующую функцию $F(x, y, \text{знак}, z)$, позволяющую выполнять расчеты по следующим формулам:

$$F(x, y, "+", z) = x + y + z$$

$$F(x, y, "*", z) = x + y * z.$$

Генерацию обучающей выборки проще всего произвести в Excel. Для этого можно использовать следующий макрос (предполагается, что студенты знакомы с технологией их использования):

Sub Generation()

st = 2 'Номер строки для вывода

For x = 0 To 10 ' обучающая выборка для x, y и z находится в диапазоне 0..10

For y = 0 To 10

For z = 0 To 10

Cells(st, 1) = x

Cells(st, 2) = y

Cells(st, 3) = "+"

Cells(st, 4) = z

Cells(st,5)=x+y+z

st = st + 1

Cells(st, 1) = x

Cells(st, 2) = y

Cells(st, 3) = "*"

Cells(st, 4) = z

Cells(st, 5) = x + y * z

st = st + 1

Next z

Next y

Next x

End Sub

Запустив указанный макрос получаем обучающую выборку, стоящую из 2662 строк данных – рис. 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2	0	0 +		0	0								
3	0	0 *		0	0								
4	0	0 +		1	1								
5	0	0 *		1	0								
6	0	0 +		2	2								
7	0	0 *		2	0								

Рис. 1. Рабочий лист MS Excel после генерации обучающей выборки.

В «STATISTICA Neural Networks» создадим область данных для обучающей выборки, выполнив последовательность действий:

File > New > Data Set.

Поскольку, согласно условию, имеется четыре входных параметра (x, y, «знак», z) и один выходной параметр F(x, y, «знак», z), то заполним форму «Create Data Set» как показано на рис. 2.

Рис. 2. Форма ввода параметров области данных для выборки.

Дополним, открывшуюся после нажатия на кнопку «Create», форму «Data Set Edition (new)» пустыми строками для заполнения данными обучающей выборки, выполнив последовательность действий *Edit-> Cases-> Add* и заполним форму «Add Cases» как изображено на Рис. 3, поскольку в исходной выборке имеется 2662 строки данных.

Рис. 3. Форма ввода количества добавляемых строк области данных.

Назначим для столбца «VAR3», являющегося входным параметром (знак) обучающей выборки, возможные значения «+» и «*». Для этого выделим данный столбец, выбрав пункт контекстного меню *Definition* (Рис. 4) и заполним форму «Variable Definition» для двух возможных значений входного параметра «VAR3».

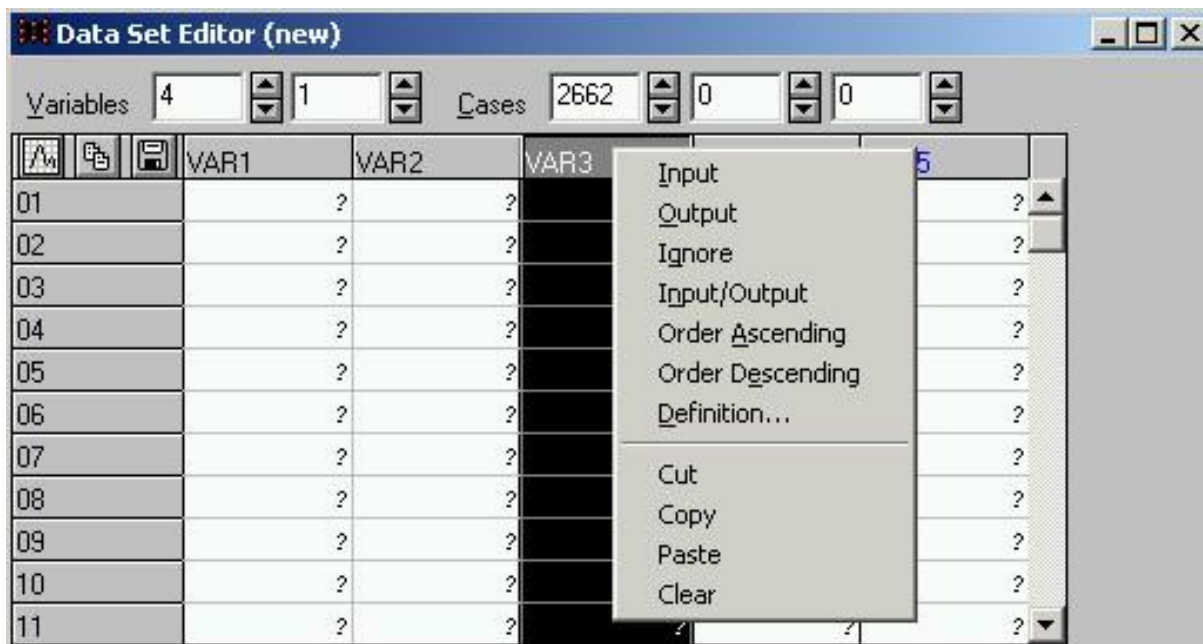


Рис. 4. Назначение возможных значений выходного параметра «VAR3».

Выполним копирование 2662 строк обучающей выборки, сгенерированных нами ранее в MS Excel, в форму «Data Set Edition (new)» (Рис. 5).

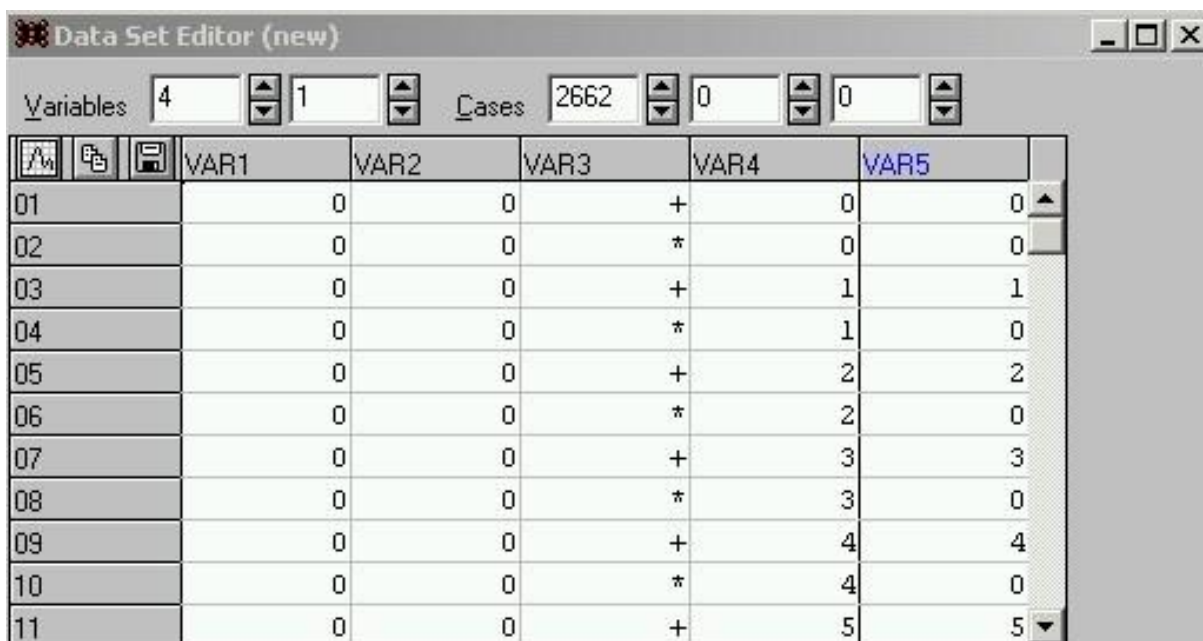


Рис. 5. Исходная выборка данных.

Пометим 10% строк обучающей выборки как «проверочные» (пункт *Verification* контекстного меню). Для этого уменьшим на ~10% количество

«обучающих» строк выборки **Cases** 2362 300 0

Перераспределим «проверочные» и «обучающие» строки обучающей выборки, выполним последовательность действий *Edit->Cases->Shuffle->Train and Verify*.

Создадим нейронную сеть, выполнив последовательность действий *File->New->Network*.

Заполним форму «Create Network» в соответствии с Рис. Ошибка! Текст указанного стиля в документе отсутствует..6 и создадим сеть (Рис. Ошибка! Текст указанного стиля в документе отсутствует..7) нажав кнопку «*Create*».

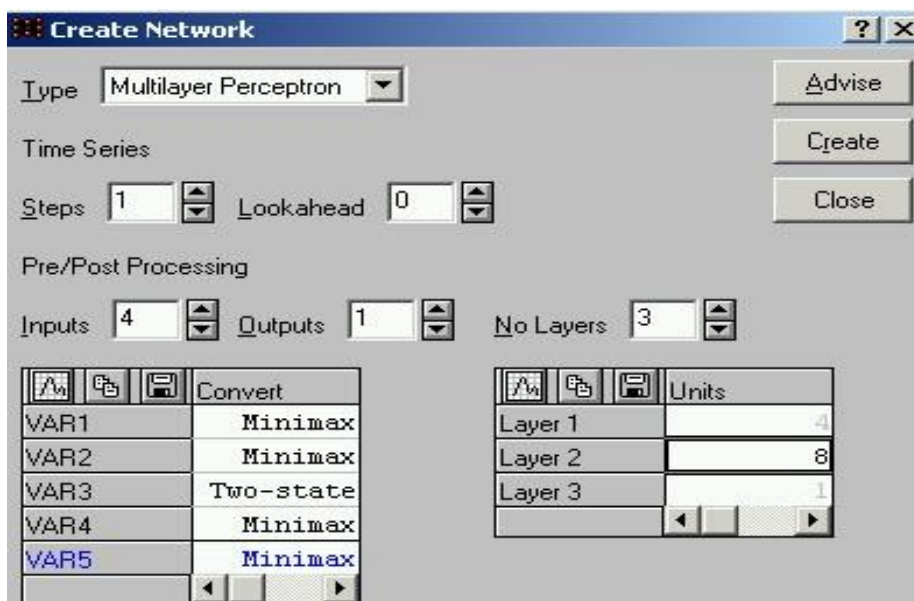


Рис. 6. Параметры создаваемой нейронной сети.

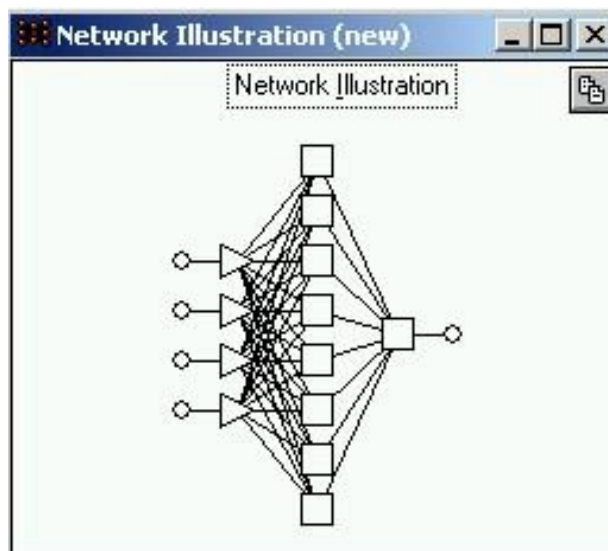


Рис. 7. Графическое представление созданной нейронной сети.

Обучаем полученную нейронную сеть, используя различные варианты обучения, представленные пунктом меню *Train->Multilayer Perceptrons*.

Для проверки уровня обучения сети удобно использовать форму «Training Error Graph» (Рис.8), вызываемую *Statistics -> Training Graph*. Особое внимание на этой форме следует обратить на количественные значения ошибки проверки и обучения нейронной сети.

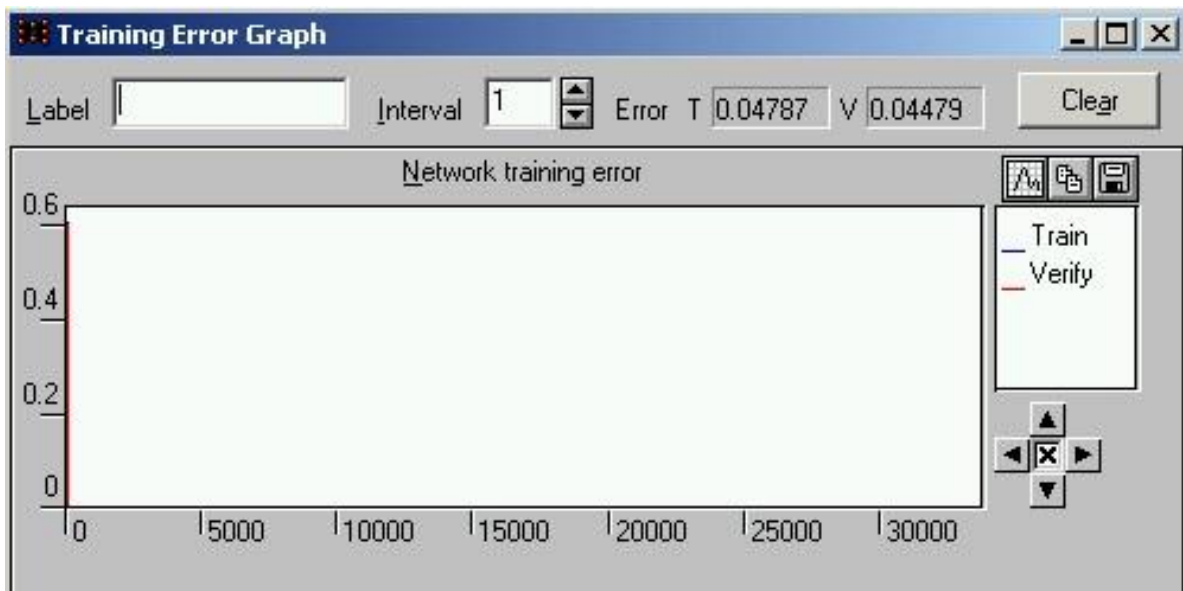


Рис. 8. Форма проверки уровня обучения сети.

После успешного обучения нейронной сети (ошибка проверки и ошибка обучения нейронной сети стремятся к 0) для вычисления с помощью созданной нами аппроксимирующей функции необходимо добавить две дополнительные строки данных, выполнив последовательность действий *Edit->Cases->Add*, пометить их как игнорируемые (пункт *Ignore* контекстного меню) и заполнить их контрольными примерами (Рис. 9).

	VAR1	VAR2	VAR3	VAR4	VAR5
01	3	4	*	10	43
02	2.3	7.7	+	5	15
03	0	0	+	0	0
04	0	0	*	0	0
05	0	0	+	1	1
06	0	0	*	1	0

Рис. 9. Заполнение контрольных примеров.

Выполнив последовательность действий *Run->Single Case* для 1-2 строки данных (Рис. 10, 11), получаем, что созданная нами нейронная сеть успешно аппроксимирует обе выбранные нами функции, погрешность вычислений при этом довольно не значительна и составляет в среднем 0,03-0,04%.

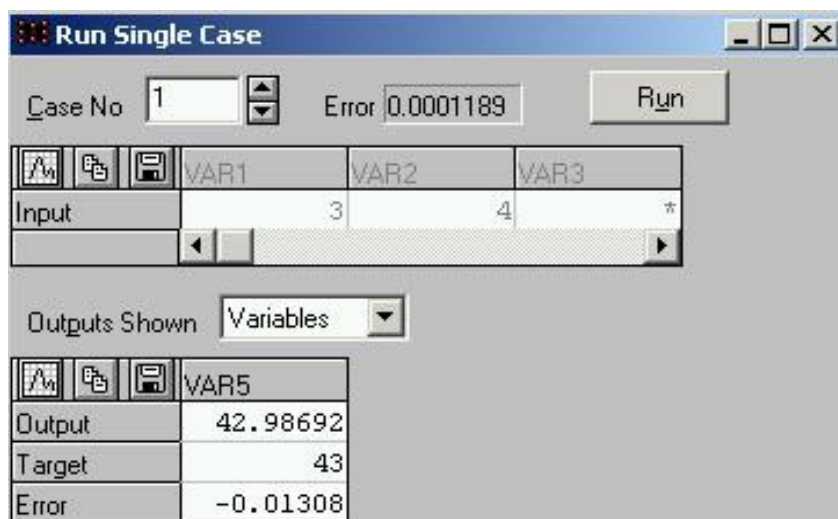


Рис. 10. Выполнение проверки аппроксимации нейронной сетью функции $F(x,y, "+" ,z)=x+y+z$

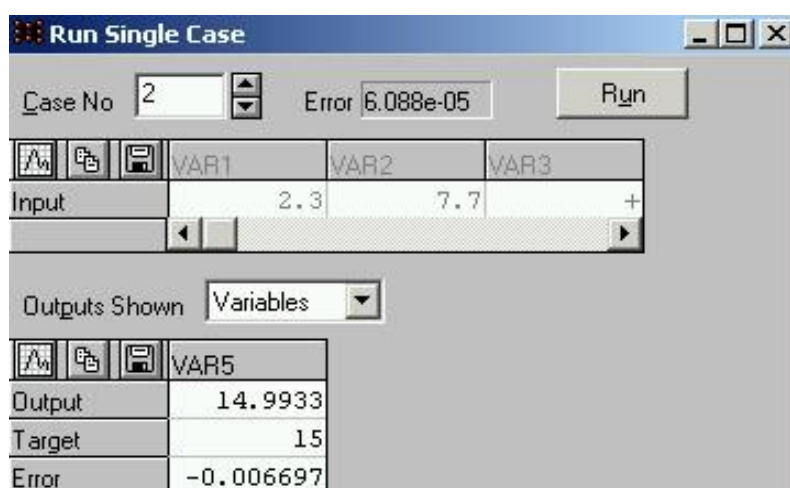


Рис. 11. Выполнение проверки аппроксимации нейронной сетью функции $F(x,y, "*" ,z)=x+y*z$

Лабораторная работа 5 Классификационный анализ с использованием нейронных сетей в пакете «STATISTICA Neural Networks»

Решение задачи классификации является одним из важнейших применений нейронных сетей. Задача классификации представляет собой задачу отнесения образца к одному из нескольких попарно не пересекающихся множеств. Примером таких задач может быть, например, задача определения кредитоспособности клиента банка, медицинские задачи, в которых необходимо определить, например, исход заболевания, решение задач управления портфелем ценных бумаг (продать купить или "придержать" акции в зависимости от ситуации на рынке), задача определения жизнеспособных и склонных к банкротству фирм.

Нейронные сети являются наиболее эффективным способом классификации, потому что генерируют фактически большое число регрессионных моделей (которые используются в решении задач классификации статистическими методами).

Однако, в применении нейронных сетей в практических задачах возникает ряд проблем. Во-первых, заранее неизвестно какой сложности (размера) может потребоваться сеть для достаточно точной реализации отображения. Эта сложность может оказаться чрезмерно высокой,

что потребует сложной архитектуры сетей.

При решении задач классификации необходимо отнести имеющиеся статические образцы (характеристики ситуации на рынке, данные медосмотра, информация о клиенте) к определенным классам. Возможно несколько способов представления данных. Наиболее распространенным является способ, при котором образец представляется вектором. Компоненты этого вектора представляют собой различные характеристики образца, которые влияют на принятие решения о том, к какому классу можно отнести данный образец. Например, для медицинских задач в качестве компонентов этого вектора могут быть данные из медицинской карты больного. Таким образом, на основании некоторой информации об образце, необходимо определить, к какому классу его можно отнести. Классификатор таким образом относит объект к одному из классов в соответствии с определенным разбиением N -мерного пространства, которое называется пространством входов, и размерность этого пространства является количеством компонент вектора.

Для построения классификатора необходимо определить, какие параметры влияют на принятие решения о том, к какому классу принадлежит образец. При этом могут возникнуть две проблемы. Во-первых, если количество параметров мало, то может возникнуть ситуация, при которой один и тот же набор исходных данных соответствует образцу, находящимся в разных классах. Тогда невозможно обучить нейронную сеть, и система не будет корректно работать (невозможно найти минимум, который соответствует такому набору исходных данных). Исходные данные обязательно должны быть непротиворечивы. Для решения этой проблемы необходимо увеличить размерность пространства признаков (количество компонент входного вектора, соответствующего образцу). Но при увеличении размерности пространства признаков может возникнуть ситуация, когда число образцов может стать недостаточным для обучения сети, и она вместо обобщения просто запомнит примеры из обучающей выборки и не сможет корректно функционировать. Таким образом, при определении признаков необходимо найти компромисс с их количеством.

Далее необходимо определить способ представления входных данных для нейронной сети, то есть определить способ нормирования. Нормировка необходима, поскольку нейронные сети работают с данными, представленными числами в диапазоне $0..1$, а исходные данные могут иметь произвольный диапазон или вообще быть нечисловыми данными. При этом возможны различные способы, начиная от простого линейного преобразования в требуемый диапазон и заканчивая многомерным анализом параметров и нелинейной нормировкой в зависимости от влияния параметров друг на друга.

Задача классификации при наличии двух классов может быть решена на сети с одним нейроном в выходном слое, который может принимать одно из двух значений 0 или 1 , в зависимости от того, к какому классу принадлежит образец. При наличии нескольких классов возникает проблема, связанная с представлением этих данных для выхода сети. Наиболее простым способом представления выходных данных в таком случае является вектор, компоненты которого соответствуют различным номерам классов. При этом i -я компонента вектора соответствует i -му классу. Все остальные компоненты при этом устанавливаются в 0 . Тогда, например, второму классу будет соответствовать 1 на 2 выходе сети и 0 на остальных. При интерпретации результата обычно считается, что номер класса определяется номером выхода сети, на котором появилось максимальное значение. Например, если в сети с тремя выходами мы имеем вектор выходных значений $(0.2, 0.6, 0.4)$, то мы видим, что максимальное значение имеет вторая компонента вектора, значит класс, к которому относится этот пример, – 2 . При таком способе кодирования иногда вводится также понятие уверенности сети в том, что пример относится к этому классу. Наиболее простой способ определения уверенности заключается в определении разности между максимальным значением выхода и значением другого выхода, которое является ближайшим к максимальному. Например, для рассмотренного выше примера уверенность сети в том, что пример относится ко второму классу, определится как разность между второй и третьей компонентой вектора и равна $0.6 - 0.4 = 0.2$. Соответственно, чем выше уверенность, тем больше вероятность того, что сеть дала правильный ответ. Этот метод кодирования является

самым простым, но не всегда самым оптимальным способом представления данных.

Следующим важным моментом при решении задач классификации является правильный выбор архитектуры нейронной сети. Построить небольшую и качественную нейронную сеть часто бывает просто невозможно, а большая нейронная сеть будет просто запоминать примеры из обучающей выборки и не производить аппроксимацию, что, естественно, приведет к некорректной работе классификатора. Существуют два основных подхода к построению нейронной сети – конструктивный и деструктивный.

При первом из них вначале берется нейронная сеть минимального размера, и постепенно увеличивают ее до достижения требуемой точности. При этом на каждом шаге ее заново обучают. Также существует так называемый метод каскадной корреляции, при котором после окончания эпохи происходит корректировка архитектуры нейронной сети с целью минимизации ошибки. При деструктивном подходе вначале берется нейронная сеть завышенного объема, и затем из нее удаляются узлы и связи, мало влияющие на решение. При этом полезно помнить следующее правило: число примеров в обучающем множестве должно быть больше числа настраиваемых весов. Иначе вместо обобщения нейронная сеть просто запомнит данные и утратит способность к классификации – результат будет не определен для образцов, которые не вошли в обучающую выборку.

При выборе архитектуры сети обычно опробуется несколько конфигураций с различным количеством элементов. При этом основным показателем является объем обучающего множества и обобщающая способность нейронной сети.

1.1.1 Применение пакета «STATISTICA Neural Networks» для классификационного анализа с использованием нейронных сетей

Рассмотрим применение классификационного анализа с использованием нейронных сетей на примере прогнозирования рентабельности потребсоюзов.

По результатам статистического анализа финансовых показателей потребсоюзов за 2009 года (использовано 2 показателя: объем продаж и прибыль) были выявлены две группы потребсоюзов: группа «А» – рентабельные, группа «В» – нерентабельные.

Несколько потребсоюзов сделали прогноз своего развития на текущий год и представили его в форме (Таблица 1). Необходимо определить, используя значения прошлого года в качестве обучающей выборки, какие из потребсоюзов окажутся рентабельными, а какие нерентабельными.

Финансовые показатели потребсоюзов

Таблица 1

Рентабельные		Нерентабельные		Прогноз	
объем продаж	прибыль	объем продаж	прибыль	объем продаж	прибыль
1,681819	0,030908	5,886363	0,665634	6,727273	0,623579
1,85	0,034141	5,213637	0,680305	4,372727	0,554121
1,513636	0,035622	5,55	0,697755	5,381818	0,768088
1,345454	0,042527	6,895454	0,698502	3,7	0,137848
2,186364	0,075902	5,045455	0,701284	5,213637	0,440325
0	0,08062	8,409091	0,707762	6,390909	0,259315

1,177273	0,097971	7,063636	0,718383
3,868181	0,110858	1,009091	0,447815
4,036364	0,113395	2,354545	0,460619
0,168182	0,122655	1,177273	0,464182
2,018182	0,123012	0,336363	0,50837
0,672728	0,140816	0,504545	0,515573
2,690909	0,248955	0,840909	0,518642
2,522727	0,252809	3,868181	0,523756
4,709091	0,267014	0,672728	0,525888
3,363636	0,270132	3,7	0,535011
6,727273	0,273323	4,204546	0,53748
6,054545	0,282188	6,222728	0,564866
2,859091	0,282605	6,559091	0,583061
6,559091	0,291241	3,531818	0,585896
5,886363	0,30606	2,522727	0,600383
6,222728	0,311093	4,540909	0,606183
3,027273	0,312837	6,390909	0,609316
3,195455	0,315252	3,363636	0,60973
6,895454	0,321926	2,690909	0,622973
5,55	0,334944	4,877272	0,62401
5,045455	0,34391	4,709091	0,624184
4,877272	0,350406	2,859091	0,624865
8,409091	0,362058	3,195455	0,63469
5,381818	0,363705	5,718182	0,636022
5,718182	0,370022	6,054545	0,648151
7,063636	0,382344	1,681819	0,351117
8,240909	0,413256	0	0,354079

8,072727	0,426519	0,168182	0,383487
7,231819	0,453037	1,85	0,386482
7,904545	0,482276	2,018182	0,390141
7,736364	0,484801	1,345454	0,398474
7,4	0,50164	1,513636	0,401233
7,568182	0,550972	2,186364	0,419636
0,336363	0,145201	4,036364	0,445616
2,354545	0,162325	3,027273	0,664423
1,009091	0,165158	8,240909	0,742959
0,504545	0,16808	8,072727	0,750651
4,204546	0,191173	7,231819	0,767444
0,840909	0,200926	7,4	0,777356
4,540909	0,218025	7,904545	0,857802
4,372727	0,220508	7,568182	0,868987
3,531818	0,236922	7,736364	0,881827

Решение данной задачи продемонстрируем с использованием пакета для всестороннего статистического анализа и нейросетевых исследований «STATISTICA Neural Networks».

Создадим область данных для обучающей выборки, выполнив последовательность действий *File->New->Data Set*.

Поскольку, согласно условию, имеется два входных параметра (объем продаж и прибыль) и один выходной параметр (рентабельность), то заполним форму «Create Data Set» как изображено на Рис. 12.

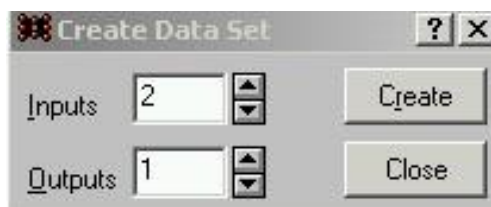


Рис. 12. Форма ввода параметров области данных для выборки

Дополним открывшуюся после нажатия на кнопку «Create» форму «Data Set Edition (new)» пустыми строками для заполнения данными обучающей выборки, выполнив последовательность действий *Edit->Cases->Add* и заполним форму «Add Cases» как изображено на Рис. Ошибка! Текст указанного стиля в документе отсутствует.**13**, поскольку в исходной выборке имеется 102 строки данных (96 исходные и 6 экспериментальные).

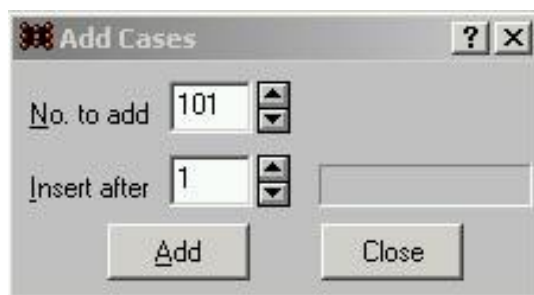


Рис. 13. Форма ввода количества добавляемых строк области данных

Преобразуем данные таблицы 1. к виду, представленному в таблице 2, предварительно заменив знак «,» на «.», поскольку в «STATISTICA Neural Networks» в качестве разделителя используется знак «.».

Преобразованная выборка данных представлена в табл. 2.

Таблица 2.

Объем продаж	Прибыль	Рентабельность
1.681819	0.030908	рент.
1.85	0.034141	рент.
1.513636	0.035622	рент.
1.345454	0.042527	рент.
2.186364	0.075902	рент.
0	0.08062	рент.
1.177273	0.097971	рент.
3.868181	0.110858	рент.
4.036364	0.113395	рент.
0.168182	0.122655	рент.
2.018182	0.123012	рент.
0.672728	0.140816	рент.
2.690909	0.248955	рент.
2.522727	0.252809	рент.
4.709091	0.267014	рент.
3.363636	0.270132	рент.
6.727273	0.273323	рент.

Объем продаж	Прибыль	Рентабельность
6.054545	0.282188	рент.
2.859091	0.282605	рент.
6.559091	0.291241	рент.
5.886363	0.30606	рент.
6.222728	0.311093	рент.
3.027273	0.312837	рент.
3.195455	0.315252	рент.
6.895454	0.321926	рент.
5.55	0.334944	рент.
5.045455	0.34391	рент.
4.877272	0.350406	рент.
8.409091	0.362058	рент.
5.381818	0.363705	рент.
5.718182	0.370022	рент.
7.063636	0.382344	рент.
8.240909	0.413256	рент.
8.072727	0.426519	рент.
7.231819	0.453037	рент.
7.904545	0.482276	рент.
7.736364	0.484801	рент.
7.4	0.50164	рент.
7.568182	0.550972	рент.
0.336363	0.145201	рент.
2.354545	0.162325	рент.

Объем продаж	Прибыль	Рентабельность
1.009091	0.165158	рент.
0.504545	0.16808	рент.
4.204546	0.191173	рент.
0.840909	0.200926	рент.
4.540909	0.218025	рент.
4.372727	0.220508	рент.
3.531818	0.236922	рент.
5.886363	0.665634	не рент.
5.213637	0.680305	не рент.
5.55	0.697755	не рент.
6.895454	0.698502	не рент.
5.045455	0.701284	не рент.
8.409091	0.707762	не рент.
7.063636	0.718383	не рент.
1.009091	0.447815	не рент.
2.354545	0.460619	не рент.
1.177273	0.464182	не рент.
0.336363	0.50837	не рент.
0.504545	0.515573	не рент.
0.840909	0.518642	не рент.
3.868181	0.523756	не рент.
0.672728	0.525888	не рент.

Объем продаж	Прибыль	Рентабельность
3.7	0.535011	не рент.
4.204546	0.53748	не рент.
6.222728	0.564866	не рент.
6.559091	0.583061	не рент.
3.531818	0.585896	не рент.
2.522727	0.600383	не рент.
4.540909	0.606183	не рент.
6.390909	0.609316	не рент.
3.363636	0.60973	не рент.
2.690909	0.622973	не рент.
4.877272	0.62401	не рент.
4.709091	0.624184	не рент.
2.859091	0.624865	не рент.
3.195455	0.63469	не рент.
5.718182	0.636022	не рент.
6.054545	0.648151	не рент.
1.681819	0.351117	не рент.
0	0.354079	не рент.
0.168182	0.383487	не рент.
1.85	0.386482	не рент.
2.018182	0.390141	не рент.
1.345454	0.398474	не рент.

Объем продаж	Прибыль	Рентабельность
1.513636	0.401233	не рент.
2.186364	0.419636	не рент.
4.036364	0.445616	не рент.
3.027273	0.664423	не рент.
8.240909	0.742959	не рент.
8.072727	0.750651	не рент.
7.231819	0.767444	не рент.
7.4	0.777356	не рент.
7.904545	0.857802	не рент.
7.568182	0.868987	не рент.
7.736364	0.881827	не рент.
6.727273	0.623579	
4.372727	0.554121	
5.381818	0.768088	
3.7	0.137848	
5.213637	0.440325	
6.390909	0.259315	

Назначим для столбца «VAR3», являющегося выходным параметром (рентабельность) обучающей выборки, возможные значения «рент.» и «не рент.». Для этого выделим данный столбец, выбрав пункт контекстного меню *Definition* (Рис. 14) и заполним форму «Variable Definition» для двух возможных значений выходного параметра «VAR3» как изображено на Рис. 15 Рис. 16.

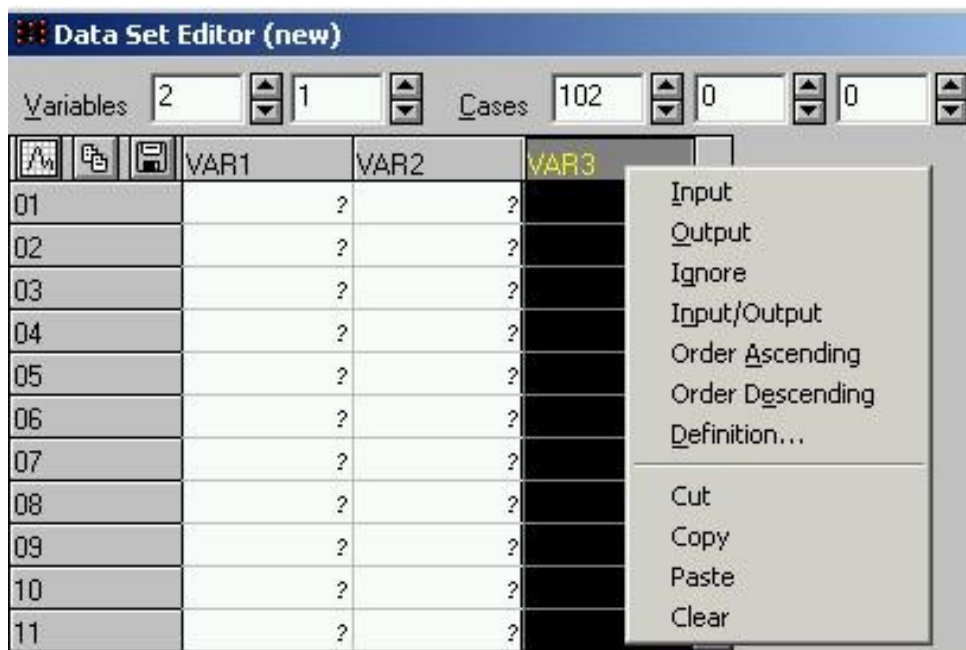


Рис. 14. Назначение возможных значений выходного параметра «VAR3»

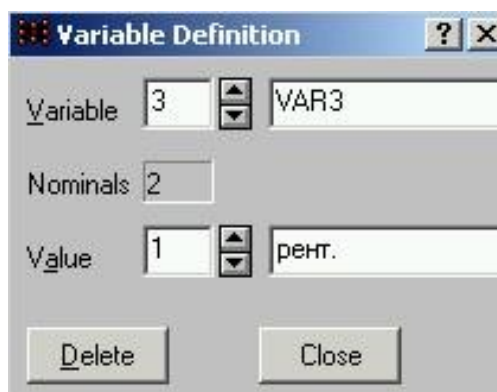


Рис. 15. Заполнение 1-го возможного значения параметра «VAR3»

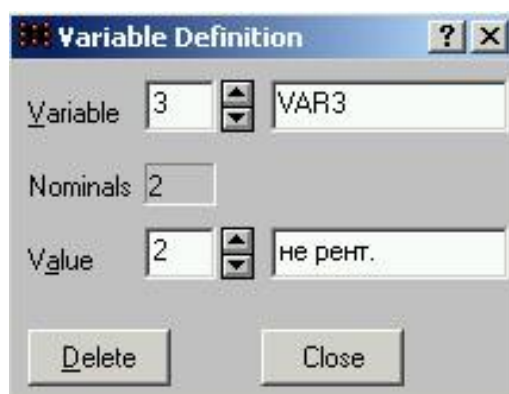


Рис. 16. Заполнение 2-го возможного значения параметра «VAR3»

Выполним копирование данных обучающей выборки таблица 2. в форму «Data Set Edition (new)» (Рис. 17).

	VAR1	VAR2	VAR3
01	1.681819	0.030908	рент.
02	1.85	0.034141	рент.
03	1.513636	0.035622	рент.
04	1.345454	0.042527	рент.
05	2.186364	0.075902	рент.
06	0	0.08062	рент.
07	1.177273	0.097971	рент.
08	3.868181	0.110858	рент.
09	4.036364	0.113395	рент.
10	0.168182	0.122655	рент.
11	2.018182	0.123012	рент.

Рис. 17. Исходная выборка данных

Пометим строки обучающей выборки с 97-102 как игнорируемые (пункт *Ignore* контекстного меню) в процессе обучения нейронной сети (Рис. 18), поскольку они будут использоваться для предсказания рентабельности после обучения нейронной сети.

	VAR1	VAR2	VAR3
92	7.231819	0.767444	не рент.
93	7.4	0.777356	не рент.
94	7.904545	0.857802	не рент.
95	7.568182	0.868987	не рент.
96	7.736364	0.881827	не рент.
97	7.273	0.623579	?
98	7.2727	0.554121	?
99	7.1818	0.768088	?
100	7.37	0.137848	?
101	7.3637	0.440325	?
102	7.0909	0.259315	?

Рис. 18. Пометка строк исходной обучающей выборки с 97-102 как игнорируемых

Пометим 20% строк обучающей выборки как «проверочные» (пункт *Verification* контекстного меню). Для этого уменьшим на ~20% количество

«обучающих» строк выборки

Перераспределим «проверочные» и «обучающие» строки обучающей выборки, выпол-

ним последовательность действий Edit->Cases->Shuffle->Train and Verify.

Создадим нейронную сеть, выполнив последовательность действий File->New->Network.

Заполним форму «Create Network» в соответствии с Рис. Ошибка! Текст указанного стиля в документе отсутствует..19 и создадим сеть (Рис. 20) нажав кнопку «Create».

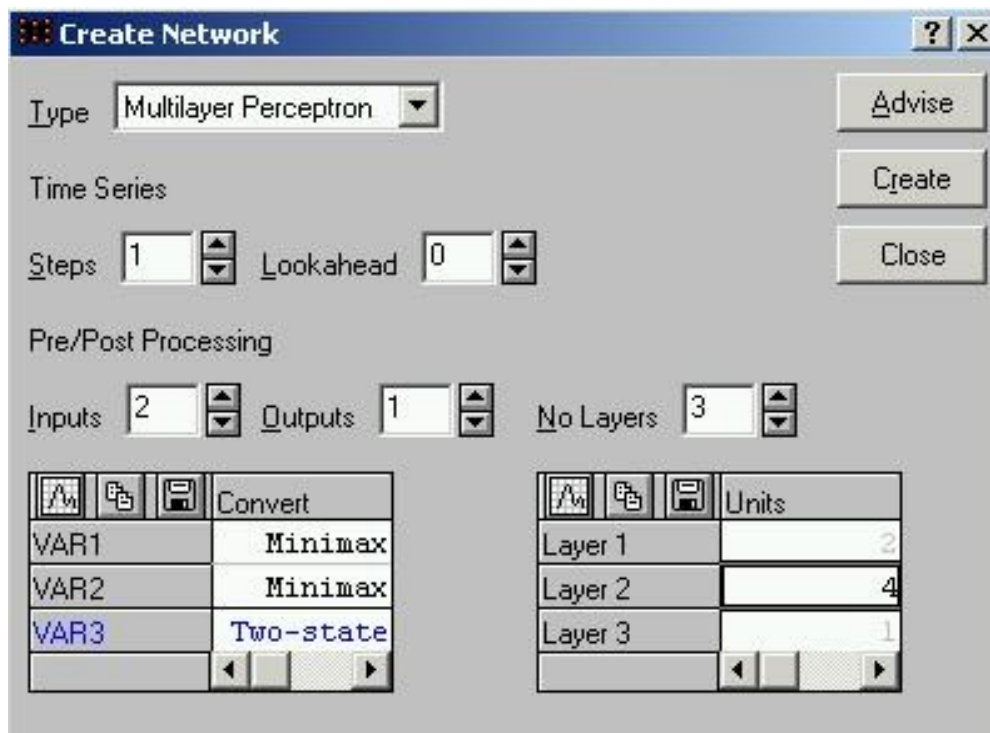


Рис. 19. Создание нейронной сети

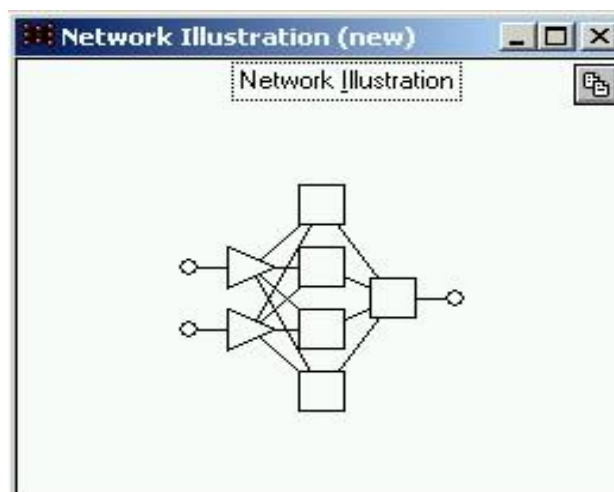


Рис. 20. Графическое представление созданной нейронной сети

Обучаем полученную нейронную сеть, используя различные варианты обучения, представленные пунктом меню Train->Multilayer Perceptrons (Рис. 21).

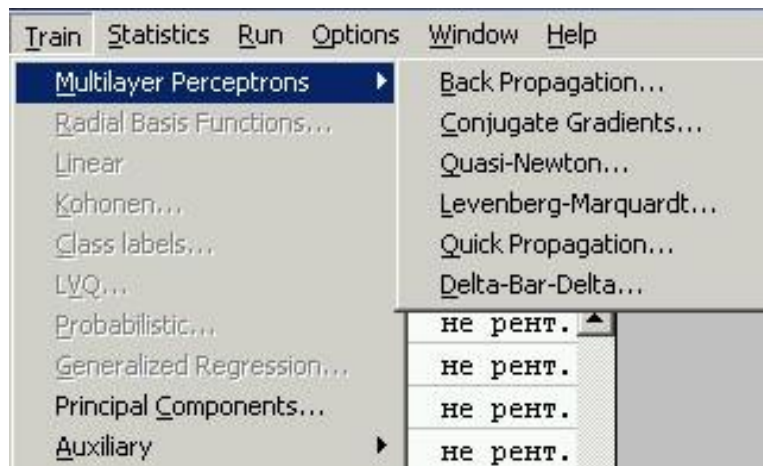


Рис. 21. Обучение нейронной сети

Для проверки уровня обучения сети удобно использовать форму «Training Error Graph», вызываемую *Statistics* -> *Training Graph*. Особое внимание на этой форме следует обратить на количественному значению ошибки проверки нейронной сети, которое не должно превосходить значение 0.001.

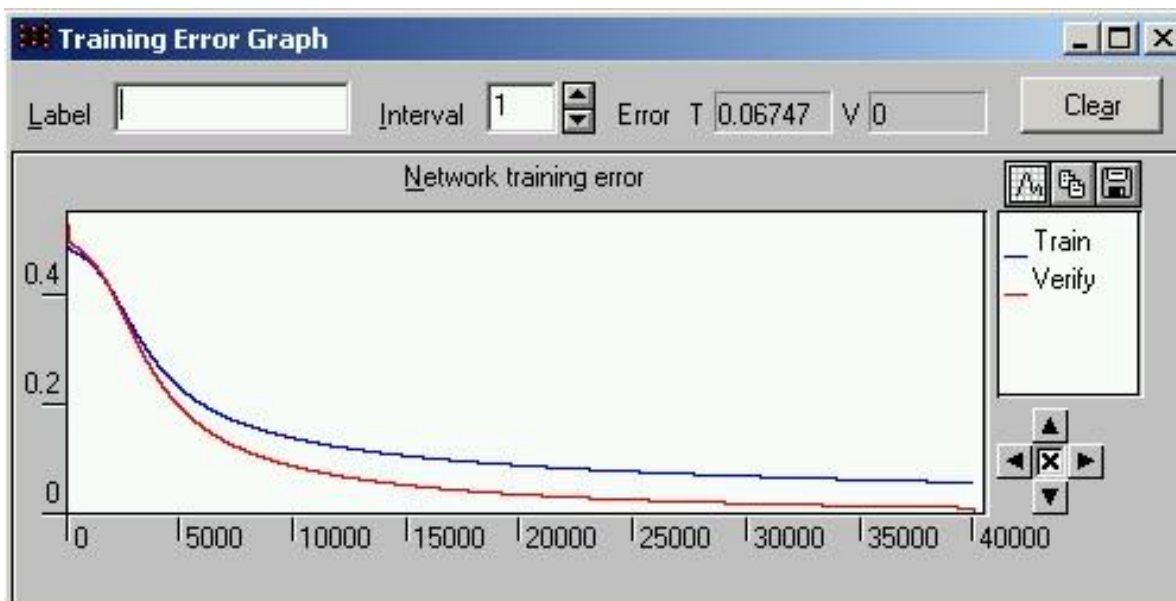


Рис. 22. Форма проверки уровня обучения сети

После успешного обучения нейронной сети (ошибка проверки нейронной сети стремится к 0) для определения какие из потребсоюзов окажутся рентабельными, а какие нерентабельными воспользуемся формой «Run Single Case» для строк 97-102, содержащих данные прогноза развития потребсоюзов на текущий год, выполнив для этого последовательность действий *Run*->*Single Case* (Рис. 23).

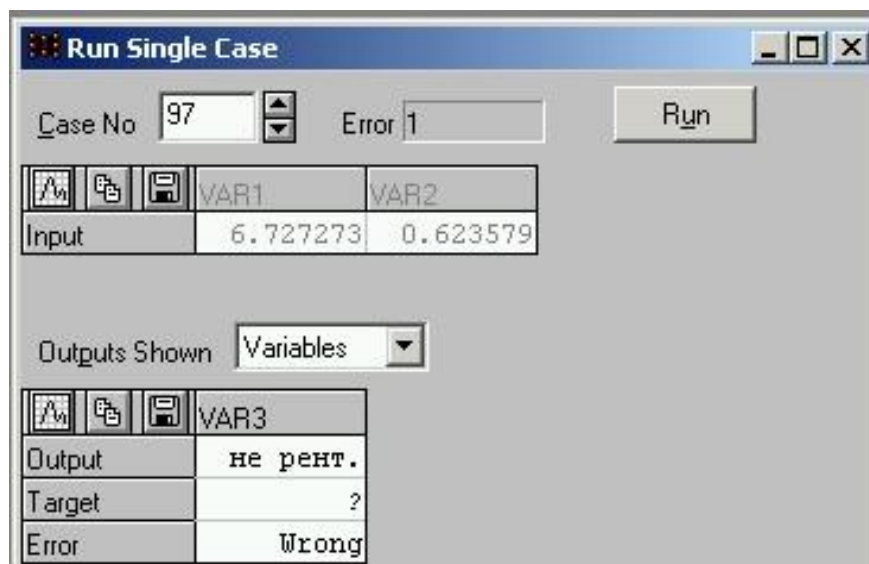


Рис. 23. Форма обработки данных обученной нейронной сетью

Используя значения «Output» выдаваемые нейронной сетью в качестве прогнозируемых для строк от 97-102 получаем, что первые 3 потребсоюза окажутся не рентабельными, а 4-6 потребсоюза будут рентабельными (Таблица 3).

Прогноз рентабельности потребсоюзов

Таблица 3

№ потребсоюза	Объем продаж	Прибыль	Рентабельность
1	6.727273	0.623579	не рент.
2	4.372727	0.554121	не рент.
3	5.381818	0.768088	не рент.
4	3.7	0.137848	рент.
5	5.213637	0.440325	рент.
6	6.390909	0.259315	рент.

Лабораторная работа № 6. Графика в Python

Краткие сведения

Python - высокоуровневый язык программирования общего назначения с динамической строгой типизацией и автоматическим управлением памятью, ориентированный на повышение производительности разработчика, читаемости кода и его качества, а также на обеспечение переносимости написанных на нём программ. Язык является полностью объектно-ориентированным в том плане, что всё является объектами.

Сам же язык известен как интерпретируемый и используется в том числе для написания скриптов. Недостатками языка являются зачастую более низкая скорость работы и более высокое потребление памяти написанных на нём программ по сравнению с аналогичным кодом,

написанным на компилируемых языках, таких как Си.

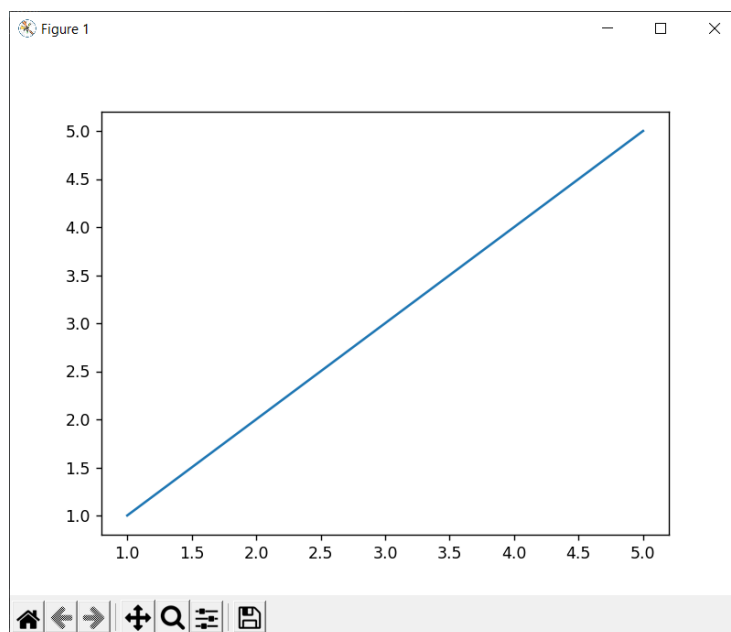
Одной из особенностей языка является наличие большого количества сторонних библиотек, содержащих большой набор полезных переносимых функций, начиная от функционала для работы с текстом и заканчивая средствами для написания сетевых приложений. Одной из таких библиотек является *Matplotlib*. С ее помощью можно решить большую часть задач, связанную с визуализацией обрабатываемых данных.

Для получения общего представления о возможностях графических представлений рассмотрим следующие примеры. При их выполнении предполагается, что обучаемые обладают как начальными навыками работы в среде Python, так и некоторым опытом программирования в указанной среде.

Пример 1

```
# Построение графика
import matplotlib.pyplot as plt
plt.plot([1, 2, 3, 4, 5], [1, 2, 3, 4, 5])
plt.show()
```

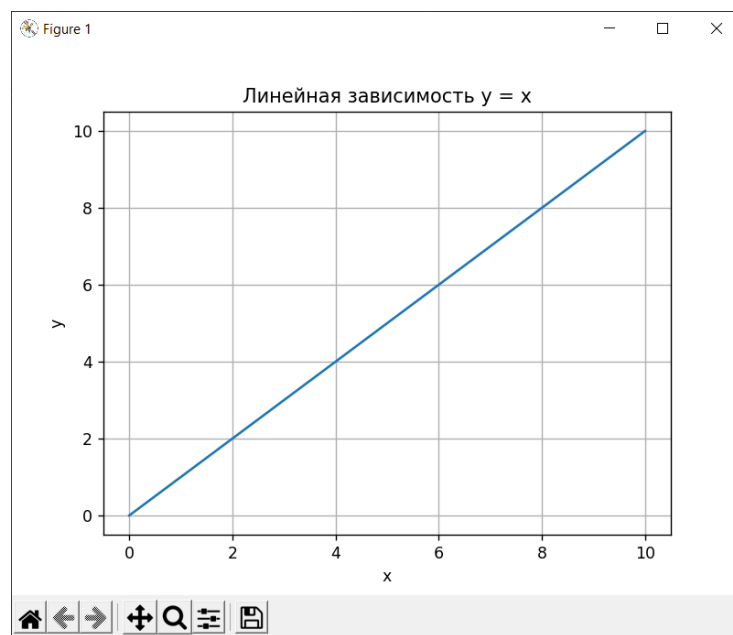
Результат работы программы показан на рисунке.



Пример 2

```
# График с элементами оформления
import matplotlib.pyplot as plt
import numpy as np
# Независимая (x) и зависимая (y) переменные
x = np.linspace(0, 10, 50)
y = x
# Построение графика
plt.title('Линейная зависимость y = x') # заголовок
plt.xlabel('x') # ось абсцисс
plt.ylabel('y') # ось ординат
plt.grid() # включение отображение сетки
plt.plot(x, y) # построение графика
plt.show()
```

Результат работы программы показан на рисунке.



Пример 3

Два графика на одной диаграмме

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

Линейная зависимость

```
x = np.linspace(0, 10, 50)
```

```
y1 = x
```

Квадратичная зависимость

```
y2 = [i**2 for i in x]
```

Построение графика

```
plt.title('Зависимости: y1 = x, y2 = x^2')
```

заголовок

```
plt.xlabel('x') # ось абсцисс
```

```
plt.ylabel('y1, y2') # ось ординат
```

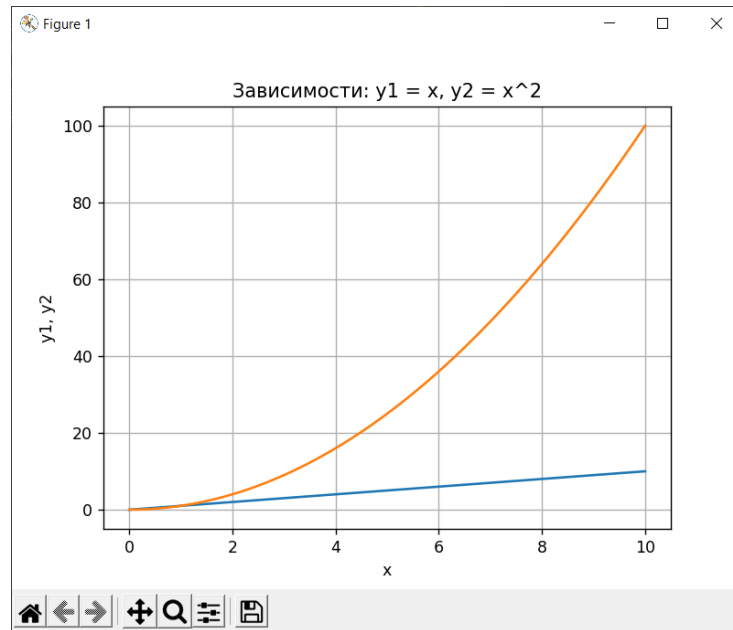
```
plt.grid() # включение отображение сетки
```

```
plt.plot(x, y1, x, y2)
```

построение графика

```
plt.show()
```

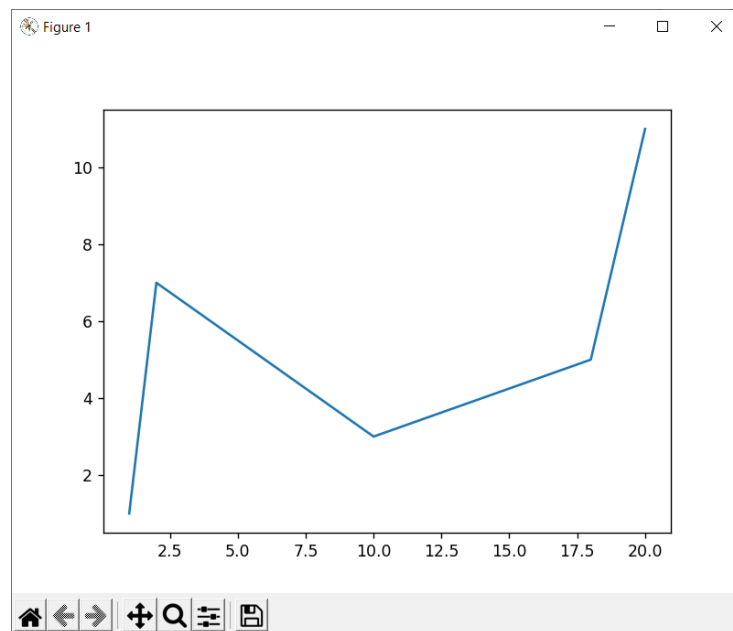
Результат работы программы показан на рисунке.



Пример 4

```
# неравномерный график
import matplotlib.pyplot as plt
plt.plot([1, 2, 10, 18, 20], [1, 7, 3, 5, 11])
plt.show()
```

Результат работы программы показан на рисунке.

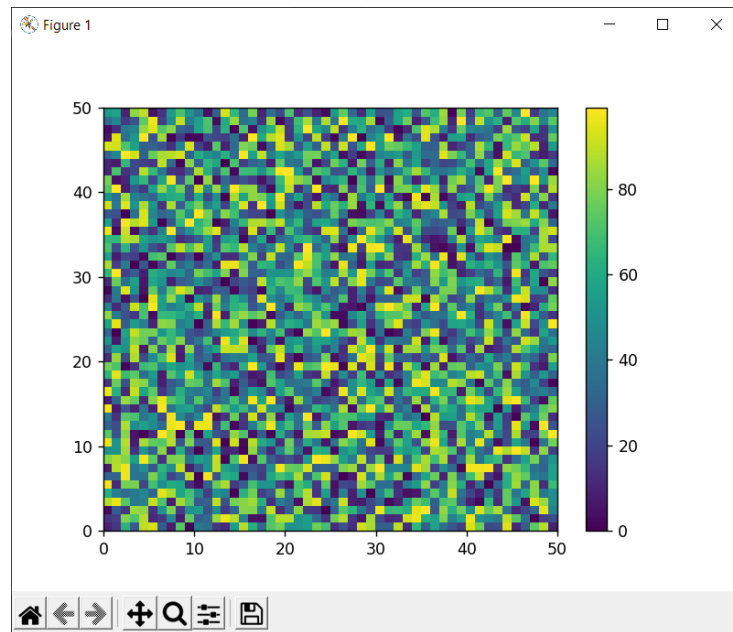


Пример 5

```
# цветовая карта
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(100)
vals = np.random.randint(10, size=(50, 50))
plt.pcolor(vals)
plt.colorbar()
```


plt.show()

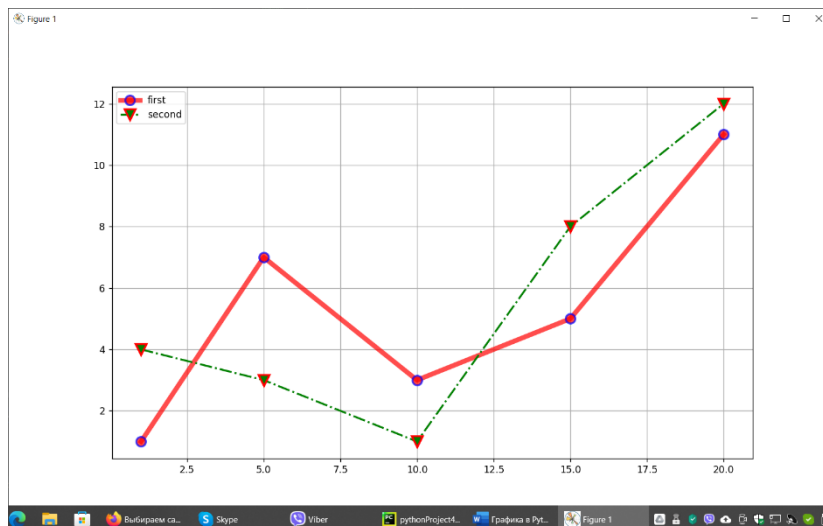
Результат работы программы показан на рисунке.



Пример 6

```
# два графика
import matplotlib.pyplot as plt
x = [1, 5, 10, 15, 20]
y1 = [1, 7, 3, 5, 11]
y2 = [4, 3, 1, 8, 12]
plt.figure(figsize=(12, 7))
plt.plot(x, y1, 'o-r', alpha=0.7, label='first', lw=5, mec='b', mew=2, ms=10)
plt.plot(x, y2, 'v-g', label='second', mec='r', lw=2, mew=2, ms=12)
plt.legend()
plt.grid(True)
plt.show()
```

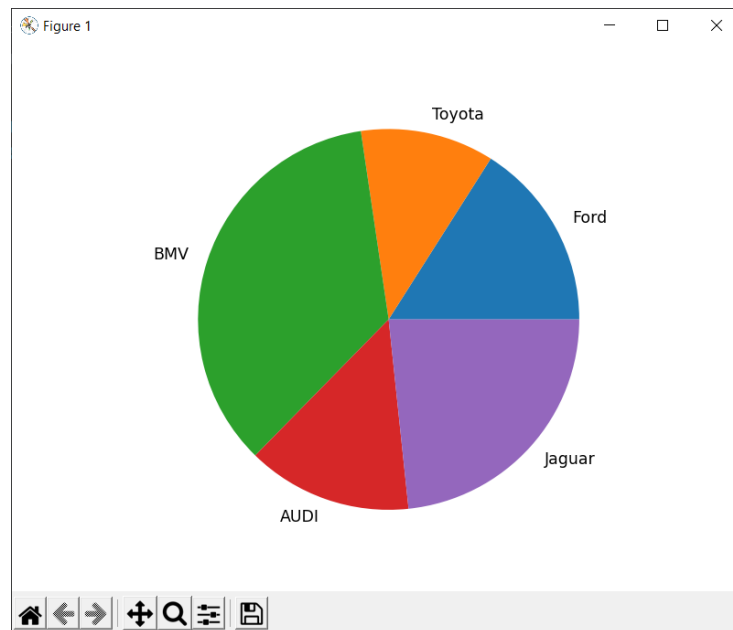
Результат работы программы показан на рисунке.



Пример 7

```
# круговая диаграмма
import matplotlib.pyplot as plt
vals = [24, 17, 53, 21, 35]
labels = ['Ford', 'Toyota', 'BMV', 'AUDI', 'Jaguar']
fig, ax = plt.subplots()
ax.pie(vals, labels=labels)
ax.axis('equal')
plt.show()
```

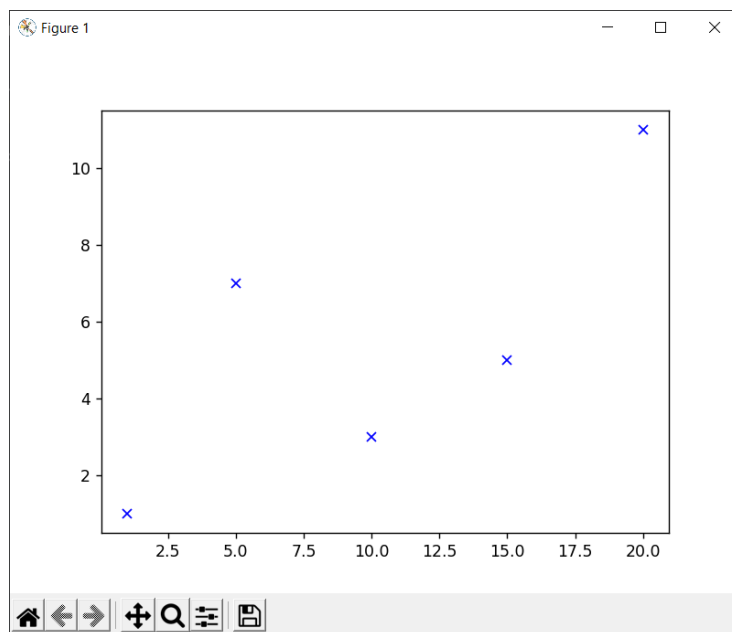
Результат работы программы показан на рисунке.



Пример 8

```
# двумерный точечный график
import matplotlib.pyplot as plt
x = [1, 5, 10, 15, 20]
y = [1, 7, 3, 5, 11]
# plt.plot(x,y,'--r')
# plt.plot(x, y, 'ro')
plt.plot(x, y, 'bx')
plt.show()
```

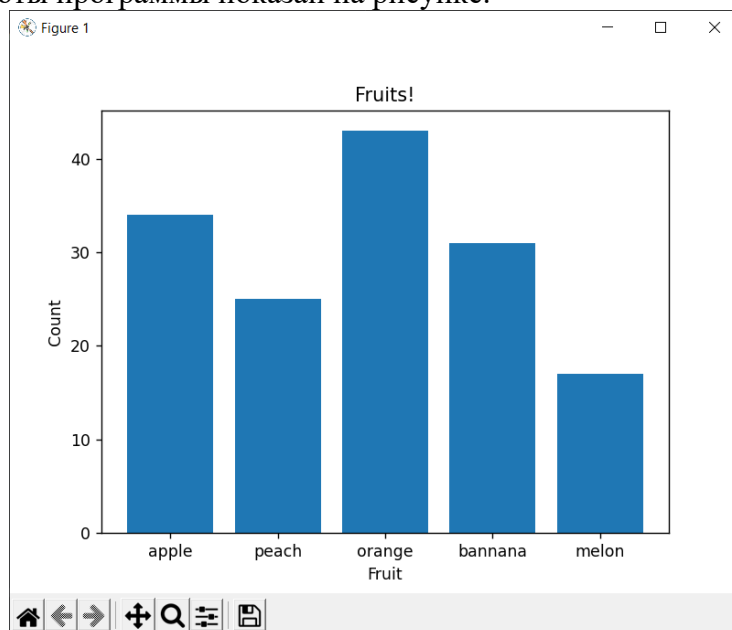
Результат работы программы показан на рисунке.



Пример 9

```
# гистограмма
import matplotlib.pyplot as plt
fruits = ['apple', 'peach', 'orange', 'bannana', 'melon']
counts = [34, 25, 43, 31, 17]
plt.bar(fruits, counts)
plt.title('Fruits!')
plt.xlabel('Fruit')
plt.ylabel('Count')
plt.show()
```

Результат работы программы показан на рисунке.



Пример 10

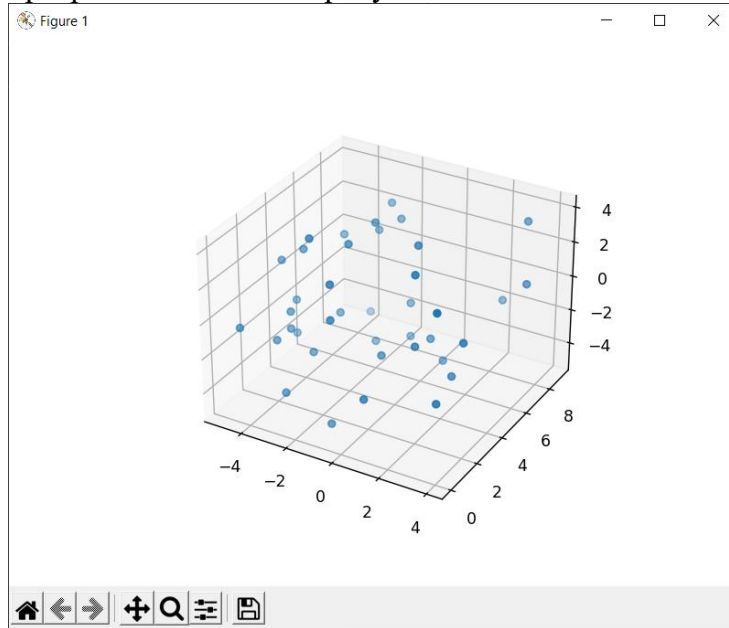
```
# трехмерный точечный график
import matplotlib.pyplot as plt #from mpl_toolkits.mplot3d
#import Axes3D
```

```

import numpy as np
np.random.seed(123)
x = np.random.randint(-5, 5, 40)
y = np.random.randint(0, 10, 40)
z = np.random.randint(-5, 5, 40)
s = np.random.randint(10, 100, 20)
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.scatter(x, y, z, s)
plt.show()

```

Результат работы программы показан на рисунке.



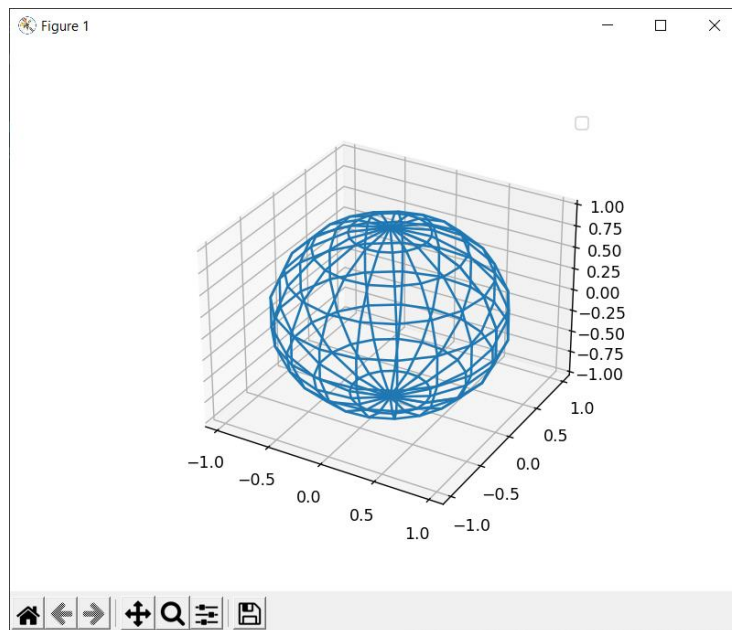
Пример 11

```

# прозрачная фигура
import matplotlib.pyplot as plt
import numpy as np
u, v = np.mgrid[0:2*np.pi:20j, 0:np.pi:10j]
x = np.cos(u)*np.sin(v)
y = np.sin(u)*np.sin(v)
z = np.cos(v)
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot_wireframe(x, y, z)
ax.legend()
plt.show()

```

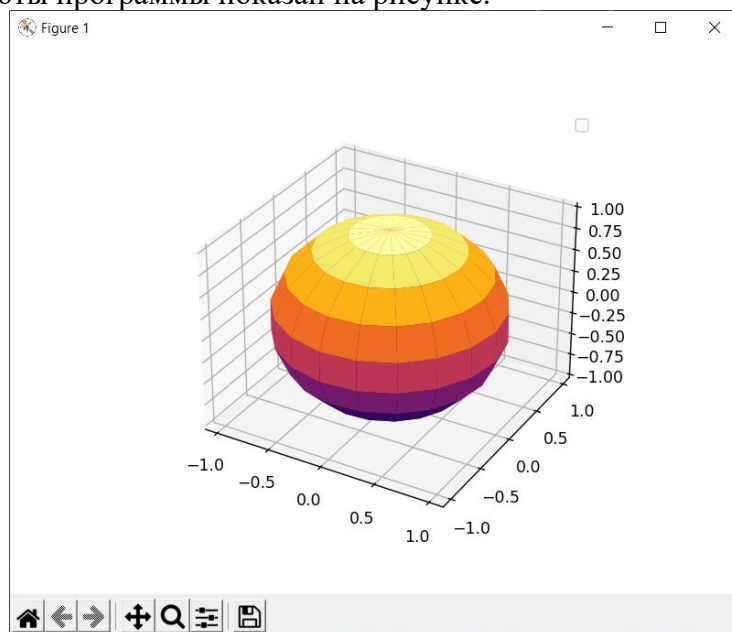
Результат работы программы показан на рисунке.



Пример 12

```
# непрозрачная фигура
import matplotlib.pyplot as plt
import numpy as np
u, v = np.mgrid[0:2*np.pi:20j, 0:np.pi:10j]
x = np.cos(u)*np.sin(v)
y = np.sin(u)*np.sin(v)
z = np.cos(v)
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
ax.plot_surface(x, y, z, cmap='inferno')
ax.legend()
plt.show()
```

Результат работы программы показан на рисунке.



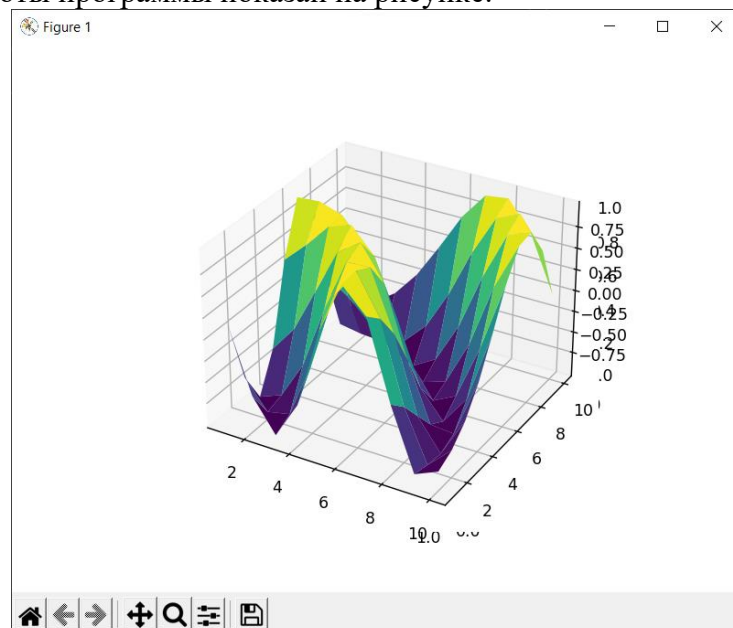
Пример 13

```

# непрозрачная поверхность
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt
# create 3d axes
fig = plt.figure()
ax = plt.axes(projection='3d')
# function for Z values
def f(x, y):
    return np.cos(np.sqrt(x ** 2 + y ** 2))
# x and y values
x = np.linspace(1, 10, 10)
y = np.linspace(1, 10, 10)
X, Y = np.meshgrid(x, y)
Z = f(X, Y)
ax = plt.axes(projection='3d')
ax.plot_surface(X, Y, Z, rstride=1, cstride=1,
               cmap='viridis')
plt.show()

```

Результат работы программы показан на рисунке.



Пример 14

```

# карскасная поверхность
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt
# create 3d axes
fig = plt.figure()
ax = plt.axes(projection='3d')
# function for Z values
def f(x, y):
    return np.sin(np.sqrt(x ** 2 + y ** 2))
# x and y values
x = np.linspace(1, 20, 10)

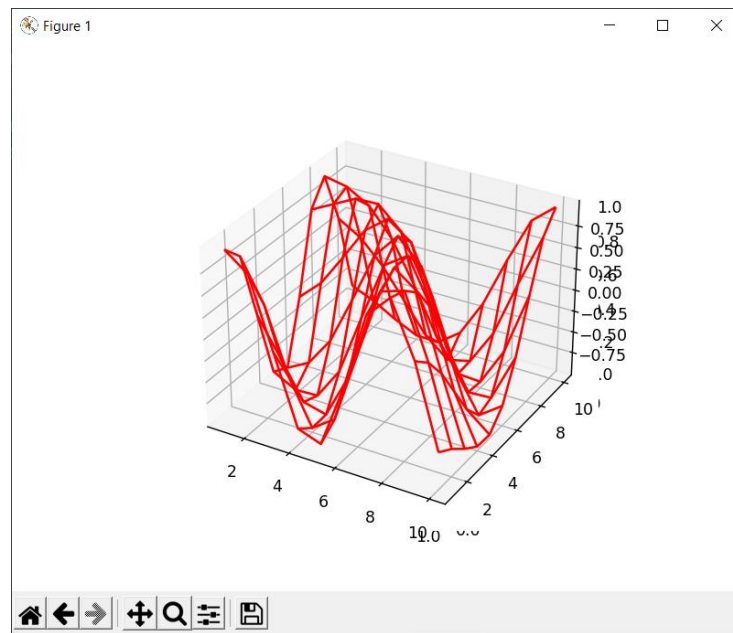
```

```

y = np.linspace(1, 20, 10)
X, Y = np.meshgrid(x, y)
Z = f(X, Y)
ax = plt.axes(projection='3d')
ax.plot_wireframe(X, Y, Z, color='red')
plt.show()

```

Результат работы программы показан на рисунке.



Лабораторная работа № 7 Методы кластеризации

1. Общие понятия

Согласно [https://ru.wikipedia.org/wiki/Кластерный_анализ] кластерный анализ — многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы

Задачей кластерного анализа является разбиение множества объектов на группы в чем-то похожих объектов - кластеры.

В качестве примера на рис.1 приведены данные об объемах сельскохозяйственного и промышленного производства некоторых регионов РФ.

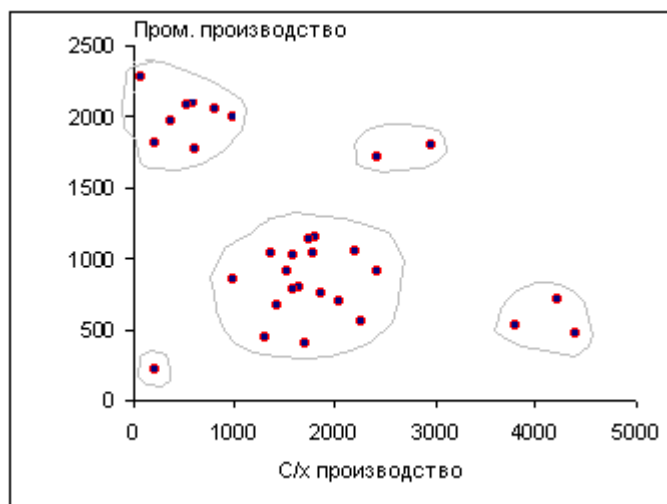


Рис.1. Графическое представление данных об объемах промышленного и с/хозяйственного производства некоторых регионов РФ

Основываясь на визуальных впечатлениях несложно "вручную" произвести группировку (кластеризацию) объектов (на рис. 5.1 обведены линиями).

Подобная группировка объектов основана на следующих признаках:

- к одному кластеру относятся те объекты, которые расположены достаточно близко друг к другу
- при этом каждый кластер отделен от других расстоянием значительно большим, чем типичное расстояние между объектами внутри кластера.

2. Методы работы с системой

Для выполнения работы используется файл *Кластерный 3D-анализ.xls*. Интерфейс программы после ее запуска показан на рис.2.

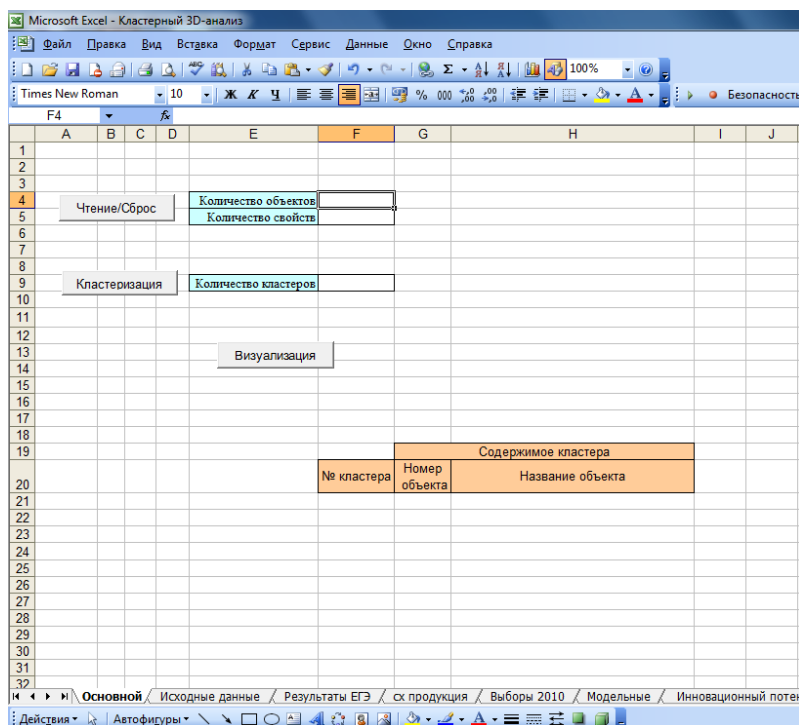


Рис.2. Интерфейс программы кластерного анализа.

Шаг 1. Загрузка данных

Скопируйте данные с листа "Модельные данные" ИЛИ Приложение 7 на лист "Исходные данные".

	A	B	C	D	E
1		x1	x2	x3	
2	Объект1	62	110	65	
3	Объект2	65	100	64	
4	Объект3	60	98	59	
5	Объект4	64	105	61	
6	Объект5	66	100	68	
7	Объект6	93	47	55	
8	Объект7	91	50	55	
9	Объект8	63	101	63	
10	Объект9	72	95	69	
11	Объект10	69	111	68	
12	Объект11	68	104	66	
13	Объект12	65	99	67	
14	Объект13	70	94	71	
15	Объект14	73	7	67	
16	Объект15	85	1	67	
17	Объект16	82	3	70	
18	Объект17	79	10	72	
19	Объект18	80	4	67	
20	Объект19	89	42	52	
21	Объект20	75	0	69	
22	Объект21	90	52	53	
23	Объект22	86	55	54	
24	Объект23	86	11	65	
25	Объект24	94	49	55	
26	Объект25	84	8	67	
27	Объект26	75	8	71	

Примечание

Если на листе "Исходные данные" уже имеются какие-то данные, то их необходимо удалить.

Перейдите на лист "Основной" и нажмите кнопку "Чтение/Сброс".

В результате в ячейке F4 должно появиться число 26 (количество объектов), а в ячейке F5 - число 3 (количество свойств).

Если появились другие числа (или ничего не появилось) то проверьте правильность размещения скопированных данных на листе "Исходные данные".

Шаг 2. Оценка количества возможных кластеров.

Для этого нажмите кнопку "Визуализация".

Запустится *Визуализатор* данных, интерфейс которого показан на рис.3.

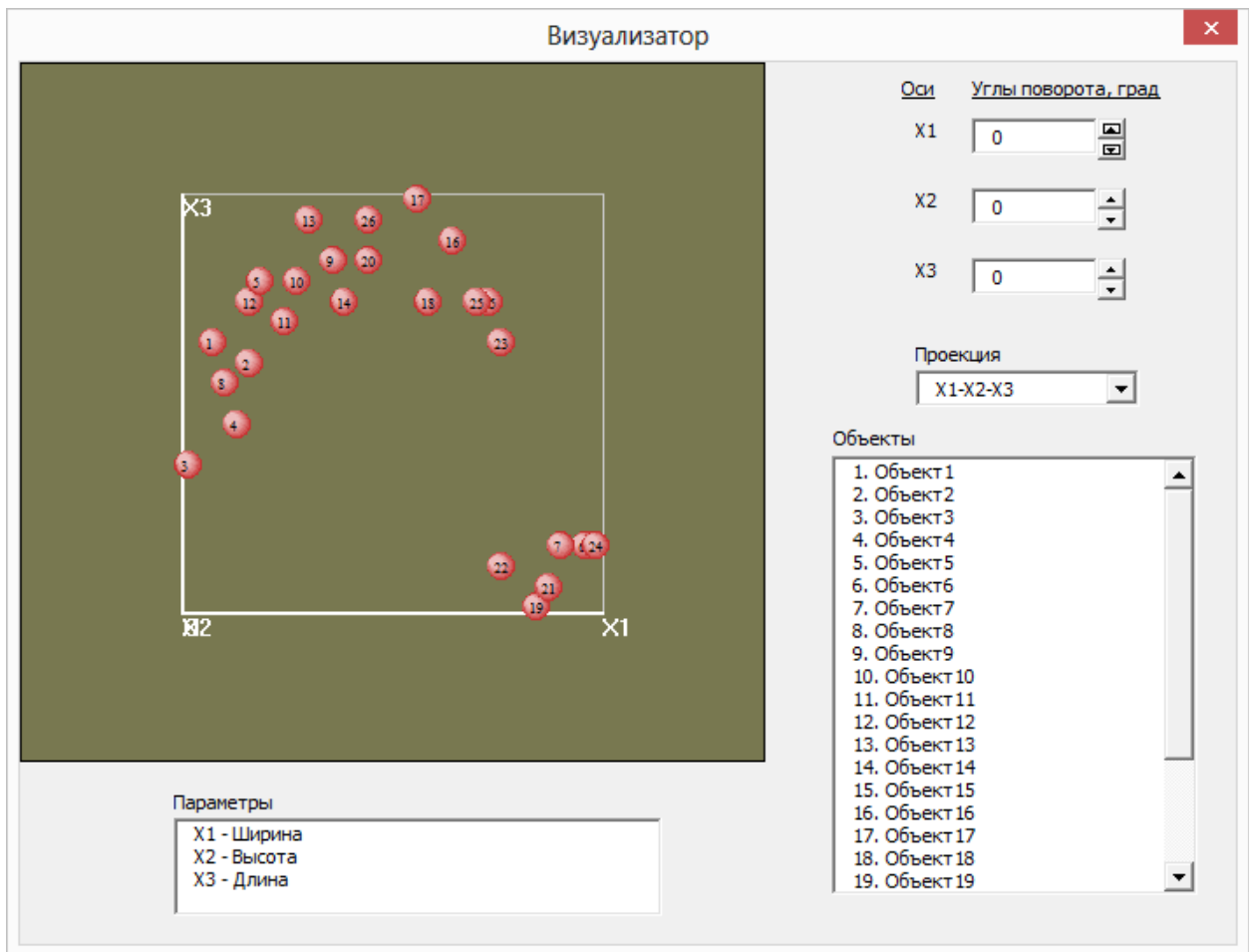



Рис. 5.3. Визуальное представление загруженных данных

С помощью кнопок  выполните вращение системы координат, так чтобы было четко видно пространственное разделение объектов на группы - примерно так, как на рис.4.

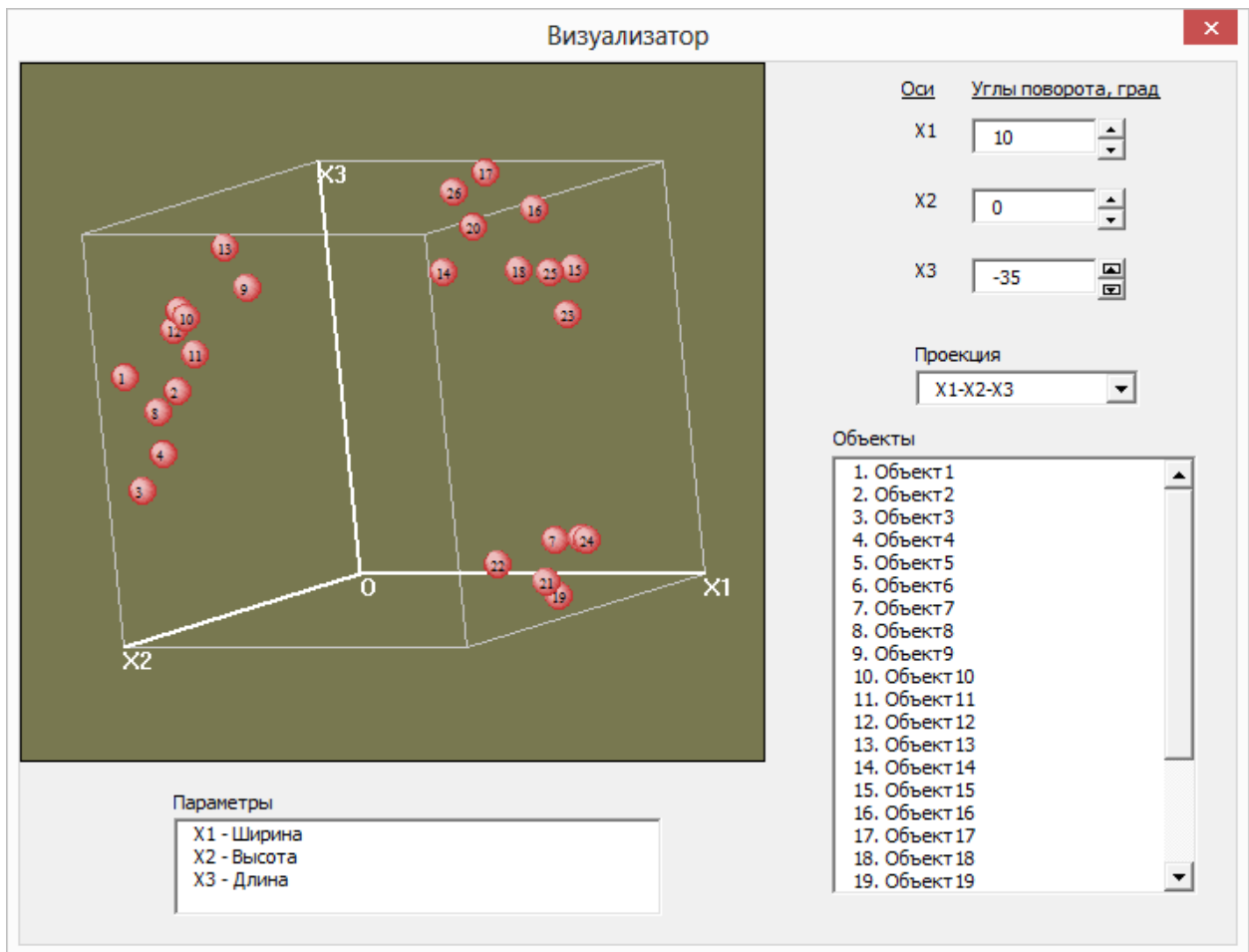


Рис.4. Визуальное представление загруженных данных после вращения координат.

Очевидно, что все множество объектов можно уверенно разделить на три группы.

Шаг 3. Кластеризация объектов

В ячейку F9 введите желаемое количество кластеров - 3 и нажмите кнопку "**Кластеризация**".

Система выполнит разделение объектов на указанное количество кластеров, распечатает это распределение и раскрасит кластеры различными цветами.

Нажмите кнопку "**Визуализация**" и выполните вращение координат до явного пространственного разделения уже раскрашенных объектов - рис.5.

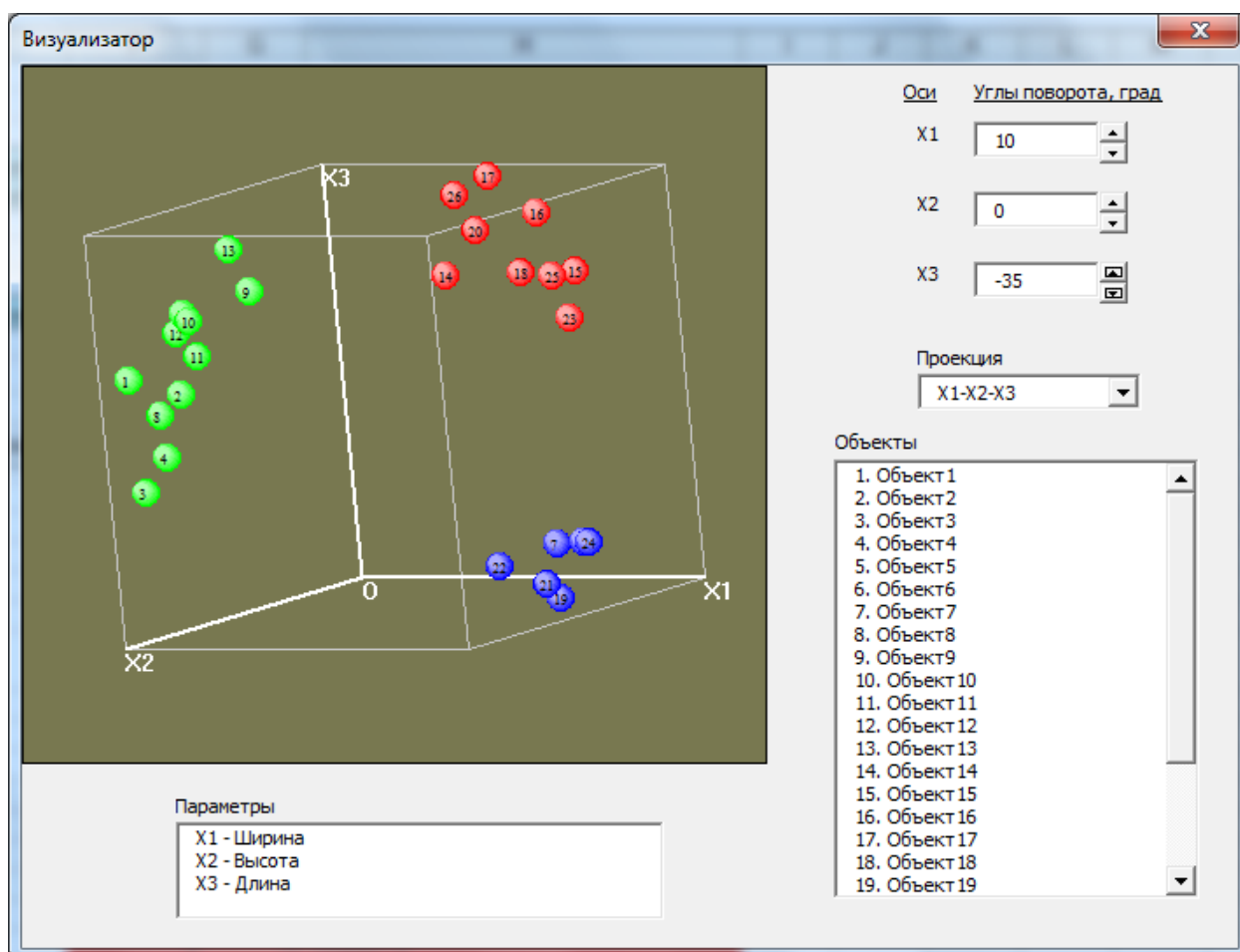


Рис.5. Визуальное представление результатов кластеризации.

Если Вас не устраивает предложенная системой цветовая гамма раскраски кластеров, то закройте *Визуализатор* и на листе "Основной" вручную перекрасьте нужные кластеры в нужные Вам цвета.

3. Основные правила кластеризации

При работе с реальными данными ситуация с четким разделением объектов на кластеры (как в разделе 5.2) является скорее исключением.

Поэтому при проведении анализа следует иметь ввиду следующие моменты:

1. Результаты анализа зависят от целей анализа
2. Различные методы кластеризации при одних и тех же исходных данных могут дать различное распределение объектов по кластерам.
3. При хорошем знании предметной области больше доверяйте своим знаниям об объектах обработки или просто здравому смыслу.

Другими словами – если результаты обработки никак не соответствуют Вашему пониманию предметной области, то тем хуже для этих результатов.

4. Человек и его зрительная аналитическая система лучше любой придуманной системы анализа. Если Ваши зрительные представления о разделении на кластеры не соответствуют полученным результатам анализа, то тем хуже для результатов.

4. Методика кластерного анализа

В общем случае кластерный анализ должен состоять из следующих этапов.

1. Определение «выбросов» данных

На рис.6 приведено окно визуализатора для некоторых данных.



Рис.6. Результаты кластерного анализа при наличии «выбросов» данных

Под «выбросами» понимаются данные, резко отличающиеся от основной группы данных. Конечно, эти данные можно рассматривать и как отдельные кластеры, состоящие из одного объекта.

Но наличие выбросов сильно искажает вид основной группы данных и не позволяет оценить ее структуру. Поэтому эти данные лучше исключить из обработки и объяснить отдельно.

2. Оценка формы кластера

В результате обработки получающиеся кластеры могут иметь самую различную форму.

В качестве примера на рис. 5.7 приведено визуальное представление некоторых данных с кластером сложной формы.

Однако по интуитивному представлению – кластер это сгусток или скопление объектов, который должен иметь шарообразную форму или, в крайнем случае, форму немного вытянутого эллипсоида.

Описанные в литературе (и особенно в интернет-публикациях) кластеры различных изощренных и причудливых форм никакого отношения к реальным данным не имеют. Это в лучшем случае результаты проверки отдельных алгоритмов на некоторых модельных данных.

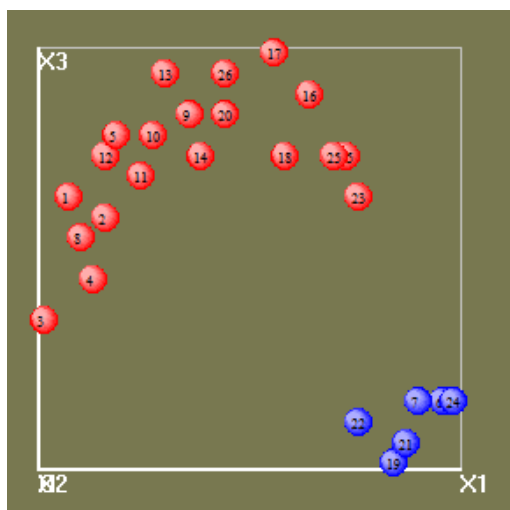


Рис.7. Пример кластера сложной формы.

Поэтому, если при анализе получается кластер слишком сложной формы, то это свидетельствует либо о неправильном выборе параметров объекта для кластеризации, либо о недостатке информации об объекте.

В примере на рис.7 достаточно информацию об объекте дополнить третьим параметром и произвести трехмерную кластеризацию, то изначальный кластер сложной формы распадается на два кластера более простой формы – рис.8.

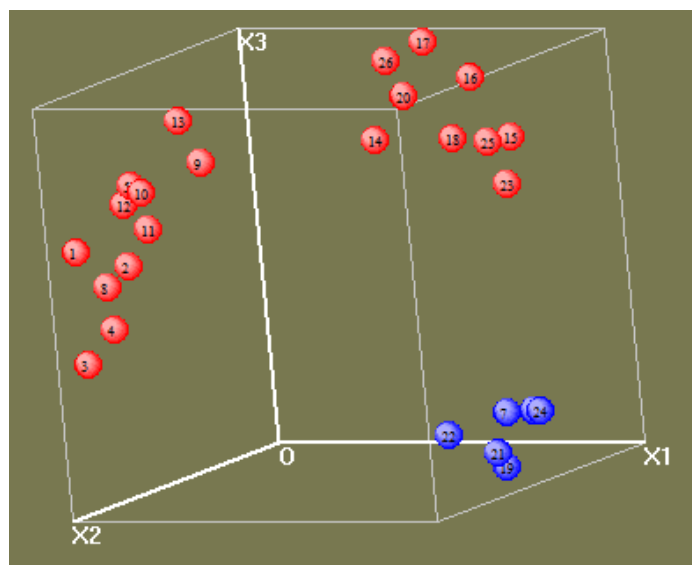


Рис.8. Распад кластера сложной формы на более простые кластеры при введении дополнительной информации об объектах.

3. *Определение уровня кластеризации*

Особенностью разработанного метода кластеризации является возможность задания желательного количества кластеров. Если обратиться к рис.5.8, то для приведенных на нем данных наиболее естественным является разбиение объектов на три кластера.

Если же для этих данных задать количество кластеров равное шести, то получим следующее визуальное представление – рис.9.

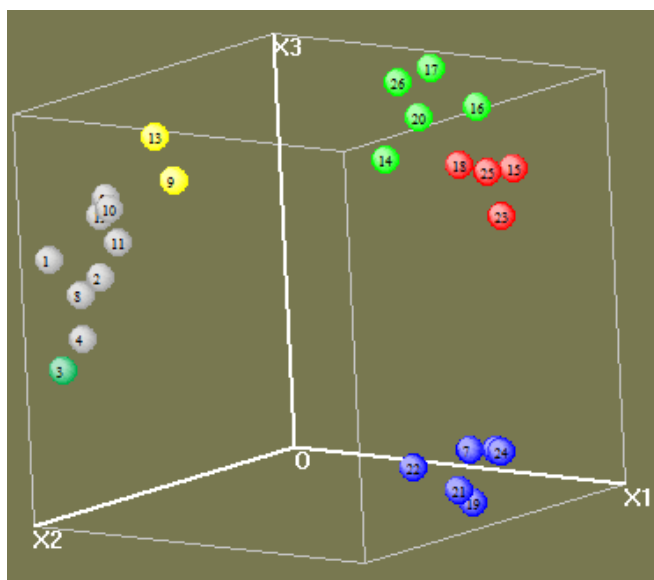


Рис.9. Разбиение объектов на шесть кластеров.

Разумеется и то, что более детальная кластеризация определяется в первую очередь целями анализа. Но основываясь на визуальном представлении эта детализация представляется избыточной.

Возможна и обратная ситуация. На рис.10 приводится визуализация разбиения объектов на восемь кластеров. Тем не менее, визуальный просмотр показывает, что этого количества недостаточно, поскольку группа синих объектов должна быть явно разделена на более мелкие группы

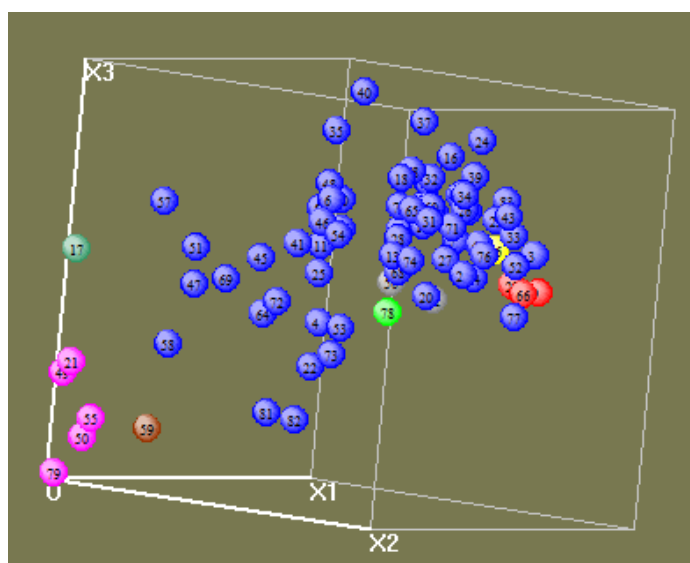


Рис.10. Недостаточный уровень кластеризации.

4. Работа с многомерными данными

Реальные данные как правило имеют размерность большую трех. Возможности же зрительного восприятия человека (как, впрочем, и компьютерных систем) ограничены тремя измерениями.

Однако и в этом случае возможна достаточно полноценная визуализация со всеми ее преимуществами. Одним из вариантов является вывод отдельных трехмерных проекций многомерного изображения.

В качестве примера на рис.11-12 приведены различные трехмерные проекции одного и того же множества объектов с одинаковым разбиением на кластеры.

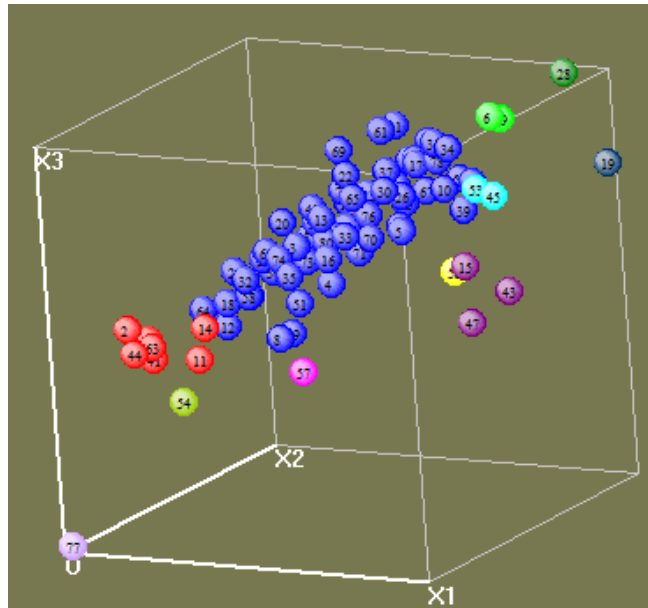


Рис.11. Проекция четырехмерного множества объектов в координатах X1-X2-X3.

На рис.11 приведена проекция четырехмерно множества объектов в координатах X1-X2-X3. Использование данной проекции не позволяет никакими вращениями визуально разделить полученные кластеры.

На рис.12 приведено то же множество, но в проекции X2-X3-X4. На данной проекции удастся подобрать такие углы поворота, при которых возможно визуальное разделение кластеров.

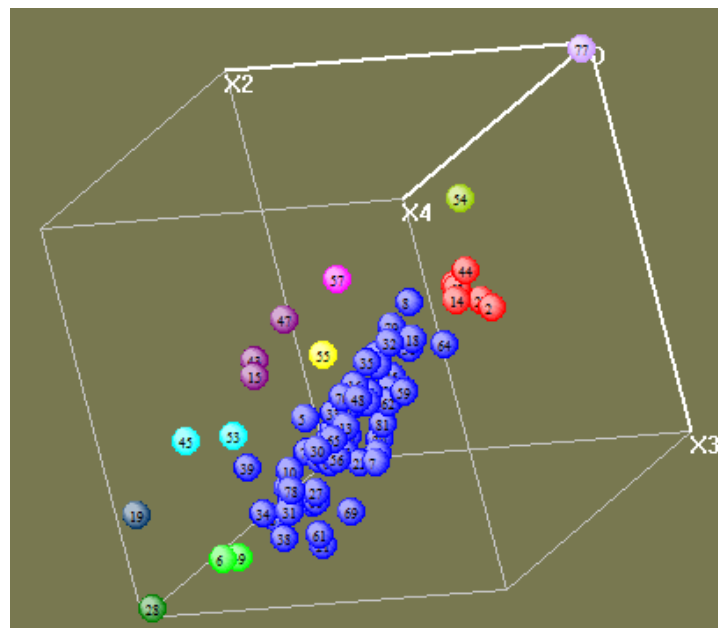


Рис.12. Проекция четырехмерного множества объектов в координатах X2-X3-X4.

5. Пример

В качестве примера рассмотрим обработку данных о производстве с/х продукции в различных регионах РФ. Согласно заданию кластеризация должна быть произведена по четырем показателям:

- производство скота и птицы;
- производство молока;
- производство яиц;

- производство картофеля.

1. Скопируем данные с листа «сх продукция» на лист «Исходные данные».

2. Перейдем на лист «Основной» и нажмем кнопку «Чтение/Сброс». В ячейках F4 и F5 должны появиться соответственно значения 70 и 4.

3. Нажмем кнопку «Визуализация». Результат визуализации приведен на рис.13.

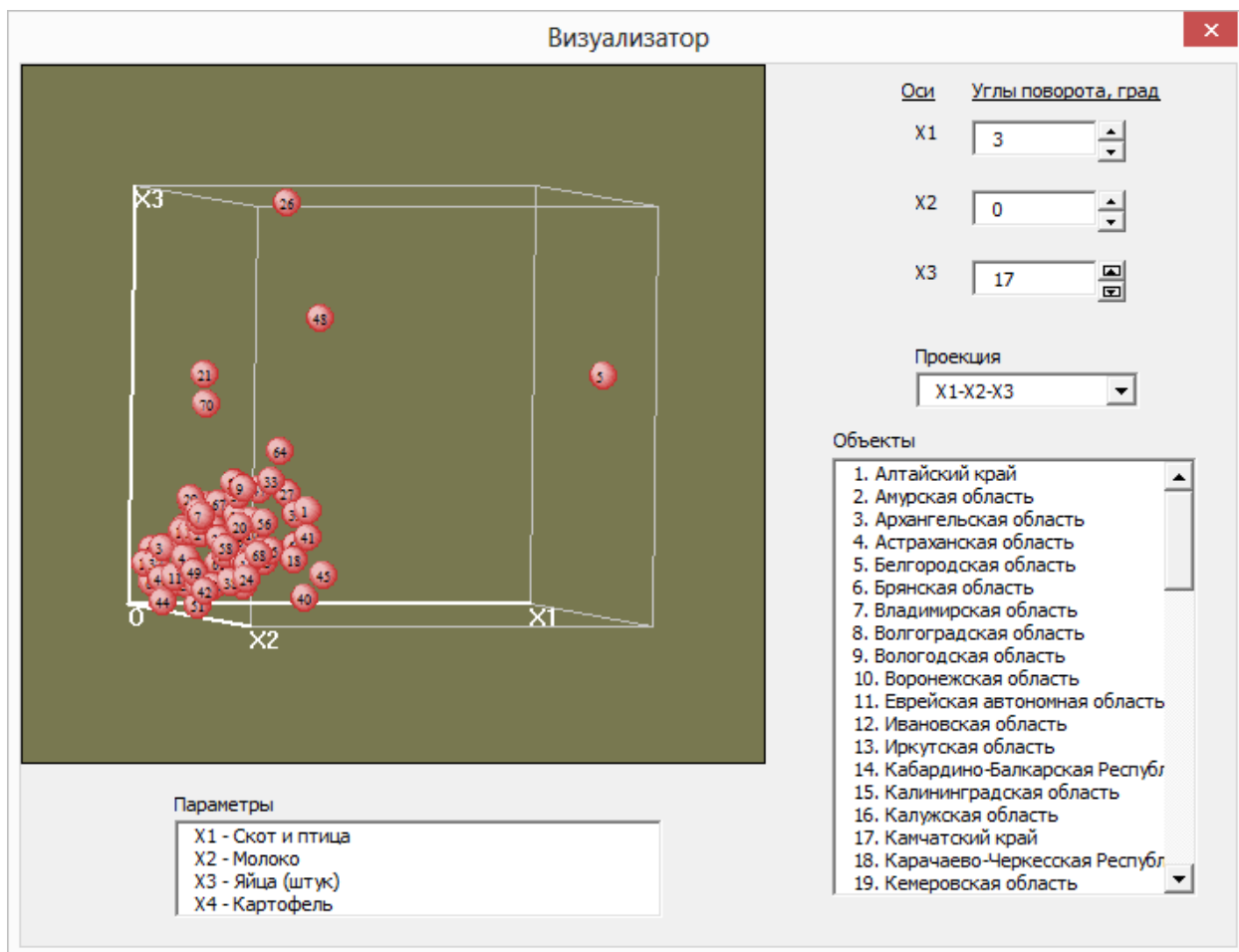


Рис.13. Визуальное представление данных об уровне выпускаемой с/х продукции различных регионов РФ.

Из представленного рисунка сразу же следует, что мы имеем 4 «выброса» из основной группы данных. Это объекты:

5 - Белгородская область;

26 – Ленинградская область;

48 – Республика Мордовия;

21 и 70 – Костромская область и Ярославская область.

Проверим наличие выбросов по четвертому параметру, которого нет на данной проекции. Для этого из списка «Проекция» выберем, например, проекцию X1-X2-X4.

На этой проекции других выбросов, кроме уже выявленных, не имеется.

4. Переходим на лист «Исходные данные» и удаляем из списка вышеуказанные объекты.

5. Переходим на лист «Основной» и нажимаем кнопку «Чтение/Сброс». В ячейках F4 и F5 должны появиться соответственно значения 65 и 4.

6. Нажимаем кнопку «Визуализация». Результат представлен на рис.14.

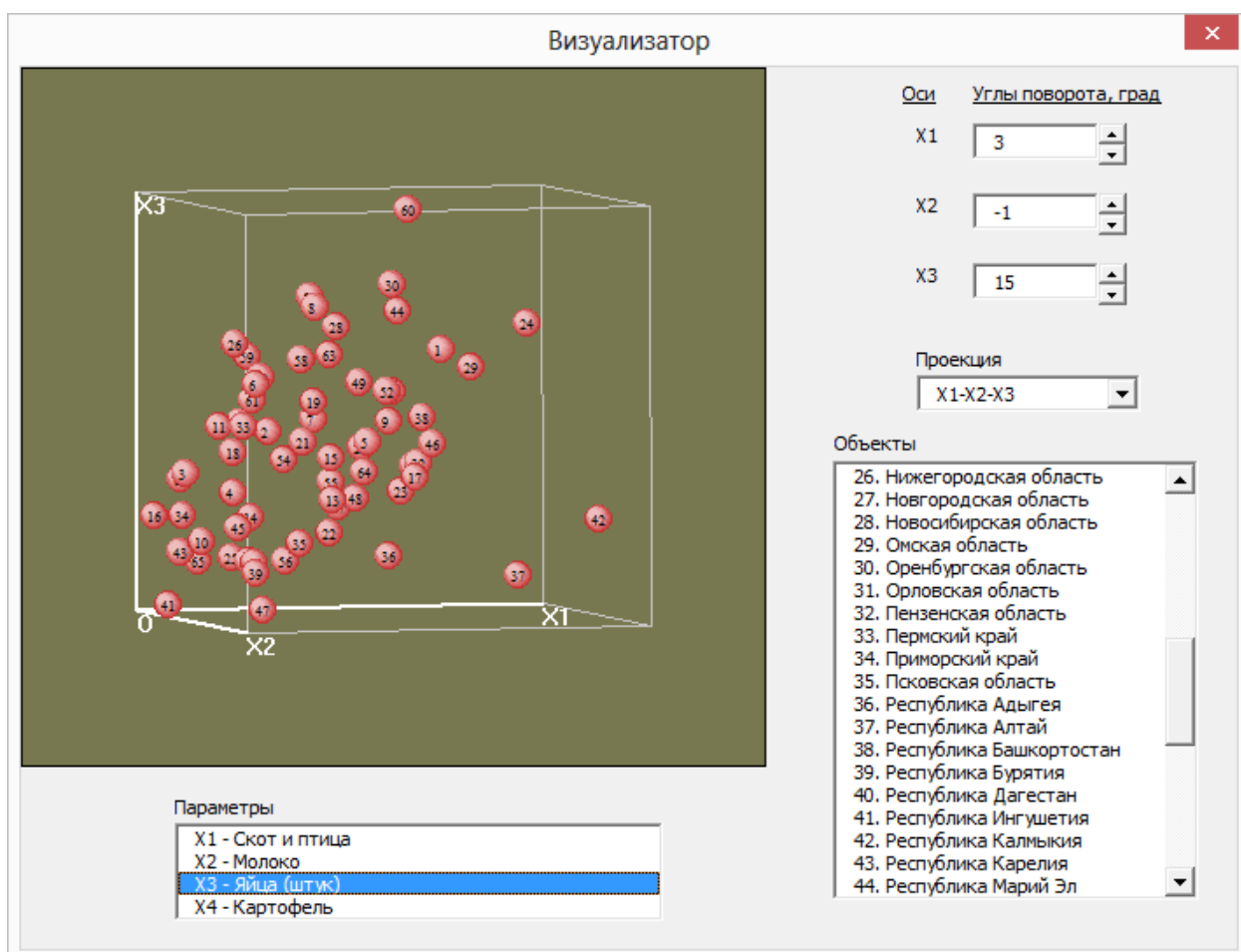


Рис.14. Визуальное представление данных об уровне выпускаемой с/х продукции различных регионов РФ после удаления «выбросов».

На этой проекции, а также на других проекциях с четвертым параметром, явных выбросов не наблюдается.

7. Согласно визуальной оценки рис.14 мы имеем следующие группы обособленных объектов: 60, 42, 37 и, возможно 24. Без этих объектов форма скопления остальных объектов близка к эллипсоидной.

8. Проверим эти предположения.

- для этого в ячейку F9 введем количество желаемых кластеров – 5;

- последовательно нажмем на кнопки «Кластеризация» и затем «Визуализация». Результат представлен на рис.15.

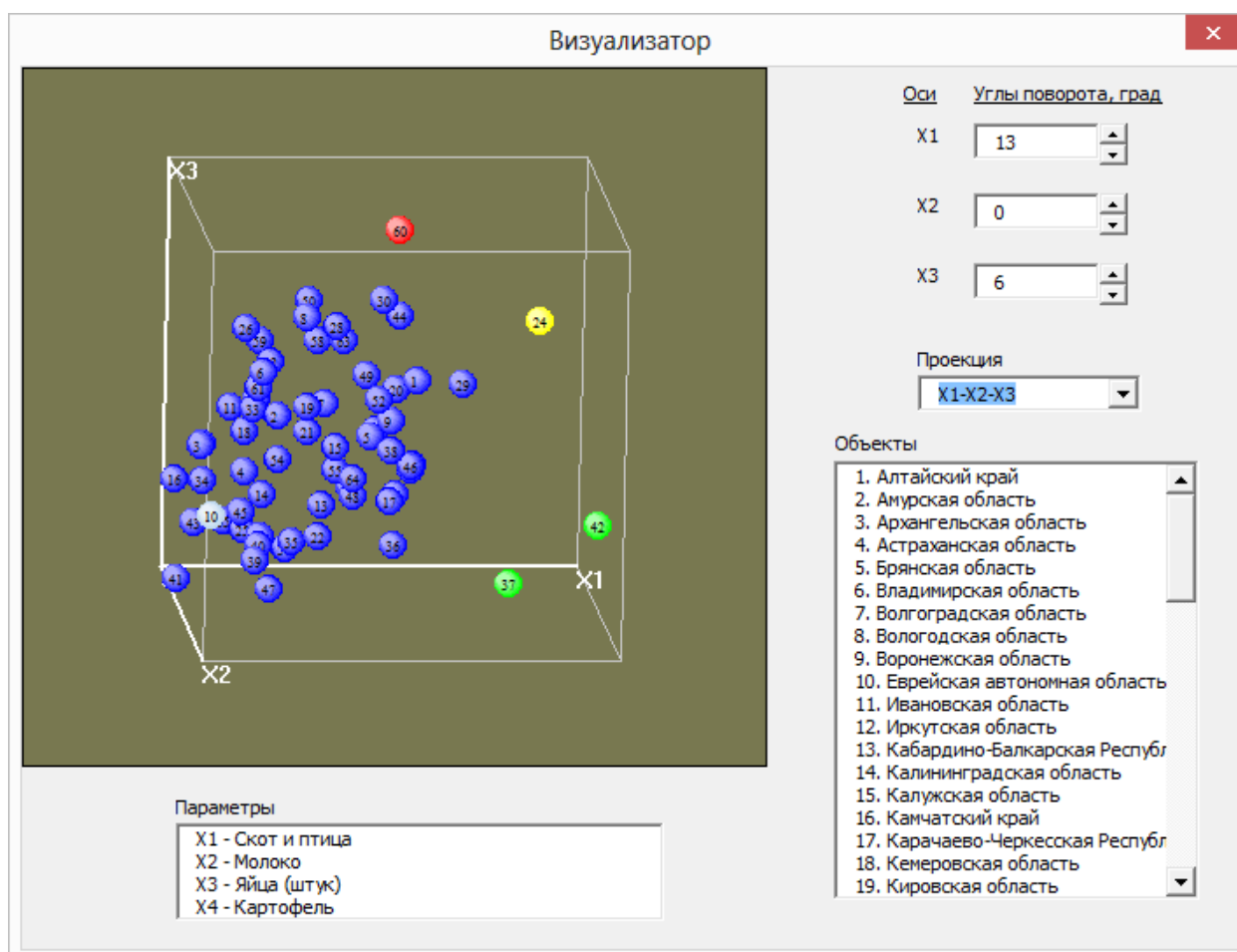


Рис.15. Результат разбиения объектов на пять кластеров.

Согласно рис.15. в отдельные кластеры выделяются:

- объект 60 (Удмуртская республика);
- объект 24 (Липецкая область);
- объекты 37 и 42 (Республика Алтай и Республика Калмыкия);
- объект 10 (Еврейская автономная область). На данной проекции выделить этот объект в отдельную группу никакими вращениями не удастся. Но если перейти на любую проекцию, связанную с четвертым параметром, то он будет явным образом отличаться от основной группы;
- все остальные объекты.

Таким образом в результате проведенного анализа выявлены следующие девять кластеров:

1. Белгородская область
2. Ленинградская область;
3. Республика Мордовия;
4. Костромская область и Ярославская область;
5. Удмуртская республика;
6. Липецкая область;
7. Республика Алтай и Республика Калмыкия;
8. Еврейская автономная область;
9. Остальные объекты.

6. Задания

По результатам выполнения работы должен быть представлен отчет, главным результа-

том которого является список кластеров с указанием объектов, входящих в каждый кластер. Одновременно должны быть описаны все шаги проведенного анализа, подтвержденные скриншотами.

Варианты заданий:

№ варианта	Местонахождение данных	Группа показателей
1	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X1-X2-X3-X4
2	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X5-X6-X7-X8
3	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X4-X5-X6-X7
4	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X8-X9-X10-X11
5	Файл "Кластерный 3D-анализ", лист "сх продукция"	X3-X4-X5-X6
6	Файл "Кластерный 3D-анализ", лист "сх продукция"	X5-X6-X7-X8
7	Файл "Кластерный 3D-анализ", лист "Выборы 2010"	X1-X2-X3-X4
8	Файл "Кластерный 3D-анализ", лист "Выборы 2010"	X5-X6-X7-X8
9	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X1-X2-X3-X4
10	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X2-X3-X4-X5
11	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X3-X4-X5-X6
12	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X1-X2-X3-X4
13	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X5-X6-X7-X8
14	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X7-X8-X9-X10

Лабораторная работа № 8 Кластеризация данных средствами языка Python

1. Краткие сведения

Одной из особенностей языка является наличие большого количества подключаемых библиотек, содержащих большой набор процедур и функций, связанных с конкретной предметной областью. Одной из таких библиотек является sklearn. В ее составе содержится набор готовых процедур, реализующих различные методы кластеризации.

Пример 1

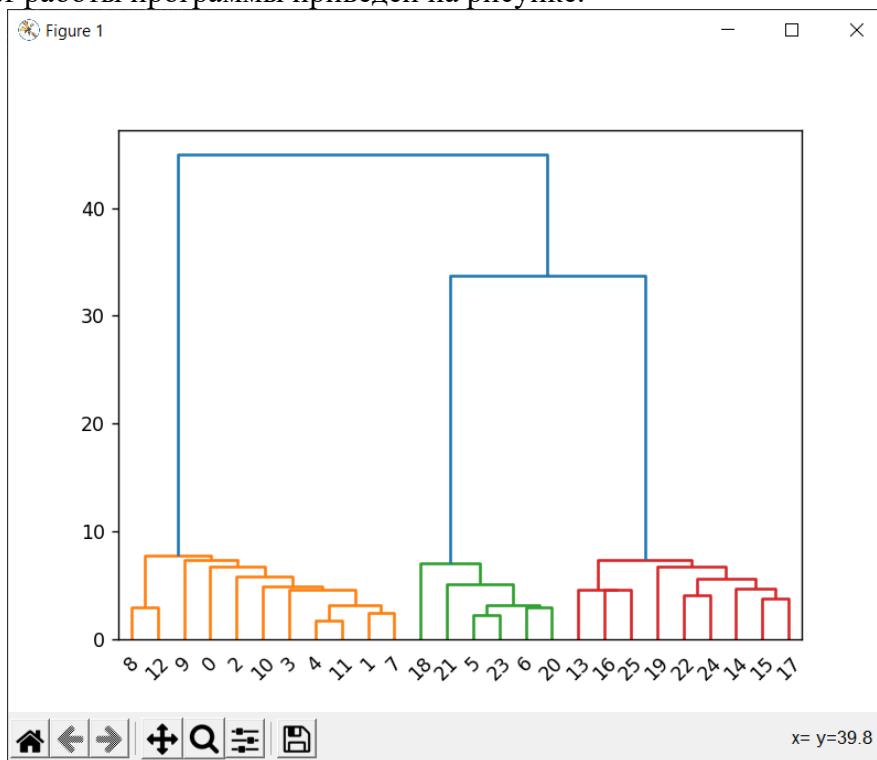
```
from scipy.cluster.hierarchy import dendrogram, linkage
from matplotlib import pyplot as plt
MyData = [[62, 110, 65], [65, 100, 64], [60, 98, 59],
[64, 105, 61], [66, 100, 68], [93, 47, 55],
[91, 50, 55], [63, 101, 63], [72, 95, 69],
[69, 111, 68], [68, 104, 66], [65, 99, 67],
[70, 94, 71], [73, 7, 67], [85, 1, 67],
```

```

[82, 3, 70], [79, 10, 72], [80, 4, 67],
[89, 42, 52], [75, 0, 69], [90, 52, 53],
[86, 55, 54], [86, 11, 65], [94, 49, 55],
[84, 8, 67], [75, 8, 71]]
Z = linkage(MyData, 'single')
dn = dendrogram(Z)
plt.show()

```

Результат работы программы приведен на рисунке.



Пример 2

```

# Метод k-средних
from sklearn.cluster import KMeans
# Задаем набор данных
MyData=[[62, 110, 65], [65, 100, 64], [60, 98, 59],
[64, 105, 61], [66, 100, 68], [93, 47, 55],
[91, 50, 55], [63, 101, 63], [72, 95, 69],
[69, 111, 68], [68, 104, 66], [65, 99, 67],
[70, 94, 71], [73, 7, 67], [85, 1, 67],
[82, 3, 70], [79, 10, 72], [80, 4, 67],
[89, 42, 52], [75, 0, 69], [90, 52, 53],
[86, 55, 54], [86, 11, 65], [94, 49, 55],
[84, 8, 67], [75, 8, 71]]
print ("количество данных=",len(MyData))
# Описываем модель
model = KMeans(n_clusters=3)
# Проводим моделирование
model.fit(MyData)
# Предсказание на всем наборе данных
all_predictions = model.predict(MyData)
# Выводим предсказания

```

```
print("распределение объектов по кластерам:")
```

```
Print(all_predictions)
```

Результат работы программы:

количество данных= 26

распределение объектов по кластерам:

```
[1 1 1 1 1 0 0 1 1 1 1 1 1 2 2 2 2 2 0 2 0 0 2 0 2 2]
```

В обоих примерах использован один и тот же набор данных, заимствованный из лабораторной работы 6 (модельные данные). обратите внимание на то, что при каждом запуске второй программы распределение объектов по кластерам меняется, т. е. при одном запуске группа объектов может оказаться в нулевом кластере, а при другом, к примеру, в первом. Но при этом группировка по группам остается правильной. Очевидно, что этот эффект обусловлен вызван применяемым алгоритмом.

2. Задания

Произвести кластеризацию данных (согласно варианту) двумя методами:

- методом иерархической кластеризации;
- методом k-средних.

Замечание

Предлагаемые для анализа данные очень объемны, а потому их ввод в программу так, как это показано в примерах является очень трудоемким. Поэтому предлагается вариант чтения данных из файла по следующему образцу:

```
file = open('Data.txt', 'r')
```

```
MyData = []
```

```
MyData = file.read()
```

```
print(MyData)
```

здесь 'Data.txt' – текстовый файл, содержащий исходные данные.

По результатам выполнения работы должен быть представлен отчет, главным результатом которого является список кластеров с указанием объектов, входящих в каждый кластер. Одновременно должны быть описаны все шаги проведенного анализа, подтвержденные скриншотами.

Варианты заданий:

№ варианта	Местонахождение данных	Группа показателей
1	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X1-X2-X3-X4
2	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X5-X6-X7-X8
3	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X4-X5-X6-X7
4	Файл "Кластерный 3D-анализ", лист "Результаты ЕГЭ"	X8-X9-X10-X11
5	Файл "Кластерный 3D-анализ", лист "сх продукция"	X3-X4-X5-X6
6	Файл "Кластерный 3D-анализ", лист "сх продукция"	X5-X6-X7-X8
7	Файл "Кластерный 3D-анализ", лист "Выборы 2010"	X1-X2-X3-X4
8	Файл "Кластерный 3D-анализ", лист "Выборы 2010"	X5-X6-X7-X8
9	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X1-X2-X3-X4
10	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X2-X3-X4-X5

11	Файл "Кластерный 3D-анализ", лист "Данные по регионам"	X3-X4-X5-X6
12	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X1-X2-X3-X4
13	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X5-X6-X7-X8
14	Файл "Кластерный 3D-анализ", лист "Инновационный потенциал "	X7-X8-X9-X10

Лабораторная работа № 9

Использование статистических методов. Выявление множеств

1. Описание задачи

Имеются данные о совместных продажах различных товаров. Необходимо выявить какие товары наиболее часто покупаются совместно.

Речь фактически идет о проведении специфического кластерного анализа, при котором товары, наиболее часто встречающиеся при совместных покупках относятся к одному множеству товаров (к одному кластеру).

2. Пример

На сетевом диске имеется файл "Множества.xls". Скопируйте его в свою рабочую папку.

В нем на листе "Продажи" приводятся сведения о примерно 6000 продаж. При этом указывается какие товары были куплены, в каждой покупке (1 – товар присутствовал в покупке, 0 – товар отсутствовал).

Расшифровка кодов товаров приведена на листе "Товары".

Для определения количества совместных покупок каждого товара с каждым создается следующая таблица (рекомендуется придерживаться указанной в таблице адресации):

	AB	AC	AD	AE	AF	AG	AH	AI	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB
4																										
5			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
6		1	2043	1262	1252	1169	1075	1011	1154	600	1238	470	478	411	322	307	475	586	201	227	296	605	432	255	237	
7		2	1262	2019	1242	1150	1036	960	1145	592	1197	465	504	420	311	255	479	606	204	223	293	573	420	239	227	
8		3	1252	1242	2016	1117	1015	964	1137	593	1218	490	507	439	310	288	483	604	206	236	279	598	450	270	255	
9		4	1169	1150	1117	1929	972	925	1081	568	1164	452	500	410	294	273	478	548	199	228	287	580	393	252	262	
10		5	1075	1036	1015	972	1835	868	1014	533	1070	433	438	375	268	258	456	558	210	218	281	530	382	247	214	
11		6	1011	960	964	925	868	1784	954	574	1024	410	441	386	265	269	402	519	198	221	270	538	388	267	242	
12		7	1154	1145	1137	1081	1014	954	1971	589	1154	458	494	397	288	272	461	584	208	237	298	616	392	272	251	
13		8	600	592	593	568	533	574	589	2281	845	894	909	850	723	695	312	394	197	216	204	948	561	1072	869	
14		9	1238	1197	1218	1164	1070	1024	1154	845	2287	579	647	580	493	466	442	594	233	253	308	742	483	421	337	
15		10	470	465	490	452	433	410	458	894	579	1592	809	613	458	456	200	272	128	146	154	643	410	849	755	
16		11	478	504	507	500	438	441	494	909	647	809	1619	643	466	450	195	283	119	116	134	630	359	826	700	
17		12	411	420	439	410	375	386	397	850	580	613	643	1493	461	464	232	296	156	146	140	701	454	712	594	
18		13	322	311	310	294	268	265	288	723	493	458	466	461	1177	422	153	224	113	123	119	618	381	535	426	
19		14	307	255	288	273	258	269	272	695	466	456	450	464	422	1153	152	216	103	117	111	596	361	540	433	
20		15	475	479	483	478	456	402	461	312	442	200	195	232	153	152	1477	716	507	557	635	390	479	187	185	
21		16	586	606	604	548	558	519	584	394	594	272	283	296	224	216	716	1683	538	578	647	458	535	211	214	
22		17	201	204	206	199	210	198	208	197	233	128	119	156	113	103	507	538	1071	510	584	265	439	157	154	
23		18	227	223	236	228	218	221	237	216	253	146	116	146	123	117	557	578	510	1145	605	294	442	164	170	
24		19	296	293	279	287	281	270	298	204	308	154	134	140	119	111	635	647	584	605	1266	307	476	151	161	
25		20	605	573	598	580	530	538	616	948	742	643	630	701	618	596	390	458	265	294	307	2065	642	835	687	
26		21	432	420	450	393	382	388	392	561	483	410	359	454	381	361	479	535	439	442	476	642	1630	526	435	
27		22	255	239	270	252	247	267	272	1072	421	849	826	712	535	540	187	211	157	164	151	835	526	1656	906	
28		23	237	227	255	262	214	242	251	869	337	755	700	594	426	433	185	214	154	170	161	687	435	906	1463	
29																										

При этом в ней должно быть 23 строки и 23 столбца.

Непосредственно для расчетов в ячейку AD6 вводится формула:

=СЧЁТЕСЛИМН(\$D5:\$D\$5814;\$AD\$5;D\$5:D\$5814;\$AD\$5), которая затем копируется во все ячейки таблицы.

Если присмотреться к полученным результатам, то можно отметить следующее:

- Элементы на главной диагонали представляют собой общее количество покупок i -ого товара. В этом нетрудно убедиться, если подсчитать суммы по каждому товару в исходной таблице;

- Данные по частоте совместных покупок симметричны относительно главной диагонали. Такая симметрия позволяет просматривать таблицу как по столбцам, так и по строкам (исходя из соображений удобства);

- Специалисты, знакомые с основами кластерного анализа, сразу увидят, что таблица очень напоминает матрицу расстояний, используемую при стандартном кластерном анализе. Там элементами таблицы является расстояния между объектами и чем меньше эти расстояния, тем с большей вероятностью объекты можно отнести к одному кластеру.

Здесь же вместо расстояний находятся частоты покупок и чем эти частоты больше, тем с большей вероятностью объекты можно отнести к одному кластеру.

Последнее обстоятельство позволяет произвести кластеризацию вручную по следующей схеме:

1. Берется первый объект (в первом столбце) и заносится в первый кластер. Чтобы затем не ошибиться ячейку AD6 можно залить каким-то цветом (например, красным).

2. Переходим ко второму столбцу (второму объекту) и ищем в нем ячейку с максимальной частотой (при этом ячейка, находящаяся на главной диагонали не учитывается). Такой ячейкой является ячейка AE5, находящаяся в первой строке таблицы. Поскольку максимальная частота появления рассматриваемого второго объекта наблюдается совместно с первым объектом, то второй объект включается в первый кластер, где уже находится первый объект.

Поэтому закрашиваем ячейку AE7 цветом первого кластера (т.е. красным).

3. Переходим к третьему столбцу. Здесь максимальная частота также находится в первой строке. Т. е. третий объект чаще всего встречается вместе с первым. Поэтому и третий объект относим к первому кластеру и закрашиваем ячейку AF8 красным цветом.

4. Переходим к четвертому столбцу. После аналогичных рассуждений четвертый объект будет отнесен к первому кластеру и ячейка AG9 будет закрашена красным цветом.

5. Аналогично к первому кластеру будет отнесен и пятый объект.

6. В шестом столбце максимальная частота наблюдается в девятой строке. Поскольку ни шестой ни девятый объект еще “свободны” и максимально близки по частоте, о естественно предположить, что они представляют собой новый отдельный кластер. Закрасим их, например, зеленым цветом.

7. Седьмой объект имеет максимальную совместную частоту с первым объектом, а потому относим его к первому кластеру.

8. Восьмой объект имеет максимальную совместную частоту вместе с одиннадцатым. Оба они пока еще “свободны”, а потому относим их к новому кластеру и соответствующие ячейки закрашиваем, например синим цветом.

9. Девятый столбец пропускаем, поскольку он уже включен во второй кластер.

10. Десятый объект имеет максимальную частоту в одиннадцатой строке. Одиннадцатый объект входит в третий кластер. Поэтому десятый объект также включаем в третий кластер.

11. И т.д.

В результате должно получиться, примерно, следующее:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	2043	1262	1252	1169	1075	1011	1154	600	1238	470	478	411	322	307	475	586	201	227	296	605	432	255	237
2	1262	2019	1242	1150	1036	960	1145	592	1197	465	504	420	311	255	479	606	204	223	293	573	420	239	227
3	1252	1242	2016	1117	1015	964	1137	593	1218	490	507	439	310	288	483	604	206	236	279	598	450	270	255
4	1169	1150	1117	1929	972	925	1081	568	1164	452	500	410	294	273	478	548	199	228	287	580	393	252	262
5	1075	1036	1015	972	1835	868	1014	533	1070	433	438	375	268	258	456	558	210	218	281	530	382	247	214
6	1011	960	964	925	868	1784	954	574	1024	410	441	386	265	269	402	519	198	221	270	538	388	267	242
7	1154	1145	1137	1081	1014	954	1971	589	1154	458	494	397	288	272	461	584	208	237	298	616	392	272	251
8	600	592	593	568	533	574	589	2281	845	894	909	850	723	695	312	394	197	216	204	948	561	1072	869
9	1238	1197	1218	1164	1070	1024	1154	845	2287	579	647	580	493	466	442	594	233	253	308	742	483	421	337
10	470	465	490	452	433	410	458	894	579	1592	809	613	458	456	200	272	128	146	154	643	410	849	755
11	478	504	507	500	438	441	494	909	647	809	1619	643	466	450	195	283	119	116	134	630	359	826	700
12	411	420	439	410	375	386	397	850	580	613	643	1493	461	464	232	296	156	146	140	701	454	712	594
13	322	311	310	294	268	265	288	723	493	458	466	461	1177	422	153	224	113	123	119	618	381	535	426
14	307	255	288	273	258	269	272	695	466	456	450	464	422	1153	152	216	103	117	111	596	361	540	433
15	475	479	483	478	456	402	461	312	442	200	195	232	153	152	1477	716	507	557	635	390	479	187	185
16	586	606	604	548	558	519	584	394	594	272	283	296	224	216	716	1683	538	578	647	458	535	211	214
17	201	204	206	199	210	198	208	197	233	128	119	156	113	103	507	538	1071	510	584	265	439	157	154
18	227	223	236	228	218	221	237	216	253	146	116	146	123	117	557	578	510	1145	605	294	442	164	170
19	296	293	279	287	281	270	298	204	308	154	134	140	119	111	635	647	584	605	1266	307	476	151	161
20	605	573	598	580	530	538	616	948	742	643	630	701	618	596	390	458	265	294	307	2065	642	835	687
21	432	420	450	393	382	388	392	561	483	410	359	454	381	361	479	535	439	442	476	642	1630	526	435
22	255	239	270	252	247	267	272	1072	421	849	826	712	535	540	187	211	157	164	151	835	526	1656	906
23	237	227	255	262	214	242	251	869	337	755	700	594	426	433	185	214	154	170	161	687	435	906	1463

Если перейти от кодов товаров к их названиям, то результат анализа будет выглядеть следующим образом:

Кластер 1	Кластер 2	Кластер 3	Кластер 4	Кластер 5
Молоко	Хлеб белый	Алкоголь	Печенье	Зефир
Сметана	Йогурт	Хлеб черный	Конфеты	Шоколад
Творог		Мясо		Мармелад
Кефир		Колбаса		
Сыр		Рыба		
Ряженка		Курица		
		Индейка		
		Овощи		
		Фрукты		
		Табак		

При анализе реальных данных количество товаров может достигать нескольких сотен. Поэтому вместо ручных расчетов лучше использовать соответствующие программные средства.

В данном случае в редакторе Visual Basic файла «Множества.xls» приведен макрос, который воспроизводит вышеприведенный алгоритм выделения кластеров.

Перейдите в редактор Visual Basic и запустите макрос Raschet (при активном листе «Продажи»). Результат будет напечатан в столбцах ВD:ВН.

3. Варианты заданий

№ варианта	Использовать исходные данные продаж, удалив столбец
1	1

2	3
3	5
4	6
5	7
6	9
7	11
8	13
9	15
10	17
11	19
12	21
13	23

Лабораторная работа № 10

Использование статистических методов. Поиск ассоциаций

1. Описание задачи

Во многих интернет магазинах требуется регистрация пользователей. Помимо всего прочего, данные регистрации могут быть использованы в следующих целях.

При входе на сайт магазина пользователю предоставляется возможность просмотреть каталоги товаров и цены на них. При этом количество товаров может быть настолько велико, что некоторым пользователям может просто не хватить терпения листать список в поисках нужного ему товара – и он просто уйдет с сайта, даже если нужный ему товар здесь есть и он самый дешевый и качественный.

Чтобы этого не произошло осуществляется сужение области поиска, используя параметры пользователя.

Например.

- пользователь: девушка, 18 лет, студентка, не замужняя – с высокой вероятностью закажет косметику и/или кондитерские изделия;

- пользователь мужчина, 30 лет, офисный работник, не женатый - с высокой вероятностью закажет алкогольные и/или табачные изделия и т.д.

Т.е. в сокращенный список товаров должны быть включены те товары, которые с наибольшей вероятностью купит пользователь, с характерными для него параметрами.

Отметим, что фирма MicroSoft решает подобные задачи с использованием нейросетевых технологий.

2. Пример

На сетевом диске имеется файл "Интернет магазин.xls". В нем приводятся сведения о покупаемой группе товаров пользователей с различным профилем свойств.

При этом рассматриваются только данные о пяти группах товаров:

1	Детские товары
2	Одежда
3	Кондитерские изделия
4	Косметика
5	Алкоголь и табачные изделия

Набор параметров пользователей ограничен тремя показателями:

	Пол
1	м
2	ж
	Возраст
1	до 21
2	21-40
3	более 40
	Профессия
1	Неработающие
2	Офисный
3	Рабочие

Таким образом у нас имеется три входных параметра (пол, возраст и профессия), которые составят профиль покупателя. Количество возможных комбинаций значений входных параметров будет равно $2 * 3 * 3 = 18$.

Представим эти комбинации в закодированном виде следующим образом (рекомендуется размещение данных по указанным в таблице адресам):

Таблица 1

	F	G	H	I	J	K	L	M	N	O
3					Группа товаров					
4		Пол	Возраст	Статус	1	2	3	4	5	
5		1	1	1						
6		1	1	2						
7		1	1	3						
8		1	2	1						
9		1	2	2						
10		1	2	3						
11		1	3	1						
12		1	3	2						
13		1	3	3						
14		2	1	1						
15		2	1	2						
16		2	1	3						
17		2	2	1						
18		2	2	2						
19		2	2	3						
20		2	3	1						
21		2	3	2						
22		2	3	3						
23										
24										

Подсчитаем сколько товаров разных групп купил каждый тип пользователя. Для этого в ячейку J5 вводится формула:

=СЧЁТЕСЛИМН(\$B\$4:\$B\$2745;\$G5;\$C\$4:\$C\$2745;\$H5;\$D\$4:\$D\$2745;\$I5;\$E\$4:\$E\$2745;\$J\$4) которая копируется до ячейки N22 (с форматом функции предлагается разобратся самостоятельно).

Результаты вычислений показаны в таб. 2

Таблица 2

	F	G	H	I	J	K	L	M	N	O	P	Q
2												
3					Группа товаров							
4		Пол	Возраст	Статус	1	2	3	4	5			
5		1	1	1	11	41	12	3	78	145	5	2
6		1	1	2	1	34	44	27	62	168	5	3
7		1	1	3	1	34	13	1	110	159	5	2
8		1	2	1	8	25	29	10	83	155	5	3
9		1	2	2	22	54	31	11	39	157	2	5
10		1	2	3	18	53	21	19	37	148	2	5
11		1	3	1	12	20	33	31	48	144	5	3
12		1	3	2	24	69	21	18	20	152	2	1
13		1	3	3	6	36	8	27	45	122	5	2
14		2	1	1	2	34	57	61	3	157	4	2
15		2	1	2	9	42	42	69	5	167	4	3
16		2	1	3	16	33	45	54	17	165	4	3
17		2	2	1	8	43	27	50	30	158	4	2
18		2	2	2	46	43	19	36	9	153	1	2
19		2	2	3	65	33	7	23	15	143	1	2
20		2	3	1	30	78	12	10	25	155	2	1
21		2	3	2	37	34	19	35	13	138	1	4
22		2	3	3	36	49	18	30	23	156	2	1
23					352	755	458	515	662	2742		
24										2742		

Здесь в ячейках J5:N22 вычислено количество покупок соответствующей группы товаров для каждого типа покупателя.

В ячейках O5:O22 подсчитано общее количество покупок для каждого типа покупателя.

В ячейках J23:N23 подсчитано общее количество покупок для каждого типа товара.

В ячейках O23 и O24 подсчитано общее количество покупок для строки 23 и столбца O.

В столбцах P и Q показаны номера наиболее вероятных групп товаров, которые покупаются каждым типом покупателя (в порядке убывания).

Таким образом, для придания сайту интернет-магазина описанной функции достаточно организовать вычисления с итоговой таблицей, аналогичной табл. 2.

3. Варианты заданий

Получить частоты покупок для следующих профилей:

№ варианта	Профиль
1	Пол – Возраст
2	Пол – Статус
3	Возраст – Статус

4	Пол
5	Возраст
6	Статус
7	Пол – Возраст
8	Пол – Статус
9	Возраст – Статус
10	Пол
11	Возраст
12	Статус

4. Регрессионно-дисперсионный анализ

Основные понятия теории эксперимента

Фактор – входной параметр, каким-то образом влияющий на объект исследования. Стандартное обозначение – X .

Уровень фактора – конкретное значение входного параметра. Отметим, что для выяснения влияния фактора на объект необходимо как минимум два уровня. Стандартное обозначение – X_i .

Многофакторный эксперимент – эксперимент, в котором исследуется влияние на выходной параметр сразу нескольких факторов (минимум двух).

Выходной параметр (отклик, функция отклика, параметр оптимизации) – свойство исследуемого объекта, на который предположительно оказывает влияние фактор (факторы). Стандартное обозначение – Y .

Взаимодействие факторов – взаимное влияние факторов на выходной параметр, выражающееся в том, что зависимость выходного параметра от одного фактора зависит от значения уровня другого фактора.

Взаимодействие может быть положительным, отрицательным или полностью отсутствовать.

Положительное взаимодействие выражается в том, что увеличение значения уровня одного фактора при одновременном росте значений второго фактора приводит к непропорциональному росту значений выходного параметра.

Отрицательное взаимодействие выражается в том, что с увеличением значения уровня одного фактора при одновременном росте значений второго фактора происходит непропорционально высокое падение значений выходного параметра.

Ортогональность плана эксперимента – признак ортогональности плана рассмотрим на примере.

Предположим, что имеется следующий план фвухфакторного эксперимента:

X1 (температура, °C)	X2 (давление, атм)	Y
20	2	Y_1
40	4	Y_2
60	6	Y_3

Очевидно, что приведенный план является некорректным, поскольку в результате эксперимента не будет понятно, какой фактор на самом деле влияет на выходной параметр.

И в то же время план:

X1 (температура, °C)	X2 (давление, атм)	Y
20	2	Y_1
20	4	Y_2
20	6	Y_3
40	2	Y_4

40	4	Y ₅
40	6	Y ₆
60	2	Y ₇
60	4	Y ₈
60	6	Y ₉

является корректным.

В терминах теории эксперимента такой план является полным факторным экспериментом (ПФЭ), поскольку реализованы все комбинации значений уровней факторов. При этом имеется возможность оценить влияние на выходной параметр каждого отдельного фактора и их взаимодействий.

В терминах планирования эксперимента полученная ранее таблица 2 является результатом полного факторного эксперимента, в котором имеется три входных параметра (пол, возраст, профессия) и пять выходных параметров (объем продаж пяти групп товаров). При этом факторы являются качественными, а параметры оптимизации – количественными.

Для обработки результатов подобных экспериментов применяется дисперсионный анализ. Однако в этом случае возможно применение специфической разновидности регрессионного анализа.

Суть стандартного регрессионного анализа заключается в получении параметров уравнения регрессии, которое является математической моделью какого-то процесса. В простейшем случае такая модель представляет собой линейную зависимость следующего вида:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (1).$$

В случае экспериментов с количественными факторами также можно использовать подобную модель. Но качественный характер факторов накладывает свою специфику на вид модели, метод расчета коэффициентов и на правила ее использования. Формально вид модели для объектов с качественными факторами аналогичен уравнению (1):

$$y = b_0 + \sum_{i=1}^n b_i F_i \quad (2)$$

где y – выходной параметр;
 n – количество факторов;
 b_0 – общее среднее эксперимента;
 F_i – эффект i -ого фактора.

Для рассматриваемого примера уравнение (2) принимает вид:

$$y = b_0 + b_1P + b_2V + b_3S \quad (3)$$

где P – величина эффекта фактора пола покупателя;
 V – величина эффекта фактора возраста покупателя;
 S – величина эффекта фактора профессии покупателя.

Но поскольку все три фактора являются качественными, то уравнение (3) должно быть записано в виде:

$$y_{ijk} = b_0 + b_{11}P_1 + b_{12}P_2 + b_{21}V_1 + b_{22}V_2 + b_{23}V_3 + b_{31}S_1 + b_{32}S_2 + b_{33}S_3 \quad (4)$$

где y_{ijk} – значение выходного параметра при i -ом уровне первого фактора, j -ом уровне второго фактора и k -ом уровне третьего фактора;
 P, V, S – означают присутствие соответствующего уровня данного фактора в уравнении. Эти величины равны единице для i -ого уровня первого фактора, j -ого уровня второго фактора и k -ого уровня третьего фактора. Остальные их значения равны нулю.

Например, для вычисления значения выходного параметра на втором уровне первого фактора, первом уровне второго фактора и третьем уровне третьего фактора уравнение сводится к виду:

$$y_{213} = b_0 + b_{12}P_2 + b_{21}V_1 + b_{33}S_3 \quad (4a)$$

Кажущаяся сложность в использовании уравнения компенсируется простотой вычисления его параметров.

Уравнение (4) можно переписать в следующем виде:

$$y = b_0 + \sum_{i=1}^m P_i + \sum_{j=1}^n V_j + \sum_{k=1}^p S_k, \quad (5)$$

где m, n, k - количество уровней соответствующего фактора.

Для него составляется следующая система уравнений:

$$\left\{ \begin{array}{l} \sum_{i=1}^q y_i = qb_0 \\ \sum_{i=1}^m y_i = mb_0 + mb_{11} \\ \sum_{i=1}^m y_i = mb_0 + mb_{12} \\ \sum_{j=1}^n y_j = nb_0 + nb_{21} \\ \sum_{j=1}^n y_j = nb_0 + nb_{22} \\ \sum_{j=1}^n y_j = nb_0 + nb_{23} \\ \sum_{k=1}^p y_k = pb_0 + pb_{31} \\ \sum_{k=1}^p y_k = pb_0 + pb_{32} \\ \sum_{k=1}^p y_k = pb_0 + pb_{33} \end{array} \right. \quad (6)$$

где q - общее количество данных;

m - количество уровней первого фактора;

n - количество уровней второго фактора;

p - количество уровней третьего фактора.

Полученная система является очень разреженной, благодаря свойству ортогональности плана эксперимента. Это свойство приводит к тому, что все суммы в уравнении (5) равны 0.

Рассчитаем параметры уравнения для первой группы товаров. Для удобства воспроизведем соответствующий фрагмент табл. 2.

Таблица 3

Пол	Возраст	Статус	Группа 1	Расчетные значения
1	1	1	11	-9,168
1	1	2	1	2,166
1	1	3	1	2,666

1	2	1	8	11,999
1	2	2	22	23,333
1	2	3	18	23,833
1	3	1	12	10,333
1	3	2	24	21,667
1	3	3	6	22,167
2	1	1	2	7,054
2	1	2	9	18,388
2	1	3	16	18,888
2	2	1	8	28,221
2	2	2	46	39,555
2	2	3	65	40,055
2	3	1	30	26,555
2	3	2	37	37,889
2	3	3	36	38,389
			352	

$$b_0 = 352 / 18 = 19,556;$$

$$b_{11} = (11+1+1+8+22+18+12+24+6) / 9 - 19,556 = -8,111$$

$$b_{12} = (2+9+16+8+46+65+30+37+36) / 9 - 19,556 = 8,111$$

$$b_{21} = (11+1+1+2+9+16) / 6 - 19,556 = -12,89$$

$$b_{22} = (8+22+18+8+46+65) / 6 - 19,556 = 9,556$$

$$b_{23} = (12+24+6+30+37+36) / 6 - 19,556 = 4,611$$

$$b_{31} = (11+8+12+2+8+30) / 6 - 19,556 = -7,723$$

$$b_{32} = (1+22+24+9+46+37) / 6 - 19,556 = 3,611$$

$$b_{33} = (1+18+6+16+65+36) / 6 - 19,556 = 4,111$$

Проверка адекватности уравнения

Проверка адекватности уравнения производится по схеме, применяемой для стандартного регрессионного анализа.

Для этого:

1. С помощью полученных коэффициентов рассчитываются теоретические значения частот покупки.

Например:

$$Y_{111} = 19,556 + (-8,111) + (-12,89) + (-7,723) = -9,167$$

$$Y_{112} = 19,556 + (-8,111) + (-12,89) + (9,556) = 2,116$$

и т.д.

Результаты расчетов приведены в последней колонке табл. 3.

2. С помощью функции =ДИСПА(J5:J22)*17 рассчитывается общая дисперсия исходных данных (равна 5178,444).

3. С помощью функции =СУММКВРАЗН(J5:J22;O5:O22) рассчитывается остаточная сумма квадратов данных, полученная в результате применения уравнения регрессии (равна 1945,333).

4. Рассчитывается сумма квадратов, приходящаяся на уравнение регрессии. Равна = 5178,444 - 1945,333 = 3233,111.

5. Рассчитываются дисперсии ошибки и уравнения.

Для ошибки = 1945,333 / 12 = 162,111 (12 - число степеней свободы для ошибки).

Для уравнения = 3233,111 / 5 = 646,622 (5 - число степеней свободы для уравнения).

Равно $f = k + 1$, где k - число используемых при расчетах количество коэффициентов уравнения).

6. Рассчитывается критерий Фишера = $646,622 / 162,111 = 3,99$

7. Рассчитанный критерий Фишера сравнивается с критическим значением (определяется с помощью функции =F.ОБР.ПХ(0,05;5;12) = 3,11).

8. Если рассчитанное значение критерия больше критического, то уравнение признается адекватным. В противном случае уравнение не адекватно. В данном случае уравнение адекватно.

5. Варианты заданий

№ варианта	Получить уравнение для	Профиль
1	2 группы товаров	Пол – Возраст
2	3 группы товаров	Пол – Статус
3	4 группы товаров	Возраст – Статус
4	5 группы товаров	Пол
5	2 группы товаров	Возраст
6	3 группы товаров	Статус
7	4 группы товаров	Пол – Возраст
8	5 группы товаров	Пол – Статус
9	2 группы товаров	Возраст – Статус
10	3 группы товаров	Пол
11	4 группы товаров	Возраст
12	5 группы товаров	Статус

Лабораторная работа № 11

Методы анализа текстовой информации. Анализ отзывов

1. Постановка задачи

Развитие сетевых технологий и, не в последнюю очередь появление социальных сетей, дало возможность пользователям явно выражать свое мнение относительно качества услуг, товаров, фильмов, книг, результатах деятельности компаний или политических деятелей.

Наличие подобной информации привело к необходимости ее анализа с точки зрения установления обратной связи между поставщиками услуг и их потребителями. При этом, количество отзывов может достигать десятков тысяч, и обработка отзывов вручную оказывается невозможной. В связи с этим широкое распространение получили автоматизированные подходы к анализу тональности текстов (sentiment analysis).

Однако, непосредственно перед анализом необходимо проводить процедуру отсеивания имеющихся отзывов с точки зрения выявления их дубликатов.

Особенно этим «грешат» форумы на сайтах компаний, являющихся поставщиками таких услуг. Не является секретом то, что часть отзывов (особенно положительных) пишут менеджеры данных компаний. Но, еще чаще для их написания привлекаются специальные люди – копирайтеры. В интернет имеется множество публикаций на тему как правильно писать подобные отзывы.

Задачей настоящей работы является освоение технологии отсеивания «заказных» отзывов.

2. Описание алгоритма

Для решения выше указанной задачи мало подходят стандартные методы классификации текстов (типа *bag of words*) прежде всего из-за небольших объемов сравнивае-

мых текстов – максимум 1-2 стр. формата А4. Получающиеся при этом частотные словари являются статистически неустойчивыми и, как следствие, результаты классификации являются не надежными.

Понятие семантического спектра

Для характеристики смысла текста вводится понятие семантического спектра. Данный спектр представляет собой своеобразную хэш-функцию заданной последовательности символов.

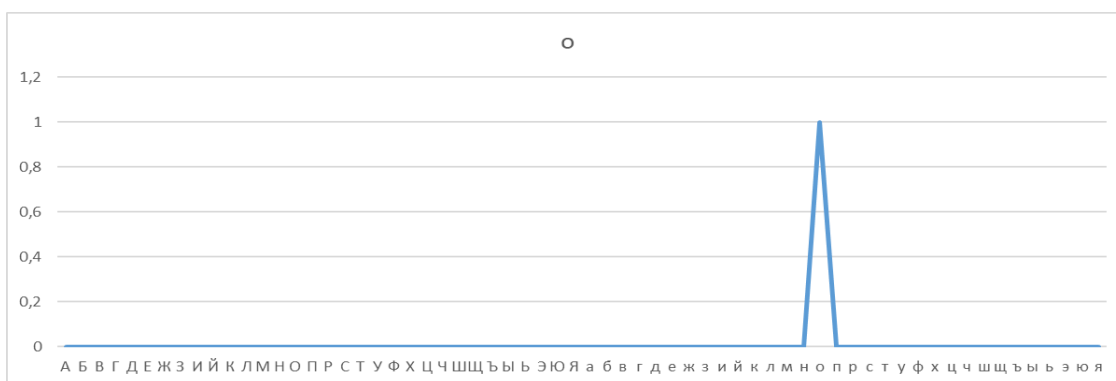
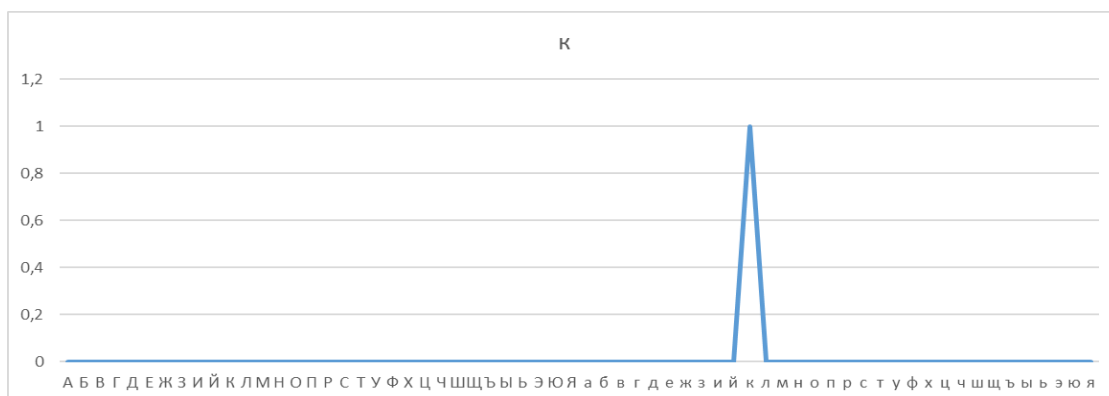
Указанная функция формируется следующим образом:

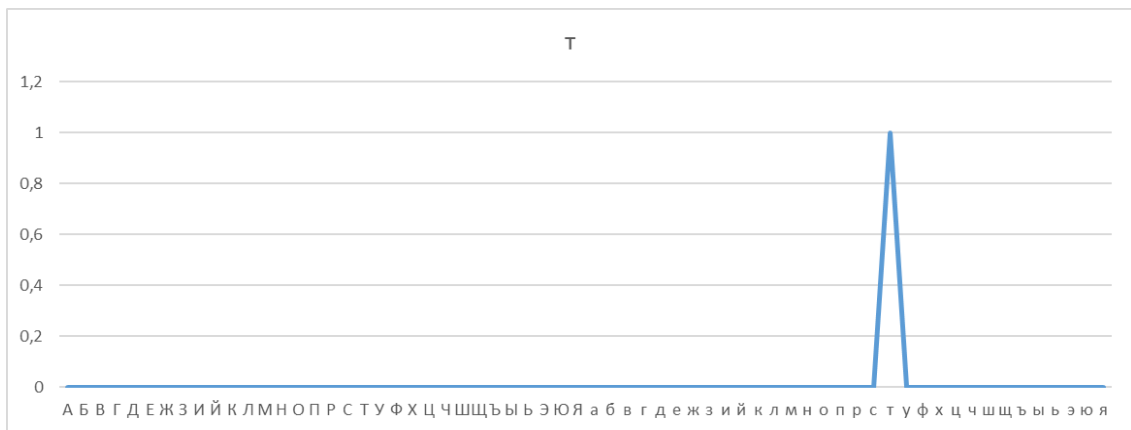
- задается набор символов, которые ожидаются в обрабатываемом тексте. Если это русскоязычные тексты, то в этот набор можно включить все строчные и прописные русские символы, цифровые символы и символы-разделители (точка, запятая, двоеточие и т.д.);

- выбранный набор символов упорядочивается. Упорядочивание наиболее естественно произвести в порядке возрастания ASCII-кодов символов;

- для каждого символа формируется стартовое значение семантического спектра. Формирование спектра производится по правилу: если номер элемента спектра равен номеру данного символа, то этот элемент спектра приравнивается 1. В противном случае элемент спектра равен 0.

Например. На рис. 5-7 показаны спектры букв «к», «о» и «т».



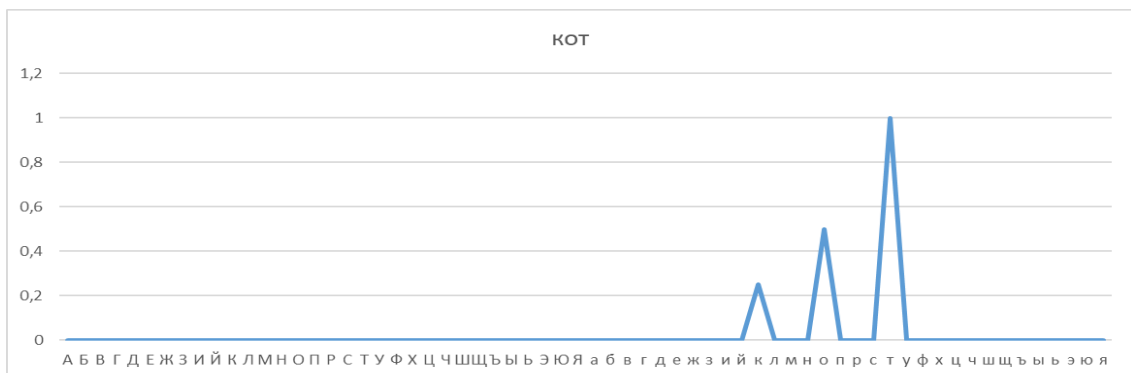


Для получения спектра последовательности символов (текста) выполняются следующие операции:

- задается начальный спектр текста, все элементы которого равны 0;
- поочередно берутся символы текста и их спектры поэлементно совмещаются с соответствующими элементами спектра текста по формуле:

$$S_{\text{текста}} = (S_{\text{текста}} + S_{\text{символа}})/2 \quad (1)$$

На рис. 8 показан полученный спектр для последовательности символов «кот».



Особенности формулы (1) приводят к следующим эффектам:

1. Получаемый спектр позволяет учитывать не только вхождение символа в последовательность, но и его порядок в этой последовательности.
2. Чем раньше символ появляется в последовательности, тем меньше его значение в спектре. Это приводит к тому, что для очень длинных последовательностей символов «сигнал» от начальных символов станет практически незаметным.

Наличие второго эффекта, в частности, означает, что нельзя механически применять уравнение (3) для очень длинных последовательностей символов.

Чтобы не потерять информацию о начальных символах длина обрабатываемой последовательности ограничивается длиной очередного слова в тексте. Типичная длина русских слов (5-6 символов) позволяет сохранять информацию о начальных символах слова в получающемся спектре. Получившийся спектр слова можно затем по той же формуле (3) объединять с начальным (нулевым) семантическим спектром текста.

Однако, нельзя и весь текст сводить к набору слов, поскольку для большого текста вклад спектров начальных слов в общую семантику текста будет постепенно нивелироваться. Чтобы этого не произошло, производится формирование промежуточных семантических спектров предложений. Для русского языка средняя длина простых предложений составляет 5-6 слов. Такой длины еще вполне достаточно, чтобы не потерять информацию о семантике начальных слов.

Если продолжить аналогичные рассуждения дольше, то просматривается следую-

шая иерархическая последовательность получения семантики текста с помощью формулы (3):

стартовая семантика символов > семантика слов > семантика простых предложений > семантика составных предложений > семантика абзацев > семантика параграфов > семантика глав (разделов) > семантика текста.

Алгоритм обработки

1. Обрабатываемый текст просматривается и все входящие в него слова заменяются базовыми словоформами.

Например:

Исходная фраза: «часто проводятся акции на разные товары»
преобразуется к виду: «часто проводить акция на разный товар»

2. Базовые словоформы заменяются условным базовым синонимом.

Например:

вышеприведенная фраза: «часто проводить акция на разный товар»
преобразуется к виду: «частый проводить поступок на отличный товар»

3. Формируется семантический спектр текста

4. С помощью шагов 1 – 3 формируются спектры других сравниваемых текстов.

5. Полученные спектры попарно сравниваются по некоторому критерию. В данной работе в качестве меры сходства семантики текстов используется косинусное сходство:

$$R = \text{ArcCos}(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

где n – размер спектра;

x_i - i -ый элемент спектра в первом тексте;

y_i - i -ый элемент спектра во втором тексте.

Значение R интерпретируется как некоторый угол между векторами X и Y в многомерном пространстве. Если вектора совпадают, угол между ними равен 0 и значение R будет равно 1. Если вектора не совпадают, то максимальный угол между ними будет равен 90 градусов и значение R будет равно 0.

3. Пример

Задание

Требования к используемому ПО:

- версия ОС: Windows 10;
- MS Office 2016
- файл «Анализ отзывов.xlsm».
- файлы «Отзыв1.txt», «Отзыв2.txt», «Отзыв3.txt», «Отзыв4.txt», «Отзыв5.txt», «Отзыв6.txt».

Подготовка файлов

Указанные выше файлы скачаны с сайта торговой сети «Лента» в формате *.txt. Однако даже такой простой формат может быть представлен в различной кодировке.

Поэтому:

1. Необходимо перевести их в кодировку Windows (или убедиться в том, что файлы уже находятся в этой кодировке).

Для этого выбранный файл загружается в Word и пересохраняется в указанной кодировке.

2. Заменить в файлах букву «ё» на букву «е».

Для этого загрузить файл в «Блокнот» выделить весь текст (Ctrl/A) и с помощью пункта меню «Заменить» произвести указанную замену.

Проведение анализа

1. Открыть файл «Анализ отзывов.xlsx».
2. Нажать кнопку «Обработать текст»
3. В появившейся форме задать имя обрабатываемого файла.

4. Нажать кнопку «Начать».

В результате в строке 5 в ячейках F18:IZ18 сформирован семантический спектр обрабатываемого текста.

5. Скопировать содержимое данных ячеек в ячейки F19:IZ19.
6. Повторить шаги 2 – 4 для другого текста и скопировать полученный спектр в строку 20.
7. Аналогично получить спектры остальных файлов, каждый раз копируя получаемый спектры в новые строки (21, 22, и т.д.).
8. Произвести расчет меры сходства полученных спектров. Для расчета используется функция CompareText (находится в категории «Определенные пользователем»).

Для более компактного размещения результатов расчетов рекомендуется следующее размещение данных:

	text1	text2	text3	text4	text5	text6
text1	1					
text2	0,444665	1				
text3	0,602006	0,7841	1			
text4	0,484063	0,6223	0,642163	1		
text5	0,288512	0,385	0,36231	0,51119	1	
text6	0,631552	0,7467	0,737299	0,67563	0,47	1

Непосредственно для расчетов необходимо установить курсор в выбранную ячейку и в строке формул вызвать функцию CompareText.

Пример вызова функции для расчета, сходства между первым и вторым текстами:
 =CompareText(F\$19:IZ\$19;F20:IZ20)

9. Строго обоснованных критических значений мер сходства не существует – все оставляется на усмотрение аналитика. В данном случае принято, что если мера сходства более 0,6, то тексты идентичны по смыслу.

10. Из данных таблицы следует, что тексты 1, 3, 4 и 6 одинаковы по смыслу и являются дубликатами. Оригинальными являются тексты 2 и 5.

4. Задания

1. Скачать из интернета отзывы о товарах или услугах выбранных компаний.
 2. Произвести процедуру сравнения полученных отзывов.
- Лишний раз напоминаем:
- скачанные тексты сохранить в формате *.txt в кодировке Windows;
 - заменить в текстах букву «ё» на букву «е».

Лабораторная работа № 12 Методы анализа текстовой информации. Классификация текстов

1. Общие сведения

Задачи обработки текстовой информации на естественном языке считаются областью такой науки как компьютерная лингвистика. Ее практическим результатом являются приложения для обработки текстов. В зависимости от целей обработки в основу приложения закладываются соответствующие модели текстов. Созданная модель и приложение на ее основе называется лингвистическим процессором (ЛП).

К настоящему времени сложились следующие основные направления:

Реферирование текста

Результатом работы системы является сокращение объема исходного текста и генерация его краткого изложения – реферата. Общий реферат может быть составлен сразу для нескольких близких по тематике документов.

Аннотирование текстов

В простейшем случае аннотация представляет собой перечень основных ключевых слов, соответствующих тематике текста.

Классификация и кластеризации текстов

При наличии создании больших коллекций электронных документов актуальны задачи с целью создания классов близких по теме документов. Классификация означает отнесение каждого документа к определенному классу с заранее известными параметрами, а кластеризация – разбиение множества документов на кластеры, т.е. подмножества тематически близких документов.

Системы машинного перевода

Системы Text Mining – это извлечение информации из текстов, или ***Information Extraction***, что необходимо при решении аналитических задач из различных предметных областей (социологии, экономики, психологии и т.д.). К направлению ***Text Mining*** относят и другие задачи. Например – выделение мнений (***Opinion Mining***) и оценка тональности текстов (***Sentiment Analysis***). В первом случае происходит поиск (в блогах, на форумах, интернет-магазинах и пр.) мнений пользователей о товарах и других объектах, а также производится анализ этих мнений. Во втором случае оценивается общая тональность и эмоциональность высказываний.

Системы поддержки диалога с пользователем

Наиболее часто эта задача решалась для интеллектуальной работы с базами данных (БД). Стандартным методом извлечения информации является применение так называемых SQL-запросов. Данный метод требует от пользователей знания этого языка.

Командные системы

Данные системы применяются для реализации голосового управления техническими устройствами (например, для управления компонентами «умного дома»).

Кроме того, стремительная информатизация всех сторон общества выдвигает все новые задачи. Например, принятая в РФ программа «Электронное правительство» подразумевает в первую очередь широкое внедрение электронного документооборота. Но ничто не мешает дополнить данный проект системами типа электронный консультант. При за-

ходе на сайт пользователь с помощью такого консультанта может получить необходимую информацию либо в виде конкретных документов, либо в виде ссылок на определенные места в этих документах.

Такие консультанты могут заметно снизить нагрузку как на сайты соответствующих госструктур (ПФР, ФНС, МФЦ и т.п.), так и на человека-консультанта. Техническая и алгоритмическая сложность создания такой системы того же порядка, что используемая в Бразилии программа «Электронный судья», способная выносить судебные решения.

2. Классификация текстов методом «Bag of Words»

Общие сведения

Модель Bag-of-words («набор слов» или «мешок слов») – самая популярная и простая модель представления текста, применяемая во многих задачах Text Mining. Модель представляет текст как набор слов без учета их взаимного расположения и взаимных связей. При ее использовании семантическая близость двух текстов (двух наборов слов) оценивается по количеству совпадающих слов. Это означает, что два текста, в которых мало общих слов или их вообще нет, считаются семантически и тематически неблизкими.

Упрощенный алгоритм построения модели выглядит следующим образом:

1. Составляется словарь из всех слов, встречающихся в сравниваемых текстах. При этом для повышения точности классификации из словаря исключаются все знаки препинания, числа и нейтральные слова (так называемые «стоп-слова»). Стоп-слова – это слова, не несущие смысловой нагрузки (союзы, предлоги и т.д.). Список нейтральных слов зависит от целей обработки.

2. Производится лексематизация словаря, т.е. различные словоформы заменяются на исходную словоформу. Например, словоформы «столы», «столе», «столов» и т. д. заменяются словоформой «стол». Кроме того, все прописные буквы заменяются строчными.

3. Для каждого документа формируется вектор, размерность которого определяется количеством слов в словаре, а значение определяется числом, сколько раз каждое слово встретилось в тексте.

4. Полученные вектора сравниваются по некоторому критерию. Наиболее часто применяется так называемое косинусное сходство:

$$R = \text{ArcCos}(\theta) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

где n – количество слов в словаре;
 x_i – частота встречаемости i -ого слова в первом тексте;
 y_i – частота встречаемости i -ого слова во втором тексте.

Значение R интерпретируется как некоторый угол между векторами X и Y в многомерном пространстве. Если вектора совпадают, угол между ними равен 0 и значение R будет равно 1 . Если вектора не совпадают, то максимальный угол между ними будет равен 90 градусов и значение R будет равно 0 .

11.3 Пример

Задание

- Создать частотные словари произведений Бунина И. А. и Паустовского К. Г.
- Определить чьим произведением является рассказ «Теплый хлеб».

Требования к используемому ПО:

- версия ОС: Windows 10;
- MS Office 2016
- файл «Классификация текстов.xlsm».

- файл «Бунин.txt»
- файл «Паустовский.txt»
- файл «Теплый хлеб.txt»

Подготовка файлов

Указанные выше файлы скачаны с интернет в формате *.txt.

Однако даже такой простой формат может быть представлен в различной кодировке.

Поэтому:

1. Необходимо перевести их в кодировку Windows (или убедиться в том, что файлы уже находятся в этой кодировке).

2. Для этого выбранный файл загружается в Word и пересохраняется в указанной кодировке.

3. Заменить в файлах букву «ё» на букву «е».

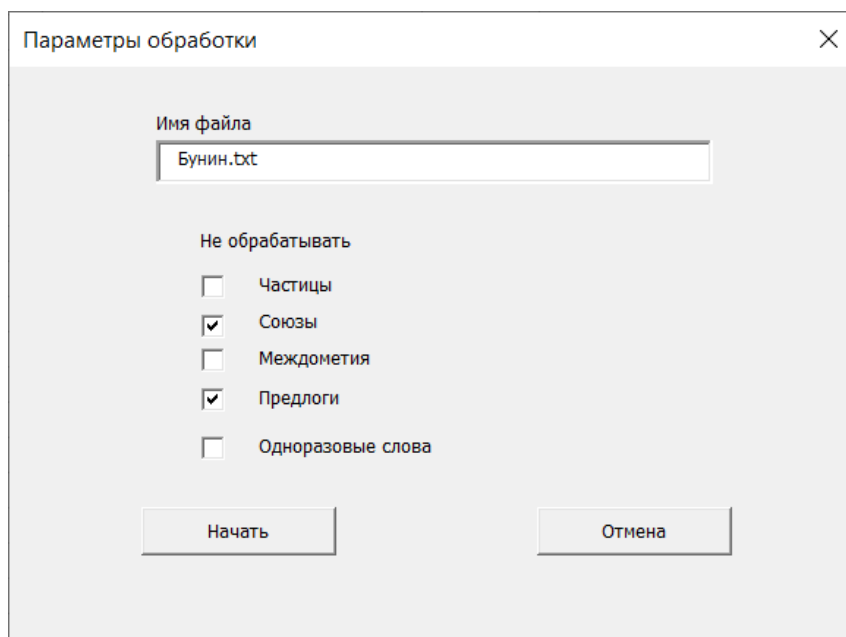
Для этого загрузить файл в «Блокнот» выделить весь текст (Ctrl/A) и с помощью пункта меню «Заменить» произвести указанную замену.

Создание частотных словарей авторов

1. Открыть файл «Классификация текстов.xlsx».

2. Нажать кнопку «Сформировать словарь»

3. В появившейся форме задать имя обрабатываемого файла и параметры обработки.



В данном случае предполагается, что частота использования союзов и предлогов у сравниваемых авторов должна быть примерно одинакова. А использование междометий и частиц специфично для каждого из них и будет характеризовать их стиль произведений.

4. Нажать кнопку «Начать».

5. Поскольку Excel не отличается быстродействием система на некоторое время «зависнет». Для указанного файла (объем 505 кб) время счета составит примерно 500 сек.

6. В результате в столбцах А и В будет сформирован частотный словарь произведений Бунина И. А. При этом в ячейке С1 будет указано общее количество слов в сформированном словаре.

7. Перенести данные из столбцов А и В куда-то в сторону (например, в столбцы К и L). Столбцы А и В – очистить.

8. Аналогично получить частотный словарь для произведений Паустовского К. Г. (объем файла – 812 кб). Время счета при этом составит, примерно, 750 сек. Общее количество слов – 13361. Перенести полученный словарь в столбцы М и N. Столбцы А и В – очистить.

9. Аналогично получить частотный словарь для файла «Теплый хлеб.txt» (объем файла 15 кб). Время счета при этом составит, примерно, 3 сек. Общее количество слов – 921. Перенести полученный словарь в столбцы I и J. Столбцы А и В – очистить.

10. Получить объединенный словарь для файла «Теплый хлеб» и произведений Бунина И. А.

11. Для этого необходимо разместить один словарь в колонках G и H, а второй – в колонках I и J. В колонках I и J уже размещен словарь для файла «Теплый хлеб». Поэтому из колонок K и L в колонки G и H переносим данные словаря по произведениям Бунина И.А.

12. Нажать кнопку «Объединить словари». В столбцах А и В будет сформирован объединенный частотный словарь по выбранным файлам.

13. Нажать кнопку «Определить сходство». В ячейке D2 появится результат расчета косинусного сходства словарей – 0,1498

14. Получить объединенный словарь для файла «Теплый хлеб» и произведений Паустовского К. Г.

15. Для этого в колонках G и H размещаем частотный словарь по произведениям Паустовского (копируем из столбцов M и N)/. В колонках I и J уже размещен словарь для файла «Теплый хлеб».

16. Нажать кнопку «Объединить словари». В столбцах А и В будет сформирован объединенный частотный словарь по выбранным файлам.

17. Нажать кнопку «Определить сходство». В ячейке D2 появится результат расчета косинусного сходства словарей – 0,3938

18. Значение косинусного сходства во втором случае больше, чем в первом. Поэтому делается вывод о том, что произведение «Теплый хлеб» относится к произведениям Паустовского К. Г.

4. Задания

1. Скачать из интернет сборникам сочинений выбранных авторов.
2. Вырезать из них по одному рассказу и записать в отдельные файлы.
3. Произвести процедуру классификации вырезанных произведений.
4. В качестве авторов может быть как русская классика (Пушкина, Лермонтов, Толстой и т. д.), так и произведения других авторов и других жанров (фантастика, детективы, любовные романы и т. п.)

Лишний раз напоминаем:

- скачанные тексты сохранить в формате *.txt в кодировке Windows;
- заменить в текстах букву «ё» на букву «е».

Раздел 2. Курсовой проект

Целью курсового проекта является закрепление и углубление теоретических знаний по дисциплине, получение практических навыков по проектированию и созданию ИИС.

Примерная тематика курсовых проектов по дисциплине (модулю) Интеллектуальные информационные системы

1. Применение экспертных систем в деятельности предприятия
2. Применение нейронных сетей в экономике.
3. Программы деловых игр. Описание и основные принципы работы.
4. Использование продукционных моделей в принятии решений.
5. Интеллектуальные интернет-технологии. Описание и принципы работы.
6. Гипертекстовые интеллектуальные информационные системы.
7. Инструментальные средства работы со знаниями.
8. Языки программирования для интеллектуальных систем и языки представления знаний.
9. Особенности естественно-языковых интеллектуальных информационных систем.
10. Использование объектно-ориентированного подхода к представлению и обработке знаний.
11. Классы прикладных систем, основанных на знаниях, и задачи, решаемые ими.
12. Основные подходы к организации баз знаний интеллектуальных систем.
13. Основные принципы работы системы распознавания текстов - Fine Reader.
14. Использование фреймовой модели представления знаний для различных аспектов деятельности.
15. Использование продукционных моделей в принятии решений.
16. Применение экспертных систем в деятельности предприятия.
17. Применение систем искусственного интеллекта в прогнозировании.
18. Системы искусственного интеллекта для распознавания образов.
19. Генетические алгоритмы.
20. Системы представления знаний в ИИС.
21. Методы представления знаний в ИИС.
22. Автоматизированные системы распознавания образов.
23. Экспертные системы.
24. Математические методы и автоматизированные системы поддержки принятия решений.
25. Нейронные сети.
26. Когнитивное моделирование.

Планируемый объем пояснительной записки – 15-25 страниц.

Законченный курсовой проект (исходный код программы в электронном виде и пояснительная записка – в бумажном виде) не позже 15-й недели семестра предъявляется руководителю. После проверки работы студенту назначается время защиты.

В случае обнаружения в программе недочетов (неоптимальное использование машинных ресурсов, недостаточно проработанный человеко-машинный интерфейс и др.), наличия в тексте пояснительной записки большого числа грамматических ошибок, а также в случае небрежного оформления текста, курсовой проект возвращается на доработку.

Общая оценка за курсовой проект проставляется с учетом работы студента в течение семестра, качества представленной работы и ее защиты.

Среднее время самостоятельной работы студента на выполнение курсового проекта 64 часа.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

А.А. ФИЛИППОВ
Ю.В. СТРОЕВА

**ПОСТРОЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ
НА ОСНОВЕ ХРАНИЛИЩ ДАННЫХ**
практикум по дисциплине
**«Интеллектуальные аналитические информационные
системы на основе хранилищ данных»**

Ульяновск
УлГТУ
2021

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.

Филиппов, Алексей Александрович

Построение интеллектуальных систем на основе хранилищ данных : практикум / А. А. Филиппов, Ю.В. Строева. – Ульяновск : УлГТУ, 2021. – 11 с.

Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Интеллектуальные аналитические информационные системы на основе хранилищ данных». Предоставлены варианты заданий, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины. Предназначен для студентов, обучающихся по направлению 09.04.03 «Прикладная информатика» (магистерская программа «Искусственный интеллект и бизнес-аналитика»).

Работа подготовлена на кафедре «Информационные системы».

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
Собеседование по лабораторным работам	5
Экзамен	5
Лабораторная работа №1	7
Контрольные вопросы к лабораторной работе	7
Лабораторная работа №2	8
Контрольные вопросы к лабораторной работе	8
Лабораторная работа №3	9
Контрольные вопросы к лабораторной работе	9
Список используемой литературы	10

ВВЕДЕНИЕ

Целью освоения дисциплины «Интеллектуальные аналитические информационные системы на основе хранилищ данных» является формирование у студентов профессиональных компетенций, связанных с использованием теоретических знаний о концепциях, принципах, подходах и моделях, положенных в основу интеллектуальных информационных систем, основанных на хранилищах данных. Особое внимание уделяется изучению архитектуры хранилищ данных, моделей хранилищ данных, моделированию хранилищ данных, проектированию хранилищ данных.

Задачами дисциплины являются:

- изучение принципов, заложенных в основу интеллектуальных информационных систем, основанных на хранилищах данных;
- формирование навыков выбора необходимого типа и состава интеллектуальных информационных систем, основанных на хранилищах данных, для решения поставленных задач;
- рассмотрение и изучение применения различных подходов к построению интеллектуальных информационных систем, основанных на хранилищах данных;
- приобретение теоретических знаний и практических навыков проектирования и разработки интеллектуальных информационных систем, основанных на хранилищах данных.

Кроме того, в результате изучения дисциплины «Интеллектуальные информационные системы на основе хранилищ данных» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне их формирования.

Тематический план дисциплины:

1. Интеллектуальные информационные системы

- 1.1. Данные.
- 1.2. Ценность и количество информации.
- 1.3. Преобразование информации.
- 1.4. Системы информационного обмена.
- 1.5. Информационные системы.
- 1.6. Информационная сеть.
- 1.7. Состав и структура информационных систем.
- 1.8. Задачи информационных систем.
- 1.9. Цели использования профилей информационной системы.
- 1.10. Профиль информационной системы.
- 1.11. Профиль пользователя информационной системы.
- 1.12. Принципы формирования профилей.
- 1.13. Структура профилей информационной системы.
- 1.14. Открытые системы.

1.15. Интеллектуальный анализ данных в информационных системах.

2. Хранилища данных

2.1. Хранилища данных.

2.2. Архитектура хранилищ данных.

2.3. Модель хранилища данных.

2.4. Жизненный цикл хранилищ данных.

2.5. Модель темпоральных (временных) данных в хранилищах данных.

2.6. Многомерное моделирование в хранилищах данных.

2.7. Физическая модель хранилища данных.

2.8. Метаданные в хранилищах данных.

2.9. Запросы к хранилищам данных.

2.10. Настройка производительности запросов к хранилищам данных.

2.11. Проектирование кубов данных.

Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки уровня знаний, умений, владений, понимания студентом основных методов и методик проектирования, разработки и поддержки информационных систем в задачах автоматизации бизнес-процессов, умения применять на практике полученных знаний. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений. Шкала оценивания имеет вид (таблица П1)

Таблица П1

Шкала и критерии оценивания решения задач на лабораторных занятиях

Оценка	Критерии
Сдано	Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку.
Не сдано	Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы.

Экзамен

Экзамен по дисциплине проводится в форме решения на компьютере (написания программы) практических задач по билетам, а также ответа на сопутствующие теоретические вопросы. Билет содержит практическое задание (задачу) для контроля освоения умений и навыков всех запланированных в ходе изучения дисциплины компетенций. Билет

формируется таким образом, чтобы в него попали практические задания, контролирующие уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица ПЗ)

Таблица ПЗ

Шкала и критерии оценивания экзамена

Оценка	Критерии
Отлично	Выставляется обучающемуся, если студент выполнил в полном объеме практическое задание и способен обосновать свои решения
Хорошо	Выставляется обучающемуся, если студент выполнил практическое задание не в полном объеме (не менее $\frac{3}{4}$) либо в полном объеме, но с некоторыми погрешностями и ошибками
Удовлетворительно	Выставляется обучающемуся, если студент выполнил практическое задание не в полном объеме (не менее $\frac{1}{2}$) либо в полном объеме, но с существенными погрешностями и ошибками
Неудовлетворительно	Выставляется обучающемуся, если студент не справился с выполнением практического задания

Лабораторная работа №1

Проектирование интеллектуальной информационной системы.

Необходимо:

1. Сформировать цель и задачи разрабатываемой системы.
2. Сформировать перечень функций разрабатываемой системы.
3. Определить состав разрабатываемой системы:
 - a. техническое обеспечение;
 - b. программное обеспечение;
 - c. лингвистическое обеспечение;
 - d. математическое обеспечение.

Контрольные вопросы к лабораторной работе

1. Что такое данные?
2. В чем различие между данными и информацией?
3. В чем заключается проблема ценности информации?
4. Что такое информационная система?
5. Какие виды обеспечения информационных систем выделяют?
6. Какие подходы к проектированию информационных систем выделяют?

Лабораторная работа №2

Проектирование хранилища данных для интеллектуальной информационной системы.

Необходимо:

1. Разработать модель данных хранилища.
2. Выбрать тип хранилища: реляционная, NoSQL.
3. Определить состав информационного обеспечения разрабатываемой системы.
4. Сформировать проект разрабатываемой системы в виде множества UML-диаграмм.

Контрольные вопросы к лабораторной работе

1. Что такое хранилище данных?
2. Какие архитектурные решения используются при построении хранилищ данных?
3. Какие модели хранилищ данных Вы знаете?
4. Из каких этапов состоит жизненный цикл хранилищ данных?
5. В чем различия между данными и метаданными?

Лабораторная работа №3

Разработка интеллектуальной информационной системы и хранилища данных.

Необходимо реализовать систему согласно проекту, разработанному ранее.

Контрольные вопросы к лабораторной работе

1. Какие виды информационных систем выделяют?
2. Что необходимо учитывать для обеспечения безопасности информационной системы?
3. Как в хранилищах данных можно представить временные данные?
4. Для чего используется многомерное моделирование?
5. В чем особенность запросов к хранилищам данных?
6. Каким образом можно повысить скорость запросов к хранилищам данных?
1. Для чего используются кубы данных?
2. Какие методы интеллектуального анализа данных Вы знаете?
3. В чем особенность интеллектуальных информационных систем?

Список используемой литературы

1. Маглинец, Ю.А. Анализ требований к автоматизированным информационным системам [Электронный ресурс] : учебное пособие / Ю.А. Маглинец. — Электрон. дан. — Москва : , 2016. — 191 с.
<https://e.lanbook.com/book/100567>
2. Марасанов, А.М. Распределенные базы и хранилища данных [Электронный ресурс] : учебное пособие / А.М. Марасанов, Н.П. Аносова, О.О. Бородин, Е.С. Гаврилов. — Электрон. дан. — Москва : , 2016. — 254 с.
<https://e.lanbook.com/book/100445>
3. Афанасьева, Татьяна Васильевна. Введение в проектирование систем интеллектуального анализа данных : учебное пособие / Т. В. Афанасьева, А. Н. Афанасьев. — Ульяновск : УлГТУ, 2017. — 64 с.
<http://venec.ulstu.ru/lib/disk/2017/231.pdf>
4. Долженко, А.И. Управление информационными системами [Электронный ресурс] : учебное пособие / А.И. Долженко. — Электрон. дан. — Москва : , 2016. — 180 с.
<https://e.lanbook.com/book/100530>
5. Туманов, В.Е. Проектирование хранилищ данных для приложений систем деловой осведомленности (Business Intelligence Systems) [Электронный ресурс] : учебное пособие / В.Е. Туманов. — Электрон. дан. — Москва : , 2016. — 957 с.
<https://e.lanbook.com/book/100389>
6. Материалы сайта OLAP.ru - Business intelligence - effective data mining & analysis
<http://www.olap.ru/basic/alpero2i.asp>

ПОСТРОЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ
НА ОСНОВЕ ХРАНИЛИЩ ДАННЫХ:

практикум к выполнению
лабораторных работ по дисциплине
«Интеллектуальные аналитические информационные системы
на основе хранилищ данных»

ФИЛИППОВ Алексей Александрович
СТРОЕВА Юлия Владимировна

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Ознакомительная практика

Профиль подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

Практическое задание

1. Процедура выполнения практического задания

Количество проводимых практических заданий в течение всего периода освоения практики	1 индивидуальное задание
Формат проведения результатов	Устно и/или Электронный
Методические рекомендации (при необходимости)	<p>1. Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с. http://venec.ulstu.ru/lib/disk/2015/198.pdf</p> <p>2. В. Г. Тронин. Планирование и управление научными проектами с применением современных ИКТ: учебное пособие: УлГТУ, 2017. http://venec.ulstu.ru/lib/disk/2017/145.pdf</p>

2. Шкала оценивания с учетом срока сдачи¹

Критерии оценки качества решения задания	Балл
Студент четко и правильно обосновывает новизну и значимость научно-исследовательской задачи, выбранных методов и инструментальных средств для решения поставленной задачи. Решение позволяет выполнять поставленное задание в полном объеме. Предоставляет полный обзор литературы по исследуемой области, включая зарубежные источники.	60
Студент недостаточно четко и полно обосновывает новизну и значимость научно-исследовательской задачи, выбранных методов и инструментальных средств для решения поставленной задачи. Либо решение позволяет выполнять поставленное задание не в полном (не менее $\frac{3}{4}$) объеме. Предоставляет недостаточно полный обзор литературы по исследуемой области, включая зарубежные источники.	45
Студент не может четко и полно обосновать новизну и значимость научно-исследовательской задачи, выбранных методов и инструментальных средств для решения поставленной задачи. Либо решение позволяет выполнять поставленное задание не в полном (не менее $\frac{1}{2}$) объеме. Обзор литературы по исследуемой области предоставлен в не в полном (не менее $\frac{1}{2}$) объеме, включая зарубежные источники.	30
Студент не может обосновать новизну и значимость научно-исследовательской задачи, выбранных методов и инструментальных средств для решения поставленной задачи. Либо решение позволяет выполнять поставленное задание не в полном (менее $\frac{1}{2}$) объеме. Обзор литературы по исследуемой области предоставлен в не в полном (менее $\frac{1}{2}$) объеме, включая зарубежные источники.	20

¹ За несвоевременную сдачу обучающемуся могут быть начислены штрафные баллы.

3. Содержание практического задания

1. Исследовать и проанализировать предметную область для решения задачи научно-исследовательского характера. Обследование предметной области включает выявление объекта, предмета и методов исследования, определение новизны и значимости исследования.
2. Подготовить обзор по направлению тематики магистерской диссертации с использованием электронных ресурсов: E-library, Google Scholar, Researchgate, Киберленинка и др.
3. Рассмотреть и проанализировать статьи зарубежных авторов, решающих аналогичную научно-исследовательскую задачу.
4. Разработать концептуальную модель информационной системы для исследования по тематике магистерской диссертации.
5. Описать формальную постановку и решение научной проблемы в виде обзора магистерской диссертации.

За время учебной практики студент должен в окончательном виде сформулировать тему магистерской диссертации и обосновать целесообразность ее разработки.

4. Перечень тем для практического задания

Тема для практического задания должна соответствовать тематике магистерской диссертации и согласовываться студентом с научным руководителем. Тема может быть предложена студентом или выбрана из предложенного научным руководителем списка. Тема диссертации должна соответствовать направлению подготовки 09.04.03.

Примерные направления исследований для выбора тем:

1. Анализ естественных языков методами машинного обучения
2. Обучение игрового агента играть в консольные игры
3. Система прогнозирования сроков разработки программного продукта
4. Разработка мобильного приложения распознавания и анализа графиков функций
5. Разработка автоматизированного рабочего места сотрудника лечебно-профилактического учреждения
6. Система интеграции задач для распределенных команд
7. Разработка системы сравнения текстовых сообщений для повышения релевантности ответа на поисковой запрос на основе нейронных сетей рекуррентного типа
8. Прогнозирование цен на рынке недвижимости
9. Разработка и исследование способов автоматизации проектирования систем анализа лингвистических временных рядов
10. Исследование и разработка системы автоматизации оценивания успешности программных проектов
11. Разработка системы автоматизации распределения производственных мощностей
12. Разработка и исследование лингвистических методов анализа текстовых ресурсов
13. Исследование и разработка методов динамической генерации графического интерфейса пользователя для веб-приложений
14. Методика разработки онтологии в предметной области программной инженерии
15. Кластеризация объектов, имеющих как числовые, так и категориальные данные, используя метод кластеризации, основанный на минимальном остовном дереве

16. Моделирование пользователя информационного ресурса для формирования рекомендаций
17. Представление и обработка нечетких временных рядов на основе Fuzzy OWL
18. Применение F-преобразования в задаче адаптации качества видео потока
19. Нейросетевой метод распознавания человека на множестве снимков
20. Разработка системы поддержки научного проекта с использованием ресурсов научной социальной сети
21. Сравнение методов машинного обучения в задаче бинарной классификации
22. Исследование применения методов выявления схожих объектов в задаче построения модели, позволяющей проводить идентификацию физических лиц
23. Разработка системы поддержки научного проекта с применением ресурсов электронной научной библиотеки
24. Прогнозирование временных рядов с использованием сочетания лингвистических и временных критериев
25. Разработка экспертной системы анализа биологических объектов
26. Интеллектуальный анализ метрик ИТ-проектов

Письменный отчет

Письменный отчет включает в себя отчет по индивидуальному практическому заданию и дневник по практике.

1. Процедура проведения

Формат проведения результатов	Отчет по индивидуальному практическому заданию - электронный и письменный, дневник - письменный
Сроки	4 недели
Методические рекомендации (при необходимости)	<ol style="list-style-type: none">1. Семушин, И. В. Письменная и устная научная коммуникация: учебное пособие / И. В. Семушин. – Ульяновск : УлГТУ, 2014. – 143 с. - Доступен также в Интернете: http://venec.ulstu.ru/lib/disk/2015/11.pdf2. Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с. http://venec.ulstu.ru/lib/disk/2015/198.pdf

3. Шкала оценивания с учетом срока сдачи¹

Критерии оценки качества решения задачи	Балл
Студент полно и аргументировано оформил письменный отчет и дневник в соответствии с выполненным практическим заданием; четко и полно дает ответы на дополнительные уточняющие вопросы.	20
Студент недостаточно чётко и полно осветил разделы в письменном отчете и в дневнике; ответил на дополнительные уточняющие вопросы с недочетами.	15
Студент допустил ошибки в письменном отчете и дневнике (при этом ошибки не должны иметь принципиального характера); неточности при ответе на уточняющие вопросы.	10
Студент дал неверные, содержащие фактические ошибки, ответы в письменном отчете и дневнике; не смог ответить на дополнительные и уточняющие вопросы или отказался отвечать.	5

¹ За несвоевременную сдачу обучающемуся могут быть начислены штрафные баллы.

4. Структура письменного отчета

В дневнике обязательно указывается тип практики, группа, ФИО студента, приказ, сроки практики, место прохождения практики, руководитель от университета.

В ходе прохождения практики студентом заполняются листы:

1. Календарный график
2. Индивидуальное задание
3. Рабочие записи.

Отчет по индивидуальному практическому заданию должен соответствовать обзору магистерской диссертации.

Структурно обзор включает следующие разделы:

1. Титульный лист.
2. Оглавление.
3. Список сокращений.
4. Основная часть.
5. Список цитируемой литературы.

Обзор включает в себя описание таких понятий, как область исследования, объект, предмет и актуальность исследования.

По содержанию основная часть обзора в магистерской диссертации должна включать следующие разделы:

1. Актуальность выбранной темы в теоретическом и практическом плане. Важно обосновать целесообразность решения вашей проблемы, связанной с недостаточным исследованием какого-то процесса, недостаточным уровнем автоматизации какого-то процесса, отсутствием в современных системах систем с такими функциональными возможностями и т. д.
2. Анализ существующего состояния проблемы на объектах исследования, выбор и обоснование математических и технологических инструментов решения; при этом анализ целесообразно проводить как по отдельным понятиям, отраженным в теме, так и по их сочетанию, с учетом специфики объектов исследования: модели и методы решения поставленной проблемы, технологии, методологии проектирования и разработки ПО. Обоснование может быть дано экспертно: приводятся высказывания, мнения авторитетных специалистов, с указанием ссылок на источники, где оно сформулировано. Или путем приведения результатов экспериментов.
3. Формулировка проблемы, степень ее разработанности в специальной научной литературе и методы ее разрешения на практике.
4. Постановка цели исследования. Обоснование и аргументация основных выводов и результатов исследования конкретной проблемы, приводящих к определению объекта, предмета, цели работы. Формулирование предложений и рекомендаций по разрешению изучаемой проблемы в виде комплекса задач, детализирующих цель исследования.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Факультет информационных систем и технологий

Кафедра Информационные системы

Отчет по практике

защищен с оценкой _____

(оценка прописью)

Руководитель

практики от

университета _____

(должность)

(подпись)

(ФИО)

ОТЧЕТ по _____ практике
(вид практики)

Студента _____
(ФИО)

Направление (специальность,
профиль) подготовки _____

Группа _____

Место прохождения практики _____

(наименование профильной организации, подразделение университета)

Ульяновск 20__ г.

Зачет с оценкой

1. Процедура проведения

Общее количество вопросов к зачету с оценкой	14 вопросов
Количество вопросов в билете	2 вопроса
Наличие задач в билете	нет
Формат проведения	Устно
Методические рекомендации (при необходимости)	-

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
выставляется обучающемуся, если студент показал глубокие знания материала по поставленному вопросу, грамотно логично и стройно его излагает.	Отлично
выставляется обучающемуся, если студент твердо знает материал, грамотно его излагает, не допускает существенных неточностей в ответе на вопрос.	Хорошо
выставляется обучающемуся, если студент показывает знания только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допускает отдельные неточности.	Удовлетворительно
выставляется обучающемуся, если студент допускает грубые ошибки в ответе на поставленный вопрос или отказался отвечать.	Неудовлетворительно

3. Вопросы к зачету с оценкой

1. Охарактеризуйте цель и задачи научно-исследовательской работы.
2. Приведите соотношение объекта и предмета исследования.
3. Как конспектировать работы и составлять научный отчет?
4. Охарактеризуйте системы «Антиплагиат».
5. Приведите электронные ресурсы (библиотеки, научные социальные сети), их возможности для поддержки подготовки магистерской диссертации.
6. Как презентовать научные результаты?
7. Опишите методику подготовки обзора магистерской диссертации.
8. Перечислите и охарактеризуйте этапы научного исследования.
9. Как идентифицировать научную проблему?
10. Как конкретизировать проблему магистерской диссертации?
11. В чем состоит цель, объект и предмет вашей магистерской диссертации?

12. Какие методы и модели используются или планируется использовать в вашей магистерской диссертации?
13. Какие вычислительные эксперименты планируются выполнить в вашей магистерской диссертации?
14. Какие информационные технологии планируется применить в вашей магистерской диссертации?

МИНОБРНАУКИ РОССИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

Кафедра _____

Д Н Е В Н И К

_____ (вид практики)

СТУДЕНТА _____
(фамилия, имя отчество)

ФАКУЛЬТЕТ _____ КУРС _____

ГРУППА _____

СПЕЦИАЛЬНОСТЬ _____
НАПРАВЛЕНИЕ, ПРОФИЛЬ _____ (наименование)

Ульяновск 20 ____ г.

ОФОРМЛЕНИЕ СТУДЕНТА НА ПРАКТИКУ

Приказ о направлении на практику № _____ от « _____ » _____ 20__ г.

Инструктажи по технике безопасности:

- вводный от УЛГТУ _____
(должность, Ф.И.О. ответственного лица, подпись)
« _____ » _____ 20__ г.
- вводный в профильной _____
организации (должность, Ф.И.О. ответственного лица, подпись)
« _____ » _____ 20__ г.
- на рабочем месте _____
(должность, Ф.И.О. ответственного лица, подпись)
« _____ » _____ 20__ г.
- повторный _____
(должность, Ф.И.О. ответственного лица, подпись)
« _____ » _____ 20__ г.

ПРЕДПИСАНИЕ НА ПРАКТИКУСтудент _____
(фамилия, имя, отчество)направлен на _____ практику
(вид практики)

в гор. _____ на _____

(наименование профильной организации)

Срок практики с _____ по _____
включая проезд туда и обратно**РУКОВОДИТЕЛЬ ПРАКТИКИ ОТ УНИВЕРСИТЕТА**_____
(должность, фамилия, имя, отчество)Декан факультета _____
(подпись)

М.П.

РУКОВОДИТЕЛЬ ПРАКТИКИ ОТ ПРОФИЛЬНОЙ ОРГАНИЗАЦИИ_____
(должность, фамилия, имя, отчество)

согласно приказа № _____ от _____

ПРИБЫЛ НА ПРЕДПРИЯТИЕ	УБЫЛ С ПРЕДПРИЯТИЯ
" _____ " _____ 20 г. _____	" _____ " _____ 20 г. _____
подпись	подпись
М.П.	М.П.

Памятка

1. Основные положения по прохождению практики

1. За неделю до начала практики руководитель от кафедры проводит собрание и инструктаж по ОТ и ТБ студентов и выдает:
 - заполненный дневник (командировочное удостоверение);
 - два экземпляра программ практики на группу (один для студентов и один для руководителя практики от профильной организации);
 - индивидуальные задания по практике;
 - направления на практику.
2. За 3-4 дня до начала практики студенту необходимо явиться на предприятие для оформления пропуска (при себе иметь паспорт, фото и другие документы), пройти инструктаж по технике безопасности и противопожарной безопасности. Студент должен представить руководителю от профильной организации дневник и программу, ознакомить его с содержанием индивидуальных заданий, ознакомиться с рабочим местом, правилами эксплуатации оборудования и уточнить план прохождения практики.
3. Студент во время практики обязан строго соблюдать правила внутреннего трудового распорядка профильной организации. О всех отлучках со своего места практики ставить в известность руководителя практики от профильной организации.
4. Отчет по практике составляется студентом в соответствии с указанием программ, индивидуальных заданий и дополнительными указаниями руководителей практики от университета и профильной организации.
5. Содержание преддипломной практики определяется темой выпускной квалификационной работы. Во время прохождения преддипломной практики студенты собирают необходимый материал о производственной деятельности профильной организации, о выполняемых исследовательских работах, о проведении новых разработок и используют этот материал при написании выпускной квалификационной работы.
6. По итогам аттестации практики студенту университета выставляется дифференцированная оценка (отлично, хорошо, удовлетворительно).

2. Правила ведения дневника

1. Дневник является основным документом студента во время прохождения практики.
2. Во время практики студент ведет рабочие записи и составляет отчет по практике.
3. Руководитель практики от университета контролирует оформление дневника.
4. По окончании практики дневник должен быть:
 - просмотрен и подписан руководителями практики (с составлением отзыва-характеристики руководителем от профильной организации и заключением от руководителя кафедры);
 - подписан общим руководителем от профильной организации (начальником отдела технического обучения, главным инженером или другим уполномоченным лицом);
 - получен студентом в окончательно оформленном виде и в установленный срок сдан на кафедру.

СОВМЕСТНЫЙ РАБОЧИЙ ГРАФИК(ПЛАН)

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ОРГАНИЗАЦИЯ МАГИСТЕРСКИХ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ РАБОТ

Методические рекомендации

Составитель Т. В. Афанасьева

Ульяновск
УлГТУ
2015

УДК 004.925.8 (076)
ББК 32.973.26-018.2я7
О-64

Рецензент доцент кафедры ПМИ УлГТУ,
канд. техн. наук Кувайскова Ю. Е.

*Рекомендовано научно-методической комиссией
факультета информационных систем и технологий
в качестве методических рекомендаций*

О-64 Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с.

Методические рекомендации написаны в соответствии с требованиями государственного стандарта для магистрантов по направлению «Программная инженерия» и «Прикладная информатика».

В рекомендациях приведены общие сведения о целях, задачах научного исследования, описаны основные объекты, методы и результаты научно-исследовательских работ.

Работа подготовлена на кафедре «Информационные системы» УлГТУ.

**УДК 004.925.8 (076)
ББК 32.973.26-018.2я7**

© Афанасьева Т. В., составление, 2015
© Оформление. УлГТУ, 2015

ОГЛАВЛЕНИЕ

Общие положения	4
Этапы самостоятельной работы магистров по НИР	5
Цели и задачи НИР	5
Требования к магистерской диссертации	7
Требования к обзору.....	11
Методика подготовки обзора	14
Рекомендации по теоретическому исследованию	17
Основы вычислительного эксперимента	18
Заключение	19
Список литературы	19
Приложение 1. Примеры обзоров	20
Приложение 2. Речевые клише	35

ОБЩИЕ ПОЛОЖЕНИЯ

Цель методических рекомендаций – изложение процесса организации самостоятельной работы магистров в ходе их учебной и научно-исследовательской деятельности.

Самостоятельная работа магистра существенно затрудняется без помощи, квалифицированного управления ее ходом. Такую помощь оказывает научный руководитель (консультант). Однако ведущую роль в учении и проведении исследования должен играть сам магистр. Иными словами, контроль за ходом и результатами его деятельности должен постепенно переходить в самоконтроль.

Самостоятельная работа магистров – индивидуальная учебная или исследовательская деятельность, осуществляется в рамках дисциплин учебного плана, при выполнении курсовых работ и проектов, в при выполнении НИР, в период производственной практики и в процессе подготовки и написании магистерской диссертации, в ходе которой магистр активно воспринимает, осмысливает полученную информацию, решает теоретические и практические задачи.

Наиболее распространенные виды самостоятельной работы: работа с учебником, справочной литературой или с первоисточниками, наблюдения, лабораторные занятия, проведение эксперимента, моделирование и др. Исходя из внутреннего содержания, можно выделить три ступени самостоятельной деятельности: репродуктивная (подражание образцам, изложение своими словами содержания учебника, научных статей и др.); продуктивная (самостоятельное применение приобретенных знаний для решения задач, выходящих за пределы типовых); творческая (применение знаний в совершенно новых условиях: выдвижение гипотез, проведение экспериментов, моделирование ситуаций и процессов и др.).

ЭТАПЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ МАГИСТРОВ ПО НИР

Организация самостоятельной работы осуществляется в несколько этапов.

Первый этап – постановка и обсуждение с научным руководителем целей, задач, хода работы; *второй этап* – период самоорганизации магистранта и его непосредственная деятельность по решению задач; *третий этап* – подведение итогов и оценка полученных результатов.

Приступая к обучению, магистрант составляет индивидуальный план своей работы. Индивидуальный план учебной и научной работы магистранта составляется на два года при непосредственном участии научного руководителя и утверждается на заседании кафедры. В процессе составления индивидуального плана учебной и научной работы необходимо уяснить очередность и логическую последовательность намеченных работ. Последовательность заданий индивидуального плана со временем может измениться с тем условием, чтобы все они были выполнены.

ЦЕЛИ И ЗАДАЧИ НИР

Основной целью организации и развития системы научно-исследовательской работы является повышение уровня научной подготовки магистров и выявление талантливой молодежи для последующего обучения и пополнения педагогических и научных кадров вузов, других учреждений и организаций страны на основе новейших достижений научно-технического прогресса, экономической мысли и культурного развития.

Основными задачами организации и развития системы НИР для магистерских программ обучения являются:

1. Обеспечение интеграции учебных занятий и научно-исследовательской работы студентов;
2. Осуществление органичного единства обучения и подготовки студентов к творческому, научному и педагогическому труду;
3. Создание условий для раскрытия и реализации личностных творческих способностей студенческой молодежи;
4. Расширение массовости и повышение результативности участия магистров в научной деятельности;
5. Отбор талантливой молодежи, проявившей способности и стремление к научной и педагогической деятельности;
6. Формирование и развитие у студентов качеств научно-педагогических и научных работников;
7. Формирование и развитие у будущих выпускников: умения вести научно обоснованную профессиональную работу на предприятиях и в учреждениях любых организационно-правовых форм; способности быстрой адаптации, приложения полученных знаний и умений при изменяющихся требованиях к своей деятельности; освоения методологии и практики планирования, выбора оптимальных решений в условиях рыночных отношений; готовности и способности к повышению квалификации и переподготовке;
8. Подготовка руководителей высокой квалификации – магистров, имеющих навыки проектно-конструкторской работы, умеющих грамотно разработать и реализовать конкретные научно-практические мероприятия на производстве, обладающих навыками самоуправления;
9. Повышение массовости и эффективности НИР в университете путем привлечения студентов к исследованиям по наиболее приоритетным

направлениям науки, связанным с современными потребностями общества и государства;

10. Поиск и реализация источников финансирования, в том числе за счет средств, получаемых из внебюджетных источников и инновационной деятельности вузов, совершенствование форм и методов привлечения их к НИР;

11. Развитие научных межвузовских связей как внутри страны, так и со странами ближнего и дальнего зарубежья.

Важным является выбор тематики исследования и ориентация НИР на подготовку к написанию магистерской диссертации.

ТРЕБОВАНИЯ К МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

При подготовке обзора по теме магистерской диссертации важно знать основные требования к магистерским диссертациям и основные понятия в содержательном плане.

Диссертация (от лат. — «dissertation» — рассуждение, исследование) — специальная форма научного произведения, имеющего квалификационный характер, подготовленная для публичной защиты и получения ученой степени.

Магистерская диссертация должна отличаться от бакалаврской работы глубокой теоретической проработкой проблемы, от дипломной работы специалиста – научной направленностью. Магистерская диссертация выполняется студентом по материалам, собранным им лично за период обучения и научно-исследовательской практики.

В магистерской диссертации необходимо выделить, определить и описать (см. примеры в приложении 1):

- 1) Тему исследования;

- 2) Область исследования;
- 3) Объект, предмет и прототип исследования (на основе обзора);
- 4) Цель исследования (на основе обзора);
- 5) Решаемые задачи (на основе обзора);
- 6) Научный результат;
- 7) Практический результат;
- 8) Подтверждение результатов.

Тема диссертации должна быть актуальной, представлять научный и практический интерес и соответствовать профилю магистерской программы. Обычно в теме указывается и проблема и метод ее решения.

Темы магистерских диссертаций, научные руководители и консультанты определяются выпускающими кафедрами (оформляются протоколом заседания кафедры), рассматриваются и утверждаются ученым советом (УС) факультета (оформляются протоколом заседания УС факультета) и приказом ректора после зачисления в магистратуру (по представлению отдела магистратуры на основании выписок из протокола заседания УС факультета).

Научными руководителями магистерских диссертаций назначаются ведущие преподаватели университета, имеющие ученую степень и ученое звание, ведущие научные исследования по тематике магистерских программ.

В обязанности научного руководителя входит:

1. Помощь в формировании темы диссертации и разработке плана работы над ней;
2. Систематическое консультирование магистранта;

3. Контроль за ходом работы над диссертацией, соответствия выполняемых работ плану;

4. Предоставление отзывов о результатах работы магистранта над диссертацией при промежуточной и итоговой аттестации;

5. Участие в обсуждении магистерских диссертаций на заседании кафедры.

Обычно сначала формируется обобщенная тема, которая к завершению обучения конкретизируется.

Областью исследования магистерской диссертации направления «Программная инженерия» является проектирование, разработка и исследование информационно-вычислительных систем, а магистерской диссертации направления «Прикладная информатика» – проектирование, разработка и исследование прикладных информационных процессов.

В каждой области в теме у магистранта есть своя специфика, которая определяет объект исследования.

Объект исследования – процесс или явление, порождающие проблемную ситуацию (это может быть недостаточно разработанная задача, задача, для которой нет алгоритмов и программного обеспечения (ПО)) и избранные для изучения с учетом соответствующего стандарта обучения.

Обычно объект исследования определяется в результате выполнения обзора как указание на целесообразность решения некоторой проблемы, связанной с недостаточным исследованием какого-то процесса, недостаточным уровнем автоматизации какого-то процесса, отсутствием в современных системах систем с такими функциональными возможностями и т. д.

Объектом исследования магистранта являются:

1. Алгоритмы решения задач;
2. Программный проект (проект разработки программного продукта);
3. Программный продукт (создаваемое программное обеспечение);
4. Процессы жизненного цикла программного продукта;
5. Методы и инструменты разработки программного продукта;
6. Методы управления персоналом.

Предмет исследования – все то, что находится в границах объекта исследования в определенном аспекте рассмотрения. Предмет исследования конкретизирует объект исследования путем определения средств для решения проблемы. Обзор средств, методов и технологий для обоснования предмета необходимо также привести в обзоре. Здесь формулируется прототип вашего исследования, то есть то, что уже есть, широко используется, но вы планируете этот прототип улучшить. Важно, чтобы прототип представлял апробированные решения.

Цель исследования. Обычно это одно предложение, интегрирующее тематику, объект и предмет исследования для достижения новых преимуществ.

Решаемые задачи. Для достижения цели необходимо решить ряд задач, таких как

1. Анализ объекта исследования, по сути, это выполнение обзора состояния в области определенной проблемы, методов их решения, технологий, алгоритмов и программных средств, используемых для решения проблемы.
2. Описание научного результата, который имеет отличия от прототипа в лучшую сторону.

3. Описание практического результата в виде конкретного программного обеспечения.

Научный результат. Это может быть алгоритм, методика (методика проектирования, тестирования, анализа, управления, оценивания), архитектура ПО и др. Важно обосновать, почему именно так предлагается решать обозначенную в цели проблему. И показать, чем ваше решение отличается от прототипа.

Практический результат. Здесь описывается, как вы создавали свое ПО, проектирование ПО, оценка ПО, архитектура, технологии, примеры использования. Если в рамках работы проводится реализация некоторого программного средства, то в разделе «Описание практической части» обязательно должна быть описана его программная реализация, в частности: приведены обоснования выбранного инструментария; приведена с иллюстрацией общая архитектура разработанного средства; приведена с иллюстрацией схема работы средства; если осуществляется доработка существующего средства, то должны быть описаны новые возможности/улучшения, реализованные в данной работе. обязательно должны быть приведены характеристики функционирования (например, сложность, производительность, время реакции и т. д.).

Требования к обзору

Результаты НИР могут быть оформлены в виде обзора, описания научных результатов и вычислительных экспериментов.

Обзор – научный документ, содержащий систематизированные научные данные по какой-либо теме, полученные в итоге анализа первоисточников. Знакомит с современным состоянием научной проблемы и перспективами ее развития.

Обзор литературы по теме должен показать умение магистранта систематизировать источники, критически их рассматривать, выделять существенное, определять изученность темы.

Структурно обзор включает следующие разделы:

1. Титульный лист (УлГТУ, каф.ИС, Обзор по теме магистерской диссертации направления «Программная инженерия» или «Прикладная информатика», научный руководитель, его ученая степень, ученое звание, «ФИО», выполнил «Магистрант ФИО», Ульяновск, год выполнения).
2. Оглавление.
3. Список сокращений.
4. Основная часть.
5. Список цитируемой литературы.

Магистрант должен особо подчеркнуть те вопросы, которые остались неразрешенными, определяя свое место в решении проблемы. Обзор литературы должен привести к выводу, что именно данная тема до сих пор не раскрыта (раскрыта частично, в другом аспекте) и нуждается в дальнейшей разработке.

Обзор включает в себя описание таких понятий, как область исследования, объект и предмет исследования.

По содержанию основная часть обзора в магистерской диссертации должна включать следующие разделы:

1. Актуальность выбранной темы в теоретическом и практическом плане. Важно обосновать целесообразность решения вашей проблемы, связанной с недостаточным исследованием какого-то процесса, недостаточным уровнем автоматизации какого-то процесса, отсутствием в

современных системах систем с такими функциональными возможностями и т. д.

Например, если принимается решение о создании некоторого программного средства, то необходимо показать, что не существует средства, обладающего нужными характеристиками.

2. Анализ существующего состояния проблемы на объектах исследования, выбор и обоснование математических и технологических инструментов решения; при этом анализ целесообразно проводить как по отдельным понятиям, отраженным в теме, так и по их сочетанию, с учетом специфики объектов направления «Программная инженерия»: модели и методы решения поставленной проблемы, технологии, методологии проектирования и разработки ПО. Обоснование может быть дано экспертно: приводятся высказывания, мнения авторитетных специалистов, с указанием ссылок на источники, где оно сформулировано. Или путем приведения результатов экспериментов.

3. Формулировка проблемы, степень ее разработанности в специальной научной литературе и методы ее разрешения на практике. Результаты можно представить в виде таблицы.

4. Постановка цели исследования. Обоснование и аргументация основных выводов и результатов исследования конкретной проблемы, приводящих к определению объекта, предмета, цели работы. Формулирование предложений и рекомендаций по разрешению изучаемой проблемы в виде комплекса задач, детализирующих цель исследования.

Методика подготовки обзора

Техника фиксации и обработки информации

Наиболее рациональными видами фиксирования информации большинством исследователей признаются цитаты, тезисы, конспекты, аннотации, рефераты и др.

Цитата – точная, буквальная выдержка из какого-либо текста с подробной ссылкой на источник (автор, заглавие источника, библиографические данные, цитируемые страницы). Выписки рекомендуется делать на одинакового формата карточках, лучше, плотной бумаги. На карточку, как правило, заносятся один или несколько фактов, идей, мыслей, касающихся определенного вопроса. Заполняется карточка на одной стороне. Другая может быть использована для соответствующих замечаний (комментариев, изложения другой точки зрения, противоположных фактов и др.). Карточки систематизируются и хранятся в электронном и в напечатанном виде. Для удобства пользования на карточках следует указать шифр, номер или название темы, раздела, проблемы и т. п.

Достоинства карточек видится в том, что они, во-первых, представляют собой отобранную и приведенную в систему наиболее ценную информацию; во-вторых, эта информация многократного и разнообразного применения: содержание карточки можно использовать для доклада, реферата, написания научной статьи и т. д.; в-третьих, карточками очень удобно пользоваться, так как они небольшие по размерам и не сброшюрованы.

Тезисы – кратко сформулированные основные положения, идеи доклада, научной работы, лекции.

Конспект – письменное изложение (может быть своими словами) содержания научной работы, лекции, доклада и др.

Аннотация – краткое разъяснительное или критическое изложение содержания, краткая характеристика и объявление назначения книги, статьи, рукописи.

Подготовка обзора требует большого объема времени как для поиска, так и для анализа и оформления результата. Этот процесс включает следующие этапы:

1. Формулировка того, что надо искать: программы, реализующие вашу тематику, модели, методы и алгоритмы. Выдвинуть предположения, что есть объект исследования, какие задачи решаются для объекта. Это итерационный процесс, что надо искать, уточняется после предварительного этапа.

2. Предварительный этап.

a. Поиск источников по теме и решаемым задачам. Обычно через Интернет (15-20 источников). Как только вы находите релевантный материал, сразу же копируете ссылку, затем нужный материал (текстовый, графический).

b. Затем распечатываете и прочитываете. Если недостаточно материала, то переход на первый этап или по ссылкам от найденных источников, иначе переход на третий этап.

c. Определение, какие задачи решаются для объекта исследования. Выдвижение предположения, какие задачи в настоящее время недостаточно решены для объекта исследования. Оформление в виде таблиц (задачи-способы решения, (задачи-ПО). Формулировка предварительной цели.

3. Перевод на иностранный язык темы, объекта, задач (примерно 10 источников).

а. Поиск в «нерусскоязычном» Интернете, аналогично п. 2. Можно использовать ресурсы computer.org, researchGate, eLibrary и др.

б. Перевод найденных материалов.

4. Дополнение обзора печатными материалами (15 - 20 источников). Обычно в виде книг, монографий, статей, диссертаций (Необходимо, чтобы большинство источников было издано не позднее 2009 года). В интернет-магазинах обычно приводятся для литературы небольшие рефераты, аннотации или оглавления, по которым можно составить мнение, о чем этот источник. Можно посмотреть в библиотеке через электронный каталог, через другие электронные библиотеки. Рекомендуемые журналы можно найти в eLibrary.ru.

5. Оформление списка используемых источников по алфавиту, начиная с иностранных авторов (ссылки должны быть оформлены в соответствии с ГОСТ Р 7.1-2003. Источники должны содержать как электронные ресурсы, так и печатную литературу в виде книг, монографий, статей, диссертаций.

6. Оформление текста в соответствии с общепринятыми требованиями к научной документации.

7. Распечатка обзора, его передача руководителю обсуждения.

Рекомендации по теоретическому исследованию

Теоретический анализ и синтез

Анализ заключается в разложении изучаемого явления на составляющие элементы, позволяет вскрыть внутреннюю структуру объекта.

Синтез позволяет воссоздать предмет как систему связей и взаимодействий с выделением наиболее существенных из них. Особенность метода теоретического анализа и синтеза в исследованиях, в любом варианте применения, заключается в его универсальных возможностях рассматривать явления и процессы действительности в их самых сложных сочетаниях, выделять наиболее существенные признаки и свойства, связи и отношения, устанавливать закономерности их развития.

Путем анализа и синтеза можно «увидеть» объективное содержание, объективные тенденции, несоответствия, предвидеть такие формы и стадии процесса, которые запроектированы, но пока еще реально не существуют.

Абстрагирование – процесс мысленного отвлечения какого-либо свойства или признака предмета от самого предмета, от других его свойств (например, исследователь аналитическим путем выделяет из всего многообразия признаков процесса программирования один – доступность программного продукта – и рассматривает его самостоятельно).

Моделирование – воспроизведение некоторых характеристик на другом объекте, созданном специально с целью изучения. Этот второй объект называется моделью (идеальной или материальной). Открывается возможность переноса информации, полученной при изучении модели на прототип. Необходимо помнить, что любая модель беднее прототипа, что она отражает лишь его отдельные стороны и связи. Исследователь, изучив

характерные черты реальных процессов и их тенденций, делает их мысленную компоновку, то есть моделирует состояние изучаемой системы. Можно выделить множество методов исследования, опирающихся на теоретические основы математического моделирования:

1. Вычислительной математики (в том числе, численные методы),
2. Дискретной математики,
3. Математической логики,
4. Теории вероятности и математической статистики и др.

Основы вычислительного эксперимента

Применение вычислительных методов решения задачи – современный подход к экспериментированию в рамках НИР.

Можно выделить следующие этапы проведения вычислительного эксперимента:

1. Построение математической модели исследуемого объекта (сюда же относится и анализ модели, выяснение корректности поставленной математической задачи);
2. Построение вычислительного алгоритма – метода приближенного решения поставленной задачи и его обоснование;
3. Программирование алгоритма на ЭВМ и его тестирование;
4. Проведение серии расчетов с варьированием определяющих параметров исходной задачи и алгоритма;
5. Анализ полученных результатов и погрешностей.

ЗАКЛЮЧЕНИЕ

Рассмотрены вопросы организации самостоятельной работы магистрантов при проведении НИР. Указанные вопросы актуальны для каждого магистранта и поддержаны большим количеством часов. Научные исследования необходимы для подготовки и написания магистерской диссертации практически в каждом семестре обучения по программам магистратуры. НИР относится к дисциплинам вариативной части программы магистратуры, которые определяют направленность (профиль) программы.

Список литературы

1. ГОСТ Р 7.32-2001 СИБИД. Отчет о научно-исследовательской работе. Структура и правила оформления. – М., 2002.

2. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.04.04 Программная инженерия (уровень магистратуры).

3. ГОСТ 15.101-98 Порядок выполнения научно-исследовательских работ. – М., 1999.

5. Правила цитирования. Доступно по адресу <http://www.ulstu.ru/main/view/article/16403> (дата обращения: 12.12.2014).

6. Методические рекомендации по проектированию оценочных средств для реализации многоуровневых образовательных программ ВПО при компетентностном подходе / В.А. Богословский, Е.В. Караваева, Е.Н. Ковтун и др. – М. : Изд-во МГУ, 2007. – 148 с.

7. Научно-исследовательская деятельность в вузе: (основные понятия, этапы, требования) / Н. В. Шестак, Е. В. Чмыхова. – М.: Изд-во МГУ, 2007. – 179 с.

8. ГОСТ Р 7.0.11-2011 СИБИД. Диссертация и автореферат диссертации. Структура и правила оформления. – М., 2011.

Примеры обзоров и представление об основных понятиях диссертации

Пример 1

ТЕМА ДИССЕРТАЦИИ: ИССЛЕДОВАНИЕ И РАЗРАБОТКА НЕЧЕТКОЙ МОДЕЛИ И КОМПЛЕКСА ПРОГРАММ ЭКОЛОГИЧЕСКОЙ ЭКСПЕРТИЗЫ ГОРНОДОБЫВАЮЩЕГО ПРОИЗВОДСТВА

В системах управления деятельностью сложных явлений и процессов, таких как производственная деятельность промышленного предприятия, процессы принятия решений часто протекают в условиях неопределенности и основываются на экспертизе, которая осуществляется в рамках экспертной деятельности, и ее результатом выступают экспертные оценки. Многие реальные процессы имеют иерархическую структуру. Изучение таких структур и применение полученных результатов для анализа реальных объектов и процессов отражено в работах Саати [1] и Такахаро [2]. Методы теории нечетких множеств являются удобным средством моделирования, анализа и синтеза человеко-компьютерных систем. Поэтому изучение нечетких иерархических систем представляется актуальной задачей, имеющей не только теоретической, но и практический интерес.

Для теории нечетких множеств основополагающим понятием является понятие нечеткого множества, которое характеризуется функцией принадлежности. Посредством нечеткого множества можно строго описывать присущие языку человека расплывчатые элементы, без

Продолжение прил. 1

формализации которых нет надежды существенно продвинуться вперед в моделировании интеллектуальных процессов.

Сегодня во всем мире принято уделять большое внимание проблеме экологии. Деятельность человека по освоению природной среды породила не только новые возможности роста благосостояния человечества, но и привела к глубокому кризису состояния окружающей среды. Поскольку ни один руководитель не будет работать себе в ущерб, важно соблюсти баланс между экологической безопасностью, требующей денежных вливаний, и экономической эффективностью деятельности хозяйствующего субъекта. Вот почему в качестве прикладной области исследования был выбран анализ экологической безопасности горнопромышленного производства на основе методики оценки экологических платежей предприятия.

Актуальность проблемы

В последние годы определился и практически осуществился переход от автоматизированных систем анализа состояния сложных технических систем к экспертным системам, работающим на основе логического приближенного вывода. Отмеченная тенденция связана с резким усложнением современных технических систем, и, в частности, экологической безопасности горнодобывающих производств. В связи с этим исследование, назначением которого является разработка математической модели и комплекса программ экологической безопасности горнодобывающего производства на основе иерархического нечеткого вывода, является, несомненно, актуальной.

Цель диссертационной работы

Целью диссертационной работы является разработка математической модели экологической безопасности горнодобывающего производства и комплекса программ экспертизы факторов такого производства для нечетко заданных исходных данных.

Задачи исследования

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать факторы горнодобывающего производства, значимые для формирования математической модели экологической безопасности, выполнить сравнительный анализ существующих методов и программ экспертной оценки экологической безопасности;

2. Разработать модель экологической безопасности горнодобывающего производства, позволяющую проводить экспертную оценку при нечетко заданных исходных данных;

3. Разработать механизм иерархического нечеткого логического вывода экспертных систем;

4. Разработать комплекс программ, реализующий экспертную систему экологической безопасности горнодобывающего производства.

Методы исследования.

Имитационное моделирование, методы математической статистики, нечеткая логика, объектно-ориентированный подход при создании комплекса программ.

Продолжение прил. 1

Научная новизна положений, выносимых на защиту:

1. Создан механизм и алгоритмы иерархического нечеткого вывода, позволяющий осуществлять экспертизу объектов с нечетко заданными исходными данными.
2. Разработана объектно-ориентированная архитектура комплекса программ экспертной системы экологической безопасности.

Практическая значимость работы

Практическая значимость работы заключается в том, что разработанная модель экологической безопасности горнодобывающего производства на основе иерархического нечеткого вывода внедрена в практику работы открытого акционерного общества «Кварц» (пос. Силикатный, Сенгилеевский район, Ульяновская обл.) и эксперта, члена НП «Горнопромышленники России», Танеева Ф.Г, в форме комплекса программ.

Реализация результатов работы

Результаты работы реализованы в виде комплекса программ JFuzzyTool1.0, зарегистрированного Федеральной службой по интеллектуальной собственности, патентам и товарным знакам (№20076144448 от 23.10.2007 г.).

Апробация работы

Основные положения и результаты диссертации докладывались, обсуждались и получили одобрение на 10 международных конференциях. Неоднократно докладывались на научно-технических конференциях УлГТУ «Вузовская наука в современных условиях».

Публикации

По теме диссертации опубликованы 5 работ.

Продолжение прил. 1

Во введении рассмотрена актуальность проблемы создания экспертных систем анализа экологической безопасности горнодобывающего производства, сформулированы объекты исследования и предмет исследования, изложены цель и соответствующие ей задачи работы.

Первая глава «Обзор современного состояния экспертных систем анализа экологической безопасности» посвящена анализу существующих моделей, используемых методов вычислительного интеллекта, инструментов для создания экспертных систем. Изложено текущее состояние различных исследований, ведущихся в этом направлении. Даны основные понятия в области экологической безопасности промышленного предприятия, рассмотрены элементы структуры экспертной системы.

Таблица 1

Сравнение методов вычислительного интеллекта

Метод вычислительного интеллекта	Обязательное наличие обучающих выборок	Обязательное наличие эксперта	Способ построения функций принадлежности и правил нечеткого вывода
Нечеткие нейронные сети: J.J.Buckley, J.S. Roger Jang	+	-	Автоматический. С помощью соответствующих алгоритмов обучения сетей
Иерархический нечеткий вывод: V.Totta, M.Brown, H.Rainer, C.Wei, L.X.Wang, L.C.Lin, X.-J.Zeng, G.-Y.Lee	-	+	Вручную. С помощью методов построения функций принадлежности. Правила формируются на основании утверждений эксперта в виде слов естественного языка

Несмотря на явные преимущества использования гибридных технологий (нечетких нейронных сетей, см. табл. 1), для решения рассматриваемой задачи они не подходят ввиду отсутствия обучающих

Продолжение прил. 1

выборок. Сделаны выводы о том, что использование пакетов различных классов позволяет строить адекватные модели современных систем. Языки моделирования общего назначения обладают наиболее широкими функциональными возможностями для любых предметных областей, могут описывать любые свойства компонентов и программировать их взаимодействия. Визуальные системы моделирования более наглядны и современны, но обладают меньшей функциональностью. Специализированные системы в этой области на сегодняшний день учитывают только специфику предприятий нефтедобывающей отрасли.

Таблица 2.
Сравнение оболочек для проектирования экспертных систем

Название оболочки	Кросс-платформенность	Сложность адаптации системы	Возможность создания иерархий	Механизмы решения задачи
Эталон	+	Пользователь. Без ограничений	+	Механизмы логического вывода
FuzzyCLIPS	+	Программист. Без ограничений	+	Нечеткий логический вывод
OPS5	–	Программист. Не предназначена для решения конкретных задач	–	Цикл распознавания пишет пользователь
BABYLON	+	Программист. Только для задач диагностики	–	Фреймы
WindExS	–	Пользователь. Без ограничений	+	Любые механизмы логического вывода, в т.ч. нечеткий

Из табл. 2 становится ясно, что наилучшим образом под выбранный эталон подходят среды FuzzyCLIPS и WindExS. Недостатком первой является то, что для адаптации экспертной системы, созданной в этой оболочке, необходимо прибегать к помощи программиста. А вторая не

Продолжение прил. 1

переносима на различные платформы. Мы решаем конкретную задачу построения нечеткой экспертной системы анализа экологической безопасности горнодобывающего предприятия. Нам необходимо, чтобы система могла быстро адаптироваться к изменениям, происходящим при принятии нового Федерального закона, изменениям ГОСТа, отраслевого положения или нового порядка расчета и начисления экологических платежей, и чтобы внести необходимые изменения мог конечный пользователь нашей системы. Так как прикладная область имеет иерархическую структуру, то экспертная система тоже должна быть иерархической.

В результате проведенного исследования существующих средств проектирования экспертных систем было принято решение о создании собственного комплекса программ, сочетающего в себе лучшие черты существующих на сегодняшний день оболочек: кросс-платформенность, возможность создания иерархий, простоту адаптации системы, быстрое действие и хорошее качество получаемых в системе результатов. Учитывая то, что наибольшую трудность при создании нечетких экспертных систем представляет построение функций принадлежности, а от того, насколько адекватно построенная функция отражает знания эксперта или экспертов, во многом зависит качество принимаемых решений, был проведен тщательный анализ методов построения функций принадлежности (табл. 3).

Продолжение прил. 1

Таблица 3.
Сравнение способов построения функций принадлежности

Название метода	Способ построения	Число экспертов	Точность	Сложность алгоритма
Эталон	\forall	1	высокая	невысокая
Метод семантических дифференциалов Ч. Осгуда	прямой	1	невысокая	низкая
Способ вычисления частичной принадлежности друг другу строгих множеств	прямой	1	невысокая	высокая
Метод парных сравнений Т.Саати	косвенный	1	высокая	высокая
Модифицированный метод Т.Саати	косвенный	1	высокая	невысокая
Метод варьирования прототипов (Н. Скала)	косвенный	1	высокая	высокая
Групповые методы (А.П. Шер, З.А. Киквидзе, Я.Я.Осис, статистические методы)		несколько		

Так как для построения функций принадлежности мы используем знания и опыт одного эксперта, групповые методы нам не подходят. А из оставшихся методов всем нашим критериям удовлетворяет только модифицированный метод Саати. Далее был проведен анализ схем нечеткого вывода, для выбора такой, которая бы позволяла использовать ее в иерархических системах таким образом, чтобы не происходило размытие нечетких множеств при их передаче на следующий шаг приближенного рассуждения (табл. 4).

Таблица 4.
Сравнение схем нечеткого вывода

Название схемы нечеткого вывода	Вид выходного нечеткого множества	Отсутствие накопления нечеткости в иерархических системах	Отсутствие суммирования одинаковых правил при дефаззификации
Эталон	1-го порядка	+	+
Е.Н.Мамдани Р.М.Ларсен У. Tsukamoto	2-го порядка	-	+
М. Sugeno	1-го порядка	+	+
Синглетон	1-го порядка	+	-

Для реализации в автоматизированной системе был выбран нечеткий логический вывод по Сугено, так как только он удовлетворяет всем поставленным условиям.

Пример 2

ТЕМА ДИССЕРТАЦИИ: МЕТОД ПРОГНОЗИРОВАНИЯ ПЕРЕГРУЗОК В КОМПЬЮТЕРНЫХ СЕТЯХ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Актуальность темы. В последнее время появляется множество провайдеров, которые предлагают высокоскоростное подключение к сети Интернет. Это обуславливается тем, что данные услуги актуальны и востребованы. С каждым днем количество абонентов, подключенных к Интернету, растет, соответственно растет и нагрузка на серверы и коммутаторы в ядре сети провайдера. Современные компьютерные сети характеризуются сложной физической и логической топологией и зачастую построены неоптимальным образом. Приложения абонентов могут генерировать трафик, превышающий пропускную способность

Продолжение прил. 1

каналов связи. Все это приводит к возникновению перегрузок на участках сети, а следовательно, к нарушению целостности, возникновению угроз потери данных, ошибкам, отказам в обслуживании и замедленной работе в сети всех абонентов.

Вышеперечисленное является причиной нарушения безопасности передачи информации и снижения качества услуг в целом.

Разработка механизмов мониторинга и прогнозирования перегрузок является актуальной, важной задачей и неотъемлемым элементом комплекса мер по обеспечению безопасной эксплуатации компьютерной сети. Мониторинг сети позволяет отслеживать угрозы, определять время, место и причины возникновения неполадок. Прогнозирование интенсивности трафика дает возможность заранее предпринимать необходимые меры по предупреждению последствий.

Целью работы является решение научной проблемы по мониторингу и прогнозированию перегрузок в каналах связи современных вычислительных сетей и разработке программного комплекса для повышения защищенности и эффективности функционирования компьютерных сетей.

Для достижения поставленной цели необходимо решение следующих задач:

1. Анализ существующих методов и базовых средств мониторинга вычислительных сетей.
2. Разработка модели сегмента сети, которая позволяет определять время обработки пакета данных в любом элементе исследуемого участка.
3. Экспериментальная проверка выбранного подхода мониторинга. Сбор и анализ полученных измерений.

Продолжение прил. 1

4. Разработка программного продукта, позволяющего осуществлять краткосрочное прогнозирование трафика.

Объектом исследования являются компьютерные сети.

Предметом исследования является применение методов и моделей для описания поведения трафика в сегменте сети.

Методы исследования базируются на теории математической статистики (анализ временных рядов), теории вычислительных сетей и систем.

Научная новизна работы состоит в следующем:

1. Разработана модель сегмента сети для вычисления времени обработки пакета данных в любом элементе этой сети.

2. Разработана методика прогнозирования перегрузок в компьютерных сетях на основе анализа сетевой статистики.

Практическая ценность исследования состоит в разработке и программной реализации методов мониторинга и прогнозирования перегрузок в компьютерных сетях. Разработанное приложение может быть использовано в локальных сетях различных предприятий с целью повышения уровня безопасности передачи данных. Отдельные теоретические и практические положения могут быть включены в процессе преподавания дисциплин «Информационная безопасность».

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на научно-практических конференциях и семинарах Санкт-Петербургского государственного университета информационных технологий, механики и оптики и других научно-практических конференциях России.

Продолжение прил. 1

Публикации. По теме диссертационной работы опубликовано 4 статьи в научных журналах.

Во введении обосновывается актуальность темы. Сформулированы ее цель, научная новизна, приведены сведения о практическом использовании полученных научных результатов и представлены основные положения, выносимые на защиту.

В первой главе рассматриваются структуры и принципы работы современных компьютерных сетей. Формулируется понятие и описывается структура сетевой модели. Рассказывается о проблемах безопасности в вычислительных сетях и способах их решения.

Рассматриваются различия, особенности и принципы работы распространенных сетевых протоколов. Описываются их типы и функции. Рассматриваются основные программные и аппаратные средства, осуществляющие связь элементов сети между собой. Приведена многослойная модель всего комплекса средств, обеспечивающего взаимосвязь. Дано понятие трафика и определены его характеристики. Исследуются причины возникновения перегрузок в сети, следствием которых является снижение уровня безопасности передачи данных. Описаны основные алгоритмы борьбы с перегрузками, их недостатки и преимущества относительно друг друга.

Представлены основные цели и задачи мониторинга трафика:

Цели:

1. Выявление перегрузок;
2. Выявление аномалий.

Задачи:

1. Контроль загрузки каналов связи;

Продолжение прил. 1

2. Обнаружение перегрузок;
3. Обнаружение выхода из строя оборудования или программных сервисов;
4. Выявление активности пользователей, приложений, обнаружение атак;
5. Выявление узких мест сети, неоптимальности конфигурации сетевого оборудования;
6. Определение характеристик сетевого трафика;
7. Прогнозирование состояния сети.

Методы мониторинга разделены на два основных класса: пассивный и активный. При пассивном мониторинге сетевое оборудование или приложения периодически посылают данные о своем состоянии монитору. Монитор сохраняет данные для последующего использования, пример – протокол SNMP.

При активном мониторинге оборудование или программы отвечают на поступающие от администратора запросы, пример – протокол ICMP.

Преимущества пассивного подхода:

- 1) большая информативность в силу большего количества показателей;
- 2) в некоторых случаях потенциально меньшая нагрузка на сеть.

Недостатки пассивного подхода:

- 1) часто оборудование не поддерживает стандарты типа SNMP;
- 2) требуется административный доступ к оборудованию или программам для конфигурирования параметров мониторинга;
- 3) большие интервалы сэмплирования.

Преимущества активного подхода:

- 1) точное, направленное измерение, которое выполняется для различных объектов по требованию;
- 2) интервалы сэмплирования могут быть малы (от мсек);
- 3) возможность осуществлять мониторинг сетей с закрытым доступом к оборудованию.

Недостатки активного подхода:

- 1) потенциально большая нагрузка на сеть;
- 2) меньшая информативность, так как мониторинг выполняется по меньшему количеству показателей, чем в случае пассивного мониторинга;
- 3) сильно подвержен влиянию брандмауэров.

Выполнена классификация средств мониторинга и анализа компьютерных сетей. Описывается каждый из классов, рассказывается об их характеристиках и функциях. Рассматривается протокол управления сетями SNMP и его взаимодействие со специальной базой данных сетевых устройств, называемой MIB (Management Information Base). Описаны средства, использующие данный протокол. Затем описывается протокол управляющих сообщений ICMP и его характеристики. Показан формат ICMP-сообщения. Приводится таблица возможных типов ICMP-сообщения. Представлен обзор базовых программ мониторинга, использующих данный протокол. Рассказываем об их принципах работы, малоизвестных возможностях.

В процессе исследования было создано приложение PingTool, работающее по принципу утилиты ping и предназначенное для

Окончание прил. 1

определения времени задержки пакета. Преимущества средства перед другими утилитами заключаются в возможности одновременного «пингования» узла с помощью различных протоколов и периодического определения временных меток, соответствующих интерфейсу указанного узла. Так же приложение строит графики изменения значений задержек и сохраняет данные в удобном виде для дальнейшего анализа.

Речевые клише

Речевая функция	Лексические средства
Причина и следствие, условие и следствие	(и) поэтому, потому, так как
	поскольку
	отсюда (откуда) следует
	вследствие
	в результате
	в силу (ввиду) этого
	в зависимости от
	в связи с этим, согласно этому
	в таком (в этом) случае
	в этих (при таких же) условиях
	(а) если же..., то...
	Что свидетельствует (указывает, говорит, соответствует, дает возможность, позволяет, способствует, имеет значение)
Временная соотнесенность и порядок изложения	Сначала, прежде всего, в первую очередь
	Первым (последующим, предшествующим) шагом
	Одновременно, в то же время, здесь же
	Наряду с этим
	Предварительно, ранее, выше
	Ещё раз, вновь, снова
	Затем, далее, потом, ниже
	В дальнейшем, в последующем, впоследствии

Продолжение прил. 2

Речевая функция	Лексические средства
Сопоставление	Как..., так и ..., так же, как и
	Не только, но и ...
	По сравнению, если..., то...
	В отличие, в противоположность, наоборот,
	Аналогично, также, таким же образом
	С одной стороны, с другой стороны
	В то время, как, между тем, вместе с тем
	Тем не менее
Дополнение и уточнение	Также и, причем, при этом, вместе с тем
	Кроме (сверх, более) того
	Главным образом, особенно
Ссылка на предыдущее или последующее высказывание	Тем более, что...
	В том числе, в случае, то есть, а именно
	Как было сказано (показано, упомянуто, отмечено, установлено, получено, обнаружено, найдено)
	Как говорилось (указывалось, отмечалось, подчеркивалось) выше
	Согласно (сообразно, соответственно) этому
	В соответствии с этим, в связи с этим
	В связи с вышеизложенным
	Данный, названный, рассматриваемый
	Такой, такой же, подобный, аналогичный, сходный, подобного рода, подобного типа
	Следующий, последующий, некоторый
	Многие (одни, некоторые) из них
	Большая часть, большинство

Окончание прил. 2

Речевая функция	Лексические средства
Обобщение вывод	Таким образом, итак, следовательно
	В результате, в итоге, в конечном счете
	Отсюда (из этого) следует (вытекает)
	Это позволит сделать вывод (сводится к следующему, свидетельствует)
	Наконец, в заключение
Иллюстрация сказанного	Например, так, в качестве примера
	Примером может служить
	Такой пример (пример) в случае
	в (для) случае
	О чем можно судить, что очевидно
Введение новой информации	Рассмотрим следующие случаи
	Остановимся подробно на...
	Приведем несколько примеров
	Основные преимущества этого метода...
	Некоторые дополнительные замечания...
	Несколько слов о перспективах исследования

Учебное издание

**ОРГАНИЗАЦИЯ МАГИСТЕРСКИХ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ РАБОТ**

Составитель **АФАНАСЬЕВА** Татьяна Васильевна

Редактор Н. А. Евдокимова

Подписано в печать 29.05.2015. Формат 60×84/16.

Усл. печ. л. 2,32. Тираж 100 экз. Заказ 468. ЭИ № 558.

Ульяновский государственный технический университет
432027, Ульяновск, ул. Сев. Венец, 32.

ИПК «Венец» УлГТУ, 432027, Ульяновск, ул. Сев. Венец, д. 32

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Научно-исследовательская работа

Профиль подготовки
Искусственный интеллект и бизнес-аналитика

Квалификация выпускника
Магистр

Формы обучения
очная

г. Ульяновск, 2021

Выполнение практического задания

1. Процедура выполнения практического задания

Количество проводимых практических заданий в течение всего периода освоения практики	1 задание
Формат проведения результатов	Устно и/или Электронный
Методические рекомендации (при необходимости)	<p>1. Написание и презентация научной работы. Существенные навыки для студентов, магистрантов и аспирантов: учебное пособие / составитель. И. В. Семушин; - Ульяновск: УлГТУ, 2013. - 312 фреймов.: зв., видео URL: http://venec.ulstu.ru/lib/disk/2014/Semushin-root/ (дата обращения: 09.10.2021).</p> <p>2. Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с. URL: http://venec.ulstu.ru/lib/disk/2015/198.pdf</p>

2. Шкала оценивания с учетом срока сдачи

Критерии оценки качества решения задания	Балл
Студент четко и правильно обосновывает использование выбранных методов и инструментальных средств для решения практического задания. Решение позволяет выполнять поставленное задание в полном объеме.	Отлично
Студент недостаточно четко и полно обосновывает использование выбранных методов и инструментальных средств для решения практического задания. Либо решение позволяет выполнять поставленное задание не в полном (не менее $\frac{3}{4}$) объеме.	Хорошо
Студент не может четко и полно обосновать использование выбранных методов и инструментальных средств для решения практического задания. Либо решение позволяет выполнять поставленное задание не в полном (не менее $\frac{1}{2}$) объеме.	Удовлетворительно
Студент не может обосновать использование выбранных методов и инструментальных средств для решения практического задания. Либо решение позволяет выполнять поставленное задание не в полном (менее $\frac{1}{2}$) объеме.	Неудовлетворительно

3. Содержание практического задания

1. Подготовить обзор по направлению тематики магистерской диссертации с использованием электронных ресурсов: E-library, Google Scholar, Researchgate, Киберленинка и др.
2. Выполнить перевод зарубежных статей авторов, решающих аналогичную научно-исследовательскую задачу.
3. Формализовать поставленную научно-исследовательскую задачу, выявить объект, предмет и методы исследования, определить новизну и значимость исследования.
4. Провести патентное исследование по тематике магистерской диссертации.
5. Разработать концептуальную модель прототипа интеллектуальной аналитической системы согласно тематике магистерской диссертации.
6. Описать формальную постановку и решение научной проблемы на основе теоретико-множественной модели.

4. Перечень вопросов к практическим заданиям

1. Опишите понятие «четырёх-блочник» и его структуру.
2. Как конспектировать работы и составлять научный отчет?
3. Как проводятся патентные исследования?
4. Приведите электронные ресурсы (библиотеки, научные социальные сети), их возможности для поддержки подготовки магистерской диссертации.
5. Как презентовать научные результаты?
6. Перечислите и охарактеризуйте этапы научного исследования.
7. Как идентифицировать научную проблему?
8. Как конкретизировать проблему магистерской диссертации?
9. В чем состоит цель, объект и предмет вашей магистерской диссертации?
10. Какие методы и модели используются или планируется использовать в вашей магистерской диссертации?
11. Какие вычислительные эксперименты планируются выполнить в вашей магистерской диссертации?
12. Какие информационные технологии и средства планируется применить в вашей магистерской диссертации?

5. Перечень тем для практического задания

Тема для практического задания должна соответствовать тематике магистерской диссертации и согласовываться студентом с научным руководителем. Тема может быть предложена студентом или выбрана из предложенного научным руководителем списка. Тема диссертации должна соответствовать направлению подготовки 09.04.03 «Прикладная информатика» программа «Искусственный интеллект и бизнес-аналитика».

Научно-технический отчет

1. Процедура проведения

Формат проведения результатов	Электронный
Сроки	10 недель
Методические рекомендации (при необходимости)	1. Письменная и устная научная коммуникация: учебное пособие / И. В. Семушин. – Ульяновск : УлГТУ, 2014. – 143 с. URL: http://venec.ulstu.ru/lib/disk/2015/11.pdf (дата обращения: 09.10.2021). 2. Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с. URL: http://venec.ulstu.ru/lib/disk/2015/198.pdf

2. Шкала оценивания с учетом срока сдачи

Критерии оценки качества решения задачи	Балл
Студент полно и аргументировано оформил научно-технический отчет в соответствии с выполненным практическим заданием; четко и полно дает ответы на дополнительные уточняющие вопросы.	Отлично
Студент недостаточно четко и полно осветил разделы в научно-техническом отчете; ответил на дополнительные уточняющие вопросы с недочетами.	Хорошо
Студент допустил ошибки в научно-техническом отчете (при этом ошибки не должны иметь принципиального характера); неточности при ответе на уточняющие вопросы.	Удовлетворительно
Студент дал неверные, содержащие фактические ошибки ответы в научно-техническом отчете; не смог ответить на дополнительные и уточняющие вопросы или отказался отвечать.	Неудовлетворительно

3. Структура научно-технического отчета

Отчет по практике должен быть оформлен в виде научно-технического отчета в соответствии с требованиями ГОСТ 7.32-2017 «Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления».

Зачет с оценкой

1. Процедура проведения

Общее количество вопросов к зачету с оценкой	17 вопросов
Количество вопросов в билете	1 вопрос
Наличие задач в билете	нет
Формат проведения	Устно
Методические рекомендации (при необходимости)	-

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
выставляется обучающемуся, если студент показал глубокие знания теоретического материала по поставленному вопросу, грамотно логично и стройно его излагает.	Отлично
выставляется обучающемуся, если студент твердо знает теоретический материал, грамотно его излагает, не допускает существенных неточностей в ответе на вопрос.	Хорошо
выставляется обучающемуся, если студент показывает знания только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допускает отдельные неточности.	Удовлетворительно
выставляется обучающемуся, если студент допускает грубые ошибки в ответе на поставленный вопрос или отказался отвечать.	Неудовлетворительно

3. Вопросы к зачету с оценкой

1. Охарактеризуйте цель и задачи научно-исследовательской работы.
2. Приведите основные требования к магистерской диссертации.
3. Охарактеризуйте основные компоненты магистерской диссертации.
4. Что является областью и объектом исследования магистерской диссертации?
5. Приведите соотношение объекта и предмета исследования.
6. Перечислите и охарактеризуйте этапы научного исследования.
7. Охарактеризуйте и приведите примеры научного результата.
8. Опишите методику подготовки обзорной части магистерской диссертации.
9. Охарактеризуйте цель и методику проведения экспериментального исследования.
10. Охарактеризуйте цель и методику проведения теоретического исследования.
11. Приведите этапы вычислительного эксперимента.
12. Опишите возможности системного анализа для проведения научного исследования.
13. Приведите примеры методов моделирования, полезных в научном исследовании.
14. Охарактеризуйте понятие научно-исследовательской задачи.
15. Перечислите и охарактеризуйте этапы научно-исследовательской работы.
16. Перечислите и охарактеризуйте способы представления научно-исследовательской работы.
17. Охарактеризуйте этапы и результат патентных исследований.

Дневник практики

1. Процедура проведения

Формат проведения результатов	Письменный
Сроки	10 недель
Методические рекомендации (при необходимости)	-

2. Шкала оценивания с учетом срока сдачи

Критерии оценки качества оформления дневника	Балл
Студент полно и аргументировано оформил дневник в соответствии с выполненным практическим заданием.	Отлично
Студент недостаточно чётко и полно осветил задачи и график работы в дневнике.	Хорошо
Студент допустил ошибки в дневнике (при этом ошибки не должны иметь принципиального характера).	Удовлетворительно
Студент допустил неверные, содержащие фактические ошибки дневнике; отсутствуют подписи и/или печати с предприятия.	Неудовлетворительно

3. Содержание дневника практики

В дневнике обязательно указывается тип практики, группа, ФИО студента, приказ, сроки практики, место прохождения практики, руководитель от университета, руководитель от предприятия.

В ходе прохождения практики студентом заполняются листы:

1. Календарный график.
2. Индивидуальное задание.
3. Рабочие записи.

Дневник выдается студенту перед практикой на предварительном собрании.

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ОРГАНИЗАЦИЯ МАГИСТЕРСКИХ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ РАБОТ

Методические рекомендации

Составитель Т. В. Афанасьева

Ульяновск
УлГТУ
2015

УДК 004.925.8 (076)
ББК 32.973.26-018.2я7
О-64

Рецензент доцент кафедры ПМИ УлГТУ,
канд. техн. наук Кувайскова Ю. Е.

*Рекомендовано научно-методической комиссией
факультета информационных систем и технологий
в качестве методических рекомендаций*

О-64 Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с.

Методические рекомендации написаны в соответствии с требованиями государственного стандарта для магистрантов по направлению «Программная инженерия» и «Прикладная информатика».

В рекомендациях приведены общие сведения о целях, задачах научного исследования, описаны основные объекты, методы и результаты научно-исследовательских работ.

Работа подготовлена на кафедре «Информационные системы» УлГТУ.

**УДК 004.925.8 (076)
ББК 32.973.26-018.2я7**

© Афанасьева Т. В., составление, 2015
© Оформление. УлГТУ, 2015

ОГЛАВЛЕНИЕ

Общие положения	4
Этапы самостоятельной работы магистров по НИР	5
Цели и задачи НИР	5
Требования к магистерской диссертации	7
Требования к обзору.....	11
Методика подготовки обзора	14
Рекомендации по теоретическому исследованию	17
Основы вычислительного эксперимента	18
Заключение	19
Список литературы	19
Приложение 1. Примеры обзоров	20
Приложение 2. Речевые клише	35

ОБЩИЕ ПОЛОЖЕНИЯ

Цель методических рекомендаций – изложение процесса организации самостоятельной работы магистров в ходе их учебной и научно-исследовательской деятельности.

Самостоятельная работа магистра существенно затрудняется без помощи, квалифицированного управления ее ходом. Такую помощь оказывает научный руководитель (консультант). Однако ведущую роль в учении и проведении исследования должен играть сам магистр. Иными словами, контроль за ходом и результатами его деятельности должен постепенно переходить в самоконтроль.

Самостоятельная работа магистров – индивидуальная учебная или исследовательская деятельность, осуществляется в рамках дисциплин учебного плана, при выполнении курсовых работ и проектов, в при выполнении НИР, в период производственной практики и в процессе подготовки и написании магистерской диссертации, в ходе которой магистр активно воспринимает, осмысливает полученную информацию, решает теоретические и практические задачи.

Наиболее распространенные виды самостоятельной работы: работа с учебником, справочной литературой или с первоисточниками, наблюдения, лабораторные занятия, проведение эксперимента, моделирование и др. Исходя из внутреннего содержания, можно выделить три ступени самостоятельной деятельности: репродуктивная (подражание образцам, изложение своими словами содержания учебника, научных статей и др.); продуктивная (самостоятельное применение приобретенных знаний для решения задач, выходящих за пределы типовых); творческая (применение знаний в совершенно новых условиях: выдвижение гипотез, проведение экспериментов, моделирование ситуаций и процессов и др.).

ЭТАПЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ МАГИСТРОВ ПО НИР

Организация самостоятельной работы осуществляется в несколько этапов.

Первый этап – постановка и обсуждение с научным руководителем целей, задач, хода работы; *второй этап* – период самоорганизации магистранта и его непосредственная деятельность по решению задач; *третий этап* – подведение итогов и оценка полученных результатов.

Приступая к обучению, магистрант составляет индивидуальный план своей работы. Индивидуальный план учебной и научной работы магистранта составляется на два года при непосредственном участии научного руководителя и утверждается на заседании кафедры. В процессе составления индивидуального плана учебной и научной работы необходимо уяснить очередность и логическую последовательность намеченных работ. Последовательность заданий индивидуального плана со временем может измениться с тем условием, чтобы все они были выполнены.

ЦЕЛИ И ЗАДАЧИ НИР

Основной целью организации и развития системы научно-исследовательской работы является повышение уровня научной подготовки магистров и выявление талантливой молодежи для последующего обучения и пополнения педагогических и научных кадров вузов, других учреждений и организаций страны на основе новейших достижений научно-технического прогресса, экономической мысли и культурного развития.

Основными задачами организации и развития системы НИР для магистерских программ обучения являются:

1. Обеспечение интеграции учебных занятий и научно-исследовательской работы студентов;
2. Осуществление органичного единства обучения и подготовки студентов к творческому, научному и педагогическому труду;
3. Создание условий для раскрытия и реализации личностных творческих способностей студенческой молодежи;
4. Расширение массовости и повышение результативности участия магистров в научной деятельности;
5. Отбор талантливой молодежи, проявившей способности и стремление к научной и педагогической деятельности;
6. Формирование и развитие у студентов качеств научно-педагогических и научных работников;
7. Формирование и развитие у будущих выпускников: умения вести научно обоснованную профессиональную работу на предприятиях и в учреждениях любых организационно-правовых форм; способности быстрой адаптации, приложения полученных знаний и умений при изменяющихся требованиях к своей деятельности; освоения методологии и практики планирования, выбора оптимальных решений в условиях рыночных отношений; готовности и способности к повышению квалификации и переподготовке;
8. Подготовка руководителей высокой квалификации – магистров, имеющих навыки проектно-конструкторской работы, умеющих грамотно разработать и реализовать конкретные научно-практические мероприятия на производстве, обладающих навыками самоуправления;
9. Повышение массовости и эффективности НИР в университете путем привлечения студентов к исследованиям по наиболее приоритетным

направлениям науки, связанным с современными потребностями общества и государства;

10. Поиск и реализация источников финансирования, в том числе за счет средств, получаемых из внебюджетных источников и инновационной деятельности вузов, совершенствование форм и методов привлечения их к НИР;

11. Развитие научных межвузовских связей как внутри страны, так и со странами ближнего и дальнего зарубежья.

Важным является выбор тематики исследования и ориентация НИР на подготовку к написанию магистерской диссертации.

ТРЕБОВАНИЯ К МАГИСТЕРСКОЙ ДИССЕРТАЦИИ

При подготовке обзора по теме магистерской диссертации важно знать основные требования к магистерским диссертациям и основные понятия в содержательном плане.

Диссертация (от лат. — «dissertation» — рассуждение, исследование) — специальная форма научного произведения, имеющего квалификационный характер, подготовленная для публичной защиты и получения ученой степени.

Магистерская диссертация должна отличаться от бакалаврской работы глубокой теоретической проработкой проблемы, от дипломной работы специалиста – научной направленностью. Магистерская диссертация выполняется студентом по материалам, собранным им лично за период обучения и научно-исследовательской практики.

В магистерской диссертации необходимо выделить, определить и описать (см. примеры в приложении 1):

- 1) Тему исследования;

- 2) Область исследования;
- 3) Объект, предмет и прототип исследования (на основе обзора);
- 4) Цель исследования (на основе обзора);
- 5) Решаемые задачи (на основе обзора);
- 6) Научный результат;
- 7) Практический результат;
- 8) Подтверждение результатов.

Тема диссертации должна быть актуальной, представлять научный и практический интерес и соответствовать профилю магистерской программы. Обычно в теме указывается и проблема и метод ее решения.

Темы магистерских диссертаций, научные руководители и консультанты определяются выпускающими кафедрами (оформляются протоколом заседания кафедры), рассматриваются и утверждаются ученым советом (УС) факультета (оформляются протоколом заседания УС факультета) и приказом ректора после зачисления в магистратуру (по представлению отдела магистратуры на основании выписок из протокола заседания УС факультета).

Научными руководителями магистерских диссертаций назначаются ведущие преподаватели университета, имеющие ученую степень и ученое звание, ведущие научные исследования по тематике магистерских программ.

В обязанности научного руководителя входит:

1. Помощь в формировании темы диссертации и разработке плана работы над ней;
2. Систематическое консультирование магистранта;

3. Контроль за ходом работы над диссертацией, соответствия выполняемых работ плану;

4. Предоставление отзывов о результатах работы магистранта над диссертацией при промежуточной и итоговой аттестации;

5. Участие в обсуждении магистерских диссертаций на заседании кафедры.

Обычно сначала формируется обобщенная тема, которая к завершению обучения конкретизируется.

Областью исследования магистерской диссертации направления «Программная инженерия» является проектирование, разработка и исследование информационно-вычислительных систем, а магистерской диссертации направления «Прикладная информатика» – проектирование, разработка и исследование прикладных информационных процессов.

В каждой области в теме у магистранта есть своя специфика, которая определяет объект исследования.

Объект исследования – процесс или явление, порождающие проблемную ситуацию (это может быть недостаточно разработанная задача, задача, для которой нет алгоритмов и программного обеспечения (ПО)) и избранные для изучения с учетом соответствующего стандарта обучения.

Обычно объект исследования определяется в результате выполнения обзора как указание на целесообразность решения некоторой проблемы, связанной с недостаточным исследованием какого-то процесса, недостаточным уровнем автоматизации какого-то процесса, отсутствием в современных системах систем с такими функциональными возможностями и т. д.

Объектом исследования магистранта являются:

1. Алгоритмы решения задач;
2. Программный проект (проект разработки программного продукта);
3. Программный продукт (создаваемое программное обеспечение);
4. Процессы жизненного цикла программного продукта;
5. Методы и инструменты разработки программного продукта;
6. Методы управления персоналом.

Предмет исследования – все то, что находится в границах объекта исследования в определенном аспекте рассмотрения. Предмет исследования конкретизирует объект исследования путем определения средств для решения проблемы. Обзор средств, методов и технологий для обоснования предмета необходимо также привести в обзоре. Здесь формулируется прототип вашего исследования, то есть то, что уже есть, широко используется, но вы планируете этот прототип улучшить. Важно, чтобы прототип представлял апробированные решения.

Цель исследования. Обычно это одно предложение, интегрирующее тематику, объект и предмет исследования для достижения новых преимуществ.

Решаемые задачи. Для достижения цели необходимо решить ряд задач, таких как

1. Анализ объекта исследования, по сути, это выполнение обзора состояния в области определенной проблемы, методов их решения, технологий, алгоритмов и программных средств, используемых для решения проблемы.
2. Описание научного результата, который имеет отличия от прототипа в лучшую сторону.

3. Описание практического результата в виде конкретного программного обеспечения.

Научный результат. Это может быть алгоритм, методика (методика проектирования, тестирования, анализа, управления, оценивания), архитектура ПО и др. Важно обосновать, почему именно так предлагается решать обозначенную в цели проблему. И показать, чем ваше решение отличается от прототипа.

Практический результат. Здесь описывается, как вы создавали свое ПО, проектирование ПО, оценка ПО, архитектура, технологии, примеры использования. Если в рамках работы проводится реализация некоторого программного средства, то в разделе «Описание практической части» обязательно должна быть описана его программная реализация, в частности: приведены обоснования выбранного инструментария; приведена с иллюстрацией общая архитектура разработанного средства; приведена с иллюстрацией схема работы средства; если осуществляется доработка существующего средства, то должны быть описаны новые возможности/улучшения, реализованные в данной работе. обязательно должны быть приведены характеристики функционирования (например, сложность, производительность, время реакции и т. д.).

Требования к обзору

Результаты НИР могут быть оформлены в виде обзора, описания научных результатов и вычислительных экспериментов.

Обзор – научный документ, содержащий систематизированные научные данные по какой-либо теме, полученные в итоге анализа первоисточников. Знакомит с современным состоянием научной проблемы и перспективами ее развития.

Обзор литературы по теме должен показать умение магистранта систематизировать источники, критически их рассматривать, выделять существенное, определять изученность темы.

Структурно обзор включает следующие разделы:

1. Титульный лист (УлГТУ, каф.ИС, Обзор по теме магистерской диссертации направления «Программная инженерия» или «Прикладная информатика», научный руководитель, его ученая степень, ученое звание, «ФИО», выполнил «Магистрант ФИО», Ульяновск, год выполнения).
2. Оглавление.
3. Список сокращений.
4. Основная часть.
5. Список цитируемой литературы.

Магистрант должен особо подчеркнуть те вопросы, которые остались неразрешенными, определяя свое место в решении проблемы. Обзор литературы должен привести к выводу, что именно данная тема до сих пор не раскрыта (раскрыта частично, в другом аспекте) и нуждается в дальнейшей разработке.

Обзор включает в себя описание таких понятий, как область исследования, объект и предмет исследования.

По содержанию основная часть обзора в магистерской диссертации должна включать следующие разделы:

1. Актуальность выбранной темы в теоретическом и практическом плане. Важно обосновать целесообразность решения вашей проблемы, связанной с недостаточным исследованием какого-то процесса, недостаточным уровнем автоматизации какого-то процесса, отсутствием в

современных системах систем с такими функциональными возможностями и т. д.

Например, если принимается решение о создании некоторого программного средства, то необходимо показать, что не существует средства, обладающего нужными характеристиками.

2. Анализ существующего состояния проблемы на объектах исследования, выбор и обоснование математических и технологических инструментов решения; при этом анализ целесообразно проводить как по отдельным понятиям, отраженным в теме, так и по их сочетанию, с учетом специфики объектов направления «Программная инженерия»: модели и методы решения поставленной проблемы, технологии, методологии проектирования и разработки ПО. Обоснование может быть дано экспертно: приводятся высказывания, мнения авторитетных специалистов, с указанием ссылок на источники, где оно сформулировано. Или путем приведения результатов экспериментов.

3. Формулировка проблемы, степень ее разработанности в специальной научной литературе и методы ее разрешения на практике. Результаты можно представить в виде таблицы.

4. Постановка цели исследования. Обоснование и аргументация основных выводов и результатов исследования конкретной проблемы, приводящих к определению объекта, предмета, цели работы. Формулирование предложений и рекомендаций по разрешению изучаемой проблемы в виде комплекса задач, детализирующих цель исследования.

Методика подготовки обзора

Техника фиксации и обработки информации

Наиболее рациональными видами фиксирования информации большинством исследователей признаются цитаты, тезисы, конспекты, аннотации, рефераты и др.

Цитата – точная, буквальная выдержка из какого-либо текста с подробной ссылкой на источник (автор, заглавие источника, библиографические данные, цитируемые страницы). Выписки рекомендуется делать на одинакового формата карточках, лучше, плотной бумаги. На карточку, как правило, заносятся один или несколько фактов, идей, мыслей, касающихся определенного вопроса. Заполняется карточка на одной стороне. Другая может быть использована для соответствующих замечаний (комментариев, изложения другой точки зрения, противоположных фактов и др.). Карточки систематизируются и хранятся в электронном и в напечатанном виде. Для удобства пользования на карточках следует указать шифр, номер или название темы, раздела, проблемы и т. п.

Достоинства карточек видится в том, что они, во-первых, представляют собой отобранную и приведенную в систему наиболее ценную информацию; во-вторых, эта информация многократного и разнообразного применения: содержание карточки можно использовать для доклада, реферата, написания научной статьи и т. д.; в-третьих, карточками очень удобно пользоваться, так как они небольшие по размерам и не сброшюрованы.

Тезисы – кратко сформулированные основные положения, идеи доклада, научной работы, лекции.

Конспект – письменное изложение (может быть своими словами) содержания научной работы, лекции, доклада и др.

Аннотация – краткое разъяснительное или критическое изложение содержания, краткая характеристика и объявление назначения книги, статьи, рукописи.

Подготовка обзора требует большого объема времени как для поиска, так и для анализа и оформления результата. Этот процесс включает следующие этапы:

1. Формулировка того, что надо искать: программы, реализующие вашу тематику, модели, методы и алгоритмы. Выдвинуть предположения, что есть объект исследования, какие задачи решаются для объекта. Это итерационный процесс, что надо искать, уточняется после предварительного этапа.

2. Предварительный этап.

- a. Поиск источников по теме и решаемым задачам. Обычно через Интернет (15-20 источников). Как только вы находите релевантный материал, сразу же копируете ссылку, затем нужный материал (текстовый, графический).

- b. Затем распечатываете и прочитываете. Если недостаточно материала, то переход на первый этап или по ссылкам от найденных источников, иначе переход на третий этап.

- c. Определение, какие задачи решаются для объекта исследования. Выдвижение предположения, какие задачи в настоящее время недостаточно решены для объекта исследования. Оформление в виде таблиц (задачи-способы решения, (задачи-ПО). Формулировка предварительной цели.

3. Перевод на иностранный язык темы, объекта, задач (примерно 10 источников).

а. Поиск в «нерусскоязычном» Интернете, аналогично п. 2. Можно использовать ресурсы computer.org, researchGate, eLibrary и др.

б. Перевод найденных материалов.

4. Дополнение обзора печатными материалами (15 - 20 источников). Обычно в виде книг, монографий, статей, диссертаций (Необходимо, чтобы большинство источников было издано не позднее 2009 года). В интернет-магазинах обычно приводятся для литературы небольшие рефераты, аннотации или оглавления, по которым можно составить мнение, о чем этот источник. Можно посмотреть в библиотеке через электронный каталог, через другие электронные библиотеки. Рекомендуемые журналы можно найти в eLibrary.ru.

5. Оформление списка используемых источников по алфавиту, начиная с иностранных авторов (ссылки должны быть оформлены в соответствии с ГОСТ Р 7.1-2003. Источники должны содержать как электронные ресурсы, так и печатную литературу в виде книг, монографий, статей, диссертаций.

6. Оформление текста в соответствии с общепринятыми требованиями к научной документации.

7. Распечатка обзора, его передача руководителю обсуждения.

Рекомендации по теоретическому исследованию

Теоретический анализ и синтез

Анализ заключается в разложении изучаемого явления на составляющие элементы, позволяет вскрыть внутреннюю структуру объекта.

Синтез позволяет воссоздать предмет как систему связей и взаимодействий с выделением наиболее существенных из них. Особенность метода теоретического анализа и синтеза в исследованиях, в любом варианте применения, заключается в его универсальных возможностях рассматривать явления и процессы действительности в их самых сложных сочетаниях, выделять наиболее существенные признаки и свойства, связи и отношения, устанавливать закономерности их развития.

Путем анализа и синтеза можно «увидеть» объективное содержание, объективные тенденции, несоответствия, предвидеть такие формы и стадии процесса, которые запроектированы, но пока еще реально не существуют.

Абстрагирование – процесс мысленного отвлечения какого-либо свойства или признака предмета от самого предмета, от других его свойств (например, исследователь аналитическим путем выделяет из всего многообразия признаков процесса программирования один – доступность программного продукта – и рассматривает его самостоятельно).

Моделирование – воспроизведение некоторых характеристик на другом объекте, созданном специально с целью изучения. Этот второй объект называется моделью (идеальной или материальной). Открывается возможность переноса информации, полученной при изучении модели на прототип. Необходимо помнить, что любая модель беднее прототипа, что она отражает лишь его отдельные стороны и связи. Исследователь, изучив

характерные черты реальных процессов и их тенденций, делает их мысленную компоновку, то есть моделирует состояние изучаемой системы. Можно выделить множество методов исследования, опирающихся на теоретические основы математического моделирования:

1. Вычислительной математики (в том числе, численные методы),
2. Дискретной математики,
3. Математической логики,
4. Теории вероятности и математической статистики и др.

Основы вычислительного эксперимента

Применение вычислительных методов решения задачи – современный подход к экспериментированию в рамках НИР.

Можно выделить следующие этапы проведения вычислительного эксперимента:

1. Построение математической модели исследуемого объекта (сюда же относится и анализ модели, выяснение корректности поставленной математической задачи);
2. Построение вычислительного алгоритма – метода приближенного решения поставленной задачи и его обоснование;
3. Программирование алгоритма на ЭВМ и его тестирование;
4. Проведение серии расчетов с варьированием определяющих параметров исходной задачи и алгоритма;
5. Анализ полученных результатов и погрешностей.

ЗАКЛЮЧЕНИЕ

Рассмотрены вопросы организации самостоятельной работы магистрантов при проведении НИР. Указанные вопросы актуальны для каждого магистранта и поддержаны большим количеством часов. Научные исследования необходимы для подготовки и написания магистерской диссертации практически в каждом семестре обучения по программам магистратуры. НИР относится к дисциплинам вариативной части программы магистратуры, которые определяют направленность (профиль) программы.

Список литературы

1. ГОСТ Р 7.32-2001 СИБИД. Отчет о научно-исследовательской работе. Структура и правила оформления. – М., 2002.

2. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.04.04 Программная инженерия (уровень магистратуры).

3. ГОСТ 15.101-98 Порядок выполнения научно-исследовательских работ. – М., 1999.

5. Правила цитирования. Доступно по адресу <http://www.ulstu.ru/main/view/article/16403> (дата обращения: 12.12.2014).

6. Методические рекомендации по проектированию оценочных средств для реализации многоуровневых образовательных программ ВПО при компетентностном подходе / В.А. Богословский, Е.В. Караваева, Е.Н. Ковтун и др. – М. : Изд-во МГУ, 2007. – 148 с.

7. Научно-исследовательская деятельность в вузе: (основные понятия, этапы, требования) / Н. В. Шестак, Е. В. Чмыхова. – М.: Изд-во МГУ, 2007. – 179 с.

8. ГОСТ Р 7.0.11-2011 СИБИД. Диссертация и автореферат диссертации. Структура и правила оформления. – М., 2011.

Примеры обзоров и представление об основных понятиях диссертации

Пример 1

ТЕМА ДИССЕРТАЦИИ: ИССЛЕДОВАНИЕ И РАЗРАБОТКА НЕЧЕТКОЙ МОДЕЛИ И КОМПЛЕКСА ПРОГРАММ ЭКОЛОГИЧЕСКОЙ ЭКСПЕРТИЗЫ ГОРНОДОБЫВАЮЩЕГО ПРОИЗВОДСТВА

В системах управления деятельностью сложных явлений и процессов, таких как производственная деятельность промышленного предприятия, процессы принятия решений часто протекают в условиях неопределенности и основываются на экспертизе, которая осуществляется в рамках экспертной деятельности, и ее результатом выступают экспертные оценки. Многие реальные процессы имеют иерархическую структуру. Изучение таких структур и применение полученных результатов для анализа реальных объектов и процессов отражено в работах Саати [1] и Такахаро [2]. Методы теории нечетких множеств являются удобным средством моделирования, анализа и синтеза человеко-компьютерных систем. Поэтому изучение нечетких иерархических систем представляется актуальной задачей, имеющей не только теоретической, но и практический интерес.

Для теории нечетких множеств основополагающим понятием является понятие нечеткого множества, которое характеризуется функцией принадлежности. Посредством нечеткого множества можно строго описывать присущие языку человека расплывчатые элементы, без

Продолжение прил. 1

формализации которых нет надежды существенно продвинуться вперед в моделировании интеллектуальных процессов.

Сегодня во всем мире принято уделять большое внимание проблеме экологии. Деятельность человека по освоению природной среды породила не только новые возможности роста благосостояния человечества, но и привела к глубокому кризису состояния окружающей среды. Поскольку ни один руководитель не будет работать себе в ущерб, важно соблюсти баланс между экологической безопасностью, требующей денежных вливаний, и экономической эффективностью деятельности хозяйствующего субъекта. Вот почему в качестве прикладной области исследования был выбран анализ экологической безопасности горнопромышленного производства на основе методики оценки экологических платежей предприятия.

Актуальность проблемы

В последние годы определился и практически осуществился переход от автоматизированных систем анализа состояния сложных технических систем к экспертным системам, работающим на основе логического приближенного вывода. Отмеченная тенденция связана с резким усложнением современных технических систем, и, в частности, экологической безопасности горнодобывающих производств. В связи с этим исследование, назначением которого является разработка математической модели и комплекса программ экологической безопасности горнодобывающего производства на основе иерархического нечеткого вывода, является, несомненно, актуальной.

Цель диссертационной работы

Целью диссертационной работы является разработка математической модели экологической безопасности горнодобывающего производства и комплекса программ экспертизы факторов такого производства для нечетко заданных исходных данных.

Задачи исследования

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать факторы горнодобывающего производства, значимые для формирования математической модели экологической безопасности, выполнить сравнительный анализ существующих методов и программ экспертной оценки экологической безопасности;

2. Разработать модель экологической безопасности горнодобывающего производства, позволяющую проводить экспертную оценку при нечетко заданных исходных данных;

3. Разработать механизм иерархического нечеткого логического вывода экспертных систем;

4. Разработать комплекс программ, реализующий экспертную систему экологической безопасности горнодобывающего производства.

Методы исследования.

Имитационное моделирование, методы математической статистики, нечеткая логика, объектно-ориентированный подход при создании комплекса программ.

Продолжение прил. 1

Научная новизна положений, выносимых на защиту:

1. Создан механизм и алгоритмы иерархического нечеткого вывода, позволяющий осуществлять экспертизу объектов с нечетко заданными исходными данными.
2. Разработана объектно-ориентированная архитектура комплекса программ экспертной системы экологической безопасности.

Практическая значимость работы

Практическая значимость работы заключается в том, что разработанная модель экологической безопасности горнодобывающего производства на основе иерархического нечеткого вывода внедрена в практику работы открытого акционерного общества «Кварц» (пос. Силикатный, Сенгилеевский район, Ульяновская обл.) и эксперта, члена НП «Горнопромышленники России», Танеева Ф.Г, в форме комплекса программ.

Реализация результатов работы

Результаты работы реализованы в виде комплекса программ JFuzzyTool1.0, зарегистрированного Федеральной службой по интеллектуальной собственности, патентам и товарным знакам (№20076144448 от 23.10.2007 г.).

Апробация работы

Основные положения и результаты диссертации докладывались, обсуждались и получили одобрение на 10 международных конференциях. Неоднократно докладывались на научно-технических конференциях УлГТУ «Вузовская наука в современных условиях».

Публикации

По теме диссертации опубликованы 5 работ.

Продолжение прил. 1

Во введении рассмотрена актуальность проблемы создания экспертных систем анализа экологической безопасности горнодобывающего производства, сформулированы объекты исследования и предмет исследования, изложены цель и соответствующие ей задачи работы.

Первая глава «Обзор современного состояния экспертных систем анализа экологической безопасности» посвящена анализу существующих моделей, используемых методов вычислительного интеллекта, инструментов для создания экспертных систем. Изложено текущее состояние различных исследований, ведущихся в этом направлении. Даны основные понятия в области экологической безопасности промышленного предприятия, рассмотрены элементы структуры экспертной системы.

Таблица 1

Сравнение методов вычислительного интеллекта

Метод вычислительного интеллекта	Обязательное наличие обучающих выборок	Обязательное наличие эксперта	Способ построения функций принадлежности и правил нечеткого вывода
Нечеткие нейронные сети: J.J.Buckley, J.S. Roger Jang	+	-	Автоматический. С помощью соответствующих алгоритмов обучения сетей
Иерархический нечеткий вывод: V.Totta, M.Brown, H.Rainer, C.Wei, L.X.Wang, L.C.Lin, X.-J.Zeng, G.-Y.Lee	-	+	Вручную. С помощью методов построения функций принадлежности. Правила формируются на основании утверждений эксперта в виде слов естественного языка

Несмотря на явные преимущества использования гибридных технологий (нечетких нейронных сетей, см. табл. 1), для решения рассматриваемой задачи они не подходят ввиду отсутствия обучающих

Продолжение прил. 1

выборок. Сделаны выводы о том, что использование пакетов различных классов позволяет строить адекватные модели современных систем. Языки моделирования общего назначения обладают наиболее широкими функциональными возможностями для любых предметных областей, могут описывать любые свойства компонентов и программировать их взаимодействия. Визуальные системы моделирования более наглядны и современны, но обладают меньшей функциональностью. Специализированные системы в этой области на сегодняшний день учитывают только специфику предприятий нефтедобывающей отрасли.

Таблица 2.
Сравнение оболочек для проектирования экспертных систем

Название оболочки	Кросс-платформенность	Сложность адаптации системы	Возможность создания иерархий	Механизмы решения задачи
Эталон	+	Пользователь. Без ограничений	+	Механизмы логического вывода
FuzzyCLIPS	+	Программист. Без ограничений	+	Нечеткий логический вывод
OPS5	–	Программист. Не предназначена для решения конкретных задач	–	Цикл распознавания пишет пользователь
BABYLON	+	Программист. Только для задач диагностики	–	Фреймы
WindExS	–	Пользователь. Без ограничений	+	Любые механизмы логического вывода, в т.ч. нечеткий

Из табл. 2 становится ясно, что наилучшим образом под выбранный эталон подходят среды FuzzyCLIPS и WindExS. Недостатком первой является то, что для адаптации экспертной системы, созданной в этой оболочке, необходимо прибегать к помощи программиста. А вторая не

Продолжение прил. 1

переносима на различные платформы. Мы решаем конкретную задачу построения нечеткой экспертной системы анализа экологической безопасности горнодобывающего предприятия. Нам необходимо, чтобы система могла быстро адаптироваться к изменениям, происходящим при принятии нового Федерального закона, изменениям ГОСТа, отраслевого положения или нового порядка расчета и начисления экологических платежей, и чтобы внести необходимые изменения мог конечный пользователь нашей системы. Так как прикладная область имеет иерархическую структуру, то экспертная система тоже должна быть иерархической.

В результате проведенного исследования существующих средств проектирования экспертных систем было принято решение о создании собственного комплекса программ, сочетающего в себе лучшие черты существующих на сегодняшний день оболочек: кросс-платформенность, возможность создания иерархий, простоту адаптации системы, быстрое действие и хорошее качество получаемых в системе результатов. Учитывая то, что наибольшую трудность при создании нечетких экспертных систем представляет построение функций принадлежности, а от того, насколько адекватно построенная функция отражает знания эксперта или экспертов, во многом зависит качество принимаемых решений, был проведен тщательный анализ методов построения функций принадлежности (табл. 3).

Продолжение прил. 1

Таблица 3.
Сравнение способов построения функций принадлежности

Название метода	Способ построения	Число экспертов	Точность	Сложность алгоритма
Эталон	\forall	1	высокая	невысокая
Метод семантических дифференциалов Ч. Осгуда	прямой	1	невысокая	низкая
Способ вычисления частичной принадлежности друг другу строгих множеств	прямой	1	невысокая	высокая
Метод парных сравнений Т.Саати	косвенный	1	высокая	высокая
Модифицированный метод Т.Саати	косвенный	1	высокая	невысокая
Метод варьирования прототипов (Н. Скала)	косвенный	1	высокая	высокая
Групповые методы (А.П. Шер, З.А. Киквидзе, Я.Я.Осис, статистические методы)		несколько		

Так как для построения функций принадлежности мы используем знания и опыт одного эксперта, групповые методы нам не подходят. А из оставшихся методов всем нашим критериям удовлетворяет только модифицированный метод Саати. Далее был проведен анализ схем нечеткого вывода, для выбора такой, которая бы позволяла использовать ее в иерархических системах таким образом, чтобы не происходило размытие нечетких множеств при их передаче на следующий шаг приближенного рассуждения (табл. 4).

Таблица 4.
Сравнение схем нечеткого вывода

Название схемы нечеткого вывода	Вид выходного нечеткого множества	Отсутствие накопления нечеткости в иерархических системах	Отсутствие суммирования одинаковых правил при дефаззификации
Эталон	1-го порядка	+	+
Е.Н.Мамдани Р.М.Ларсен У. Tsukamoto	2-го порядка	-	+
М. Sugeno	1-го порядка	+	+
Синглетон	1-го порядка	+	-

Для реализации в автоматизированной системе был выбран нечеткий логический вывод по Сугено, так как только он удовлетворяет всем поставленным условиям.

Пример 2

ТЕМА ДИССЕРТАЦИИ: МЕТОД ПРОГНОЗИРОВАНИЯ ПЕРЕГРУЗОК В КОМПЬЮТЕРНЫХ СЕТЯХ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Актуальность темы. В последнее время появляется множество провайдеров, которые предлагают высокоскоростное подключение к сети Интернет. Это обуславливается тем, что данные услуги актуальны и востребованы. С каждым днем количество абонентов, подключенных к Интернету, растет, соответственно растет и нагрузка на серверы и коммутаторы в ядре сети провайдера. Современные компьютерные сети характеризуются сложной физической и логической топологией и зачастую построены неоптимальным образом. Приложения абонентов могут генерировать трафик, превышающий пропускную способность

Продолжение прил. 1

каналов связи. Все это приводит к возникновению перегрузок на участках сети, а следовательно, к нарушению целостности, возникновению угроз потери данных, ошибкам, отказам в обслуживании и замедленной работе в сети всех абонентов.

Вышеперечисленное является причиной нарушения безопасности передачи информации и снижения качества услуг в целом.

Разработка механизмов мониторинга и прогнозирования перегрузок является актуальной, важной задачей и неотъемлемым элементом комплекса мер по обеспечению безопасной эксплуатации компьютерной сети. Мониторинг сети позволяет отслеживать угрозы, определять время, место и причины возникновения неполадок. Прогнозирование интенсивности трафика дает возможность заранее предпринимать необходимые меры по предупреждению последствий.

Целью работы является решение научной проблемы по мониторингу и прогнозированию перегрузок в каналах связи современных вычислительных сетей и разработке программного комплекса для повышения защищенности и эффективности функционирования компьютерных сетей.

Для достижения поставленной цели необходимо решение следующих задач:

1. Анализ существующих методов и базовых средств мониторинга вычислительных сетей.
2. Разработка модели сегмента сети, которая позволяет определять время обработки пакета данных в любом элементе исследуемого участка.
3. Экспериментальная проверка выбранного подхода мониторинга. Сбор и анализ полученных измерений.

Продолжение прил. 1

4. Разработка программного продукта, позволяющего осуществлять краткосрочное прогнозирование трафика.

Объектом исследования являются компьютерные сети.

Предметом исследования является применение методов и моделей для описания поведения трафика в сегменте сети.

Методы исследования базируются на теории математической статистики (анализ временных рядов), теории вычислительных сетей и систем.

Научная новизна работы состоит в следующем:

1. Разработана модель сегмента сети для вычисления времени обработки пакета данных в любом элементе этой сети.

2. Разработана методика прогнозирования перегрузок в компьютерных сетях на основе анализа сетевой статистики.

Практическая ценность исследования состоит в разработке и программной реализации методов мониторинга и прогнозирования перегрузок в компьютерных сетях. Разработанное приложение может быть использовано в локальных сетях различных предприятий с целью повышения уровня безопасности передачи данных. Отдельные теоретические и практические положения могут быть включены в процессе преподавания дисциплин «Информационная безопасность».

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на научно-практических конференциях и семинарах Санкт-Петербургского государственного университета информационных технологий, механики и оптики и других научно-практических конференциях России.

Продолжение прил. 1

Публикации. По теме диссертационной работы опубликовано 4 статьи в научных журналах.

Во введении обосновывается актуальность темы. Сформулированы ее цель, научная новизна, приведены сведения о практическом использовании полученных научных результатов и представлены основные положения, выносимые на защиту.

В первой главе рассматриваются структуры и принципы работы современных компьютерных сетей. Формулируется понятие и описывается структура сетевой модели. Рассказывается о проблемах безопасности в вычислительных сетях и способах их решения.

Рассматриваются различия, особенности и принципы работы распространенных сетевых протоколов. Описываются их типы и функции. Рассматриваются основные программные и аппаратные средства, осуществляющие связь элементов сети между собой. Приведена многослойная модель всего комплекса средств, обеспечивающего взаимосвязь. Дано понятие трафика и определены его характеристики. Исследуются причины возникновения перегрузок в сети, следствием которых является снижение уровня безопасности передачи данных. Описаны основные алгоритмы борьбы с перегрузками, их недостатки и преимущества относительно друг друга.

Представлены основные цели и задачи мониторинга трафика:

Цели:

1. Выявление перегрузок;
2. Выявление аномалий.

Задачи:

1. Контроль загрузки каналов связи;

Продолжение прил. 1

2. Обнаружение перегрузок;
3. Обнаружение выхода из строя оборудования или программных сервисов;
4. Выявление активности пользователей, приложений, обнаружение атак;
5. Выявление узких мест сети, неоптимальности конфигурации сетевого оборудования;
6. Определение характеристик сетевого трафика;
7. Прогнозирование состояния сети.

Методы мониторинга разделены на два основных класса: пассивный и активный. При пассивном мониторинге сетевое оборудование или приложения периодически посылают данные о своем состоянии монитору. Монитор сохраняет данные для последующего использования, пример – протокол SNMP.

При активном мониторинге оборудование или программы отвечают на поступающие от администратора запросы, пример – протокол ICMP.

Преимущества пассивного подхода:

- 1) большая информативность в силу большего количества показателей;
- 2) в некоторых случаях потенциально меньшая нагрузка на сеть.

Недостатки пассивного подхода:

- 1) часто оборудование не поддерживает стандарты типа SNMP;
- 2) требуется административный доступ к оборудованию или программам для конфигурирования параметров мониторинга;
- 3) большие интервалы сэмплирования.

Преимущества активного подхода:

- 1) точное, направленное измерение, которое выполняется для различных объектов по требованию;
- 2) интервалы сэмплирования могут быть малы (от мсек);
- 3) возможность осуществлять мониторинг сетей с закрытым доступом к оборудованию.

Недостатки активного подхода:

- 1) потенциально большая нагрузка на сеть;
- 2) меньшая информативность, так как мониторинг выполняется по меньшему количеству показателей, чем в случае пассивного мониторинга;
- 3) сильно подвержен влиянию брандмауэров.

Выполнена классификация средств мониторинга и анализа компьютерных сетей. Описывается каждый из классов, рассказывается об их характеристиках и функциях. Рассматривается протокол управления сетями SNMP и его взаимодействие со специальной базой данных сетевых устройств, называемой MIB (Management Information Base). Описаны средства, использующие данный протокол. Затем описывается протокол управляющих сообщений ICMP и его характеристики. Показан формат ICMP-сообщения. Приводится таблица возможных типов ICMP-сообщения. Представлен обзор базовых программ мониторинга, использующих данный протокол. Рассказывается об их принципах работы, малоизвестных возможностях.

В процессе исследования было создано приложение PingTool, работающее по принципу утилиты ping и предназначенное для

Окончание прил. 1

определения времени задержки пакета. Преимущества средства перед другими утилитами заключаются в возможности одновременного «пингования» узла с помощью различных протоколов и периодического определения временных меток, соответствующих интерфейсу указанного узла. Так же приложение строит графики изменения значений задержек и сохраняет данные в удобном виде для дальнейшего анализа.

Речевые клише

Речевая функция	Лексические средства
Причина и следствие, условие и следствие	(и) поэтому, потому, так как
	поскольку
	отсюда (откуда) следует
	вследствие
	в результате
	в силу (ввиду) этого
	в зависимости от
	в связи с этим, согласно этому
	в таком (в этом) случае
	в этих (при таких же) условиях
	(а) если же..., то...
	Что свидетельствует (указывает, говорит, соответствует, дает возможность, позволяет, способствует, имеет значение)
Временная соотнесенность и порядок изложения	Сначала, прежде всего, в первую очередь
	Первым (последующим, предшествующим) шагом
	Одновременно, в то же время, здесь же
	Наряду с этим
	Предварительно, ранее, выше
	Ещё раз, вновь, снова
	Затем, далее, потом, ниже
	В дальнейшем, в последующем, впоследствии

Продолжение прил. 2

Речевая функция	Лексические средства
Сопоставление	Как..., так и ..., так же, как и
	Не только, но и ...
	По сравнению, если..., то...
	В отличие, в противоположность, наоборот,
	Аналогично, также, таким же образом
	С одной стороны, с другой стороны
	В то время, как, между тем, вместе с тем
	Тем не менее
Дополнение и уточнение	Также и, причем, при этом, вместе с тем
	Кроме (сверх, более) того
	Главным образом, особенно
Ссылка на предыдущее или последующее высказывание	Тем более, что...
	В том числе, в случае, то есть, а именно
	Как было сказано (показано, упомянуто, отмечено, установлено, получено, обнаружено, найдено)
	Как говорилось (указывалось, отмечалось, подчеркивалось) выше
	Согласно (сообразно, соответственно) этому
	В соответствии с этим, в связи с этим
	В связи с вышеизложенным
	Данный, названный, рассматриваемый
	Такой, такой же, подобный, аналогичный, сходный, подобного рода, подобного типа
	Следующий, последующий, некоторый
	Многие (одни, некоторые) из них
	Большая часть, большинство

Окончание прил. 2

Речевая функция	Лексические средства
Обобщение вывод	Таким образом, итак, следовательно
	В результате, в итоге, в конечном счете
	Отсюда (из этого) следует (вытекает)
	Это позволит сделать вывод (сводится к следующему, свидетельствует)
	Наконец, в заключение
Иллюстрация сказанного	Например, так, в качестве примера
	Примером может служить
	Такой пример (пример) в случае
	в (для) случае
	О чем можно судить, что очевидно
Введение новой информации	Рассмотрим следующие случаи
	Остановимся подробно на...
	Приведем несколько примеров
	Основные преимущества этого метода...
	Некоторые дополнительные замечания...
	Несколько слов о перспективах исследования

Учебное издание

**ОРГАНИЗАЦИЯ МАГИСТЕРСКИХ
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ РАБОТ**

Составитель **АФАНАСЬЕВА** Татьяна Васильевна

Редактор Н. А. Евдокимова

Подписано в печать 29.05.2015. Формат 60×84/16.

Усл. печ. л. 2,32. Тираж 100 экз. Заказ 468. ЭИ № 558.

Ульяновский государственный технический университет
432027, Ульяновск, ул. Сев. Венец, 32.

ИПК «Венец» УлГТУ, 432027, Ульяновск, ул. Сев. Венец, д. 32

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Производственная (преддипломная) практика

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Составители: Е.В. Суркова,
Н.В. Корунова

Ульяновск
УлГТУ
2021

УДК 004.891.2 (076)

ББК 32.813.5 я73

П63

Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве методических указаний

П63 Производственная (преддипломная) практика: методические указания / сост.:Е. В. Суркова, Н. В. Корунова. – Ульяновск : УлГТУ, 2021. – 29 с.

Методические указания по прохождению преддипломной практики рекомендованы обучающимся для прохождения, сдачи и защиты преддипломной практики. Предоставлены требования к оформлению отчета по практике. Рекомендации и требования к самостоятельной работе обучающихся в ходе практики разработаны в соответствии с рабочей программой практики. Предназначены для обучающихся, обучающихся по направлениям 09.04.03 Прикладная информатика (магистерская Искусственный интеллект и бизнес-аналитика) и 09.04.04 Программная инженерия (магистерская программа Искусственный интеллект и предиктивная аналитика).

Работа подготовлена на кафедре «Информационные системы».

УДК 004.891.2 (076)

ББК 32.813.5 я73

© Суркова Е. В., Корунова Н.В.,
2021

© Оформление. УлГТУ, 2021

Содержание	
Введение	4
1.Общие положения о производственной (преддипломной) практике.....	4
2.Организация производственной (преддипломной) практики	5
2.1. Объем и продолжительность практики	7
2.2. Планируемые результаты обучения при прохождении практики, соотнесенных с планируемыми результатами освоения образовательной программы.....	7
2.3. Содержание практики.....	18
2.4. Формирование отчетных документов по практике	19
2.5. Защита отчета по практике, промежуточная аттестация.....	22
3.Перечень учебной литературы и ресурсов сети «Интернет», необходимых для проведения практики.....	23
Приложение А	27
Приложение Б.....	29

ВВЕДЕНИЕ

Методические указания по прохождению (преддипломной) практики разработаны в соответствии с ОПОП ВО с учетом требований российского законодательства в области высшего образования, Устава Ульяновского государственного технического университета (далее – УлГТУ, Университет) и Положения о порядке проведения практики обучающихся Ульяновского государственного технического университета.

1. ОБЩИЕ ПОЛОЖЕНИЯ О ПРОИЗВОДСТВЕННОЙ (ПРЕДДИПЛОМНОЙ) ПРАКТИКЕ

Преддипломная практика, как обязательная часть основной образовательной программы, является завершающим этапом обучения и проводится после освоения магистрантами программ теоретического и практического обучения. Прохождение практики в организациях, на предприятиях, в Университете связано с профессиональной ориентацией магистрантов, способствует формированию представления о применении методов и средств решения задач исследования с использованием технологий и методов искусственного интеллекта; подготовки данных для составления отчетов, а также сбора и анализа практического материала для выполнения выпускной квалификационной работы (магистерской диссертации).

Целями производственной (преддипломной) практики являются:

- формирование профессиональных компетенций в сфере проектирования, разработки и использования систем искусственного интеллекта в различных предметных областях;
- получение теоретических и практических результатов, являющихся достаточными для успешного выполнения и защиты магистерской диссертации.

Данная цель определяет следующие задачи преддипломной практики:

- приобретение обучающимися навыков разработки алгоритмов и программных средств для решения задач в области создания и применения систем искусственного интеллекта, практического решения профессиональных задач на конкретных рабочих местах в качестве исполнителей или стажёров;
- приобретение опыта исследования и разработки архитектуры систем искусственного интеллекта для различных предметных областей;
- изучение опыта разработки и проведения экспериментальной проверки работоспособности программных компонентов систем

- искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования в условиях конкретных производств, организаций или фирм;
- изучение опыта руководства созданием комплексных систем искусственного интеллекта с применением новых методов и алгоритмов машинного обучения в условиях конкретных производств;
 - приобретение обучающимся навыков информационно-аналитической работы, включая сбор необходимого материала, подготовку и проведение исследований процессов принятия управленческих решений с целью реинжиниринга бизнес-процессов путем внедрения систем искусственного интеллекта,
 - описание и анализ бизнес-процессов предприятия / подразделения / группы / исполнителя, построение и исследование их с использованием методов научных исследований и математического моделирования в области проектирования и управления системами искусственного интеллекта;
 - выработка специфических навыков и компетенций, связанных с необходимостью формирования профессиональных умений и опыта в процессе научно-исследовательской, аналитической и проектной деятельности;
 - выработка у обучающихся навыков научной дискуссии (экспертного обсуждения) и презентации исследовательских результатов, в т. ч. и через обсуждение проектов и готовых исследовательских работ обучающихся;
 - выработка у обучающихся навыков написания научных и информационно-аналитических работ.

В результате прохождения производственной (преддипломной) практики обучающиеся должны завершить анализ и обобщение материалов для магистерской диссертации.

По итогам прохождения преддипломной практики обучающиеся сдают отчетность следующих форм: письменный отчет и дневник.

2. ОРГАНИЗАЦИЯ ПРОИЗВОДСТВЕННОЙ (ПРЕДДИПЛОМНОЙ) ПРАКТИКИ

Производственная (преддипломная) практика проводится в научно-исследовательских организациях, ИТ-компаниях, на промышленных предприятиях или в структурных подразделениях университета. Выбор

организации–базы практики определяется самостоятельно обучающимся, исходя из темы магистерской диссертации.

Практика организуется:

а) на основе прямых договоров с предприятиями и организациями, заключаемыми университетом по своей инициативе;

б) на основе прямых договоров с предприятиями и организациями, заключаемыми университетом по инициативе обучающихся.

Базы практики для обучающихся должны отвечать следующим основным требованиям:

- соответствовать направлению и профилю магистерской программы;
- располагать квалифицированными кадрами для руководства практикой обучающихся.

Практика проводится в течение 4 недель в соответствии с ежегодно утверждаемым графиком учебного процесса.

Практика должна проходить в одном из подразделений предприятия (организации, учреждения), выполняющего экономические, информационные, организационные или управленческие функции, или их комплекс. Имея рабочее место в одном из таких подразделений, обучающиеся знакомятся с деятельностью других подразделений (если это необходимо по теме производственной практики) по согласованию с руководством предприятия.

До начала прохождения практики все обучающиеся проходят инструктаж по охране труда и технике безопасности в университете, по прибытию на место прохождения практики – в организации-базе практики.

Обучающиеся, совмещающие обучение с трудовой деятельностью, вправе проходить практику по месту трудовой деятельности в случаях, если профессиональная деятельность, осуществляемая ими, соответствует требованиям к содержанию практики (за исключением иностранных обучающихся, трудовая деятельность которых возможна в порядке, установленном Федеральным законом от 25.07.2002 г. №115-ФЗ «О правовом положении иностранных граждан в Российской Федерации»).

По завершении практики обучающимся составляется отчет, который утверждается руководителями практики от организации-базы практики и руководителем практики от Университета. Требования к содержанию отчета определены в Программе практики. Отчеты обучающихся по практике хранятся на кафедре в соответствии с утвержденной в Университете Номенклатурой дел.

По окончании периода прохождения практики обучающиеся сдают дифференцированный зачет.

2.1. Объем и продолжительность практики

Трудоемкость практики составляет 6 з.е. (216 ак. ч.)

Практика проводится в 4 семестре в течение 4 недель.

2.2. Планируемые результаты обучения при прохождении практики, соотнесенных с планируемыми результатами освоения образовательной программы

Прохождение преддипломной практики направлено на формирование следующих компетенций:

Код компетенции	Формулировка компетенции	Индикаторы достижения компетенции
УКи-7	Способен понимать фундаментальные принципы работы современных систем искусственного интеллекта, разрабатывать правила и стандарты взаимодействия человека и искусственного интеллекта и использовать их в социальной и профессиональной деятельности	<p>ИД-1 <small>УКи-7</small> Использует нормативно-правовую базу, правовые, этические правила, стандарты при решении задач искусственного интеллекта:</p> <ul style="list-style-type: none">- Знает правовую базу информационного законодательства, правовые нормы и стандарты в области искусственного интеллекта и смежных областей- Знает содержание нормативно-правовых документов в сфере информационных технологий, искусственного интеллекта и информационной безопасности- Умеет применять правовые нормы и стандарты в области искусственного интеллекта при создании систем искусственного интеллекта- Умеет применять этические нормы и стандарты в области искусственного интеллекта при создании систем искусственного интеллекта- Умеет использовать нормативно-правовые документы в сфере информационных технологий, искусственного интеллекта и информационной безопасности при разработке стандартов, норм и правил <p>ИД-2 <small>УКи-7</small> Разрабатывает стандарты, правила в сфере искусственного интеллекта и смежных областях и использует их в социальной и профессиональной деятельности:</p> <ul style="list-style-type: none">- Знает содержание основных международных и национальных стандартов и методологий разработки автоматизированных систем и программного обеспечения, стандартов в области

информационной безопасности, подходов к управлению и фундаментальные принципы работы, развития и использования технологий искусственного интеллекта

- Умеет использовать международные и национальные стандарты и методологии разработки автоматизированных систем программного обеспечения, стандартов в области информационной безопасности, принципы развития и использования технологий искусственного интеллекта при разработке стандартов, норм и правил в сфере искусственного интеллекта

ИД-3 УКи-7 Применяет современные методы и инструменты для представления результатов научно-исследовательской деятельности:

- Знает современные методы и инструменты для представления результатов научно-исследовательской деятельности

- Умеет применять современные методы и инструменты для представления результатов научно-исследовательской деятельности

ИД-4 УКи-7 Владеет нормами международного и российского законодательства в сфере интеллектуальной собственности:

- Знает нормы международного и российского законодательства в сфере интеллектуальной собственности

- Умеет применять нормы международного и российского законодательства в сфере интеллектуальной собственности

ИД-5 УКи-7 Проводит поиск зарегистрированных результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности:

- Знает методы выполнения поиска зарегистрированных результатов интеллектуальной деятельности и средств индивидуализации

- Умеет применять методы исследований результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной

		<p>деятельности</p> <p>ИД-6 <small>УКи-7</small> Осуществляет защиту прав результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности:</p> <ul style="list-style-type: none"> - Знает принципы защиты прав результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности - Умеет осуществлять защиту прав результатов интеллектуальной деятельности и средств индивидуализации при создании инновационных продуктов в профессиональной деятельности
ОПКи-9	Способен разрабатывать алгоритмы и программные средства для решения задач в области создания и применения искусственного интеллекта	<p>ИД-1 <small>ОПКи-9</small> Применяет инструментальные среды, программно-технические платформы для решения задач в области создания и применения искусственного интеллекта:</p> <ul style="list-style-type: none"> - Знает инструментальные среды, программно-технические платформы для решения профессиональных задач - Умеет применять инструментальные среды, программно-технические платформы для решения профессиональных задач <p>ИД-2 <small>ОПКи-9</small> Разрабатывает оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта:</p> <ul style="list-style-type: none"> - Знает принципы разработки оригинальных программных средств для решения профессиональных задач - Умеет разрабатывать оригинальные программные средства для решения задач в области создания и применения искусственного интеллекта
ОПКи-10	Способен адаптировать и применять на практике классические и новые научные принципы и методы исследований	<p>ИД-1 <small>ОПКи-10</small> Адаптирует известные научные принципы и методы исследований с целью их практического применения:</p> <ul style="list-style-type: none"> - Знает фундаментальные научные принципы и методы исследований - Умеет адаптировать с целью практического применения фундаментальные и новые научные принципы и методы исследований <p>ИД-2 <small>ОПКи-10</small> Решает профессиональные задачи на</p>

	<p>для решения задач в области создания и применения технологий и систем искусственного интеллекта и методы исследований</p>	<p>основе применения новых научных принципов и методов исследования:</p> <ul style="list-style-type: none"> - Знает особенности решения профессиональные задачи на основе применения новых научных принципов и методов исследования - Умеет разрабатывать, контролировать, оценивать и исследовать компоненты профессиональной деятельности; планировать самостоятельную деятельность в решении профессиональных задач
<p>ОПКи-11</p>	<p>Способен использовать методы научных исследований и математического моделирования в области проектирования и управления системами искусственного интеллекта</p>	<p>ИД-1 <small>ОПКи-11</small> Применяет логические методы и приемы научного исследования, методологические принципы современной науки, направления, концепции, источники знания и приемы работы с ними, основные особенности научного метода познания, программно-целевые методы решения научных проблем в профессиональной деятельности:</p> <ul style="list-style-type: none"> - Знает логические методы и приемы научного исследования; методологические принципы современной науки, направления, концепции, источники знания и приемы работы с ними; основные особенности научного метода познания; программно-целевые методы решения научных проблем; основы моделирования управленческих решений; динамические оптимизационные модели; математические модели оптимального управления для непрерывных и дискретных процессов, их сравнительный анализ; многокритериальные методы принятия решений в профессиональной деятельности - Умеет применять логические методы и приемы научного исследования; методологические принципы современной науки, концепции, источники знания и приемы работы с ними; основные метода научного познания; программно-целевые методы решения научных проблем; основы моделирования управленческих решений; динамические оптимизационные модели; математические модели оптимального управления для непрерывных и дискретных процессов, их сравнительный анализ; многокритериальные методы принятия решений в профессиональной деятельности <p>ИД-2 <small>ОПКи-11</small> Осуществляет методологическое обоснование научного исследования, создание и применение библиотек искусственного интеллекта:</p>

		<ul style="list-style-type: none"> - Знает приемы методологического обоснования научного исследования, методы организации библиотек искусственного интеллекта - Умеет проводить методологическое обоснование научного исследования, в том числе посредством создания и использования библиотек искусственного интеллекта
ОПКи-12	Способен осуществлять эффективное управление проектами по разработке и внедрению систем искусственного интеллекта	<p>ИД-1 <small>ОПКи-12</small> Исследует архитектуру информационных систем предприятий и организаций; применяет методологии и технологии реинжиниринга, проектирования и аудита информационных систем различных классов:</p> <ul style="list-style-type: none"> - Знает новые научные принципы и методы реинжиниринга, проектирования и аудита информационных систем для решения профессиональных задач - Умеет разрабатывать программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач <p>ИД-2 <small>ОПКи-12</small> Применяет инструментальные средства поддержки технологии проектирования и аудита информационных систем и сервисов; методы оценки экономической эффективности и качества, управления надежностью и информационной безопасностью:</p> <ul style="list-style-type: none"> - Знает особенности модернизации программного и аппаратного обеспечения информационных и автоматизированных систем для решения профессиональных задач - Умеет модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем для решения профессиональных задач <p>ИД-3 <small>ОПКи-12</small> Исследует особенности процессного подхода к управлению информационными системами и системами искусственного интеллекта; применяет системы управления качеством:</p> <ul style="list-style-type: none"> - Знает особенности процессного подхода к управлению информационными системами и системами искусственного интеллекта; системы управления качеством

<p>- Умеет применять системы управления качеством</p>
<p>ИД-4 <small>ОПКи-12</small> Выбирает методологию и технологию проектирования информационных систем; обосновывает архитектуру информационных систем и систем искусственного интеллекта:</p> <ul style="list-style-type: none"> - Знает методологию и технологию проектирования информационных систем - Умеет обосновывать архитектуру информационных систем и систем искусственного интеллекта
<p>ИД-5 <small>ОПКи-12</small> Управляет проектами по созданию (модификации) программного обеспечения, на всех стадиях жизненного цикла, оценивает эффективность и качество проекта; применяет современные методы управления проектами по разработке и внедрению систем искусственного интеллекта:</p> <ul style="list-style-type: none"> - Знает особенности управления проектами по созданию (модификации) программного обеспечения на всех стадиях жизненного цикла - Умеет оценивать эффективность и качество проекта; применять современные методы управления проектами и сервисами информационных систем и систем искусственного интеллекта
<p>ИД-6 <small>ОПКи-12</small> Использует инновационные подходы к проектированию информационных систем и систем искусственного интеллекта; принимает решения по информатизации предприятий в условиях неопределенности:</p> <ul style="list-style-type: none"> - Знает инновационные подходы к проектированию информационных систем и систем искусственного интеллекта - Умеет принимать решения по информатизации предприятий в условиях неопределенности
<p>ИД-7 <small>ОПКи-12</small> Проводит реинжиниринг прикладных и информационных процессов:</p> <ul style="list-style-type: none"> - Знает особенности процессного подхода, принципы реинжиниринга прикладных и информационных процессов - Умеет проводить реинжиниринг прикладных и информационных процессов

ПК-1	<p>Способен исследовать и разрабатывать архитектуры систем искусственного интеллекта для различных предметных областей на основе комплексов методов и инструментальных средств систем искусственного интеллекта</p>	<p>ИД-1_{ПК-1} Исследует и разрабатывает архитектуры систем искусственного интеллекта для различных предметных областей:</p> <ul style="list-style-type: none"> - Знает архитектурные принципы построения систем искусственного интеллекта, методы декомпозиции основных подсистем (компонентов) и реализации их взаимодействия на основе методологии предметно-ориентированного проектирования - Умеет выстраивать архитектуру системы искусственного интеллекта, осуществлять декомпозицию основных подсистем (компонентов) и реализации их взаимодействия на основе методологии предметно-ориентированного проектирования <p>ИД-2_{ПК-1} Выбирает комплексы методов и инструментальных средств искусственного интеллекта для решения задач в зависимости от особенностей предметной области:</p> <ul style="list-style-type: none"> - Знает методы и инструментальные средства систем искусственного интеллекта, критерии их выбора и методы комплексирования в рамках создания интегрированных гибридных интеллектуальных систем различного назначения - Умеет выбирать, применять и интегрировать методы и инструментальные средства систем искусственного интеллекта, критерии их выбора и методы комплексирования в рамках создания интегрированных гибридных интеллектуальных систем различного назначения
ПК-2	<p>Способен выбирать, разрабатывать и проводить экспериментальную проверку работоспособности программных компонентов систем искусственного интеллекта по обеспечению требуемых</p>	<p>ИД-1_{ПК-2} Выбирает и разрабатывает программные компоненты систем искусственного интеллекта:</p> <ul style="list-style-type: none"> - Знает основные критерии эффективности и качества функционирования системы искусственного интеллекта: точность, релевантность, достоверность, целостность, быстрота решения задач, надежность, защищенность функционирования систем искусственного интеллекта - Знает методы, языки и программные средства разработки программных компонентов систем искусственного интеллекта - Умеет выбирать, адаптировать, разрабатывать и

	критериев эффективности и качества функционирования	интегрировать программные компоненты систем искусственного интеллекта с учетом основных критериев эффективности и качества функционирования
		ИД-2 ПК-2 Проводит экспериментальную проверку работоспособности систем искусственного интеллекта: - Знает методы постановки задач, проведения и анализа тестовых и экспериментальных испытаний работоспособности систем искусственного интеллекта. - Умеет ставить задачи и проводить тестовые и экспериментальные испытания работоспособности систем искусственного интеллекта анализировать результаты и вносить изменения
ПК-3	Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач	ИД-1 ПК-3 Ставит задачи по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области: - Знает классы методов и алгоритмов машинного обучения - Умеет ставить задачи и разрабатывать новые методы и алгоритмы машинного обучения
		ИД-2 ПК-3 Руководит исследовательской группой по разработке или совершенствованию методов и алгоритмов для решения комплекса задач предметной области: - Знает методы и критерии оценки качества моделей машинного обучения - Умеет определять критерии и метрики оценки результатов моделирования при построении систем искусственного интеллекта в исследуемой области
ПК-4	Способен руководить проектами по созданию комплексных систем искусственного интеллекта	ИД-1 ПК-4 Руководит разработкой архитектуры комплексных систем искусственного интеллекта: - Знает возможности современных инструментальных средств и систем программирования для решения задач машинного обучения - Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения
		ИД-2 ПК-4 Осуществляет руководство созданием комплексных систем искусственного интеллекта

		<p>с применением новых методов и алгоритмов машинного обучения:</p> <ul style="list-style-type: none"> - Знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения - Знает принципы построения систем искусственного интеллекта, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта - Умеет применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения - Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта
ПК-5	<p>Способен руководить проектами по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов</p>	<p>ИД-1 ПК-5 Руководит работами по оценке и выбору моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи:</p> <ul style="list-style-type: none"> - Знает функциональность современных инструментальных средств и систем программирования в области создания моделей искусственных нейронных сетей - Умеет проводить оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения задач машинного обучения - Умеет применять современные инструментальные средства и системы программирования для разработки и обучения моделей искусственных нейронных сетей <p>ИД-2 ПК-5 Руководит созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств:</p> <ul style="list-style-type: none"> - Знает принципы построения систем искусственного интеллекта на основе искусственных нейронных сетей, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта - Умеет руководить выполнением коллективной

		<p>проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе искусственных нейронных сетей</p> <p>ИД-3 ПК-5 Руководит проектами по разработке систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов:</p> <ul style="list-style-type: none"> - Знает принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения (с подкреплением и без) - Знает подходы к применению моделей на основе нечеткой логики в системах искусственного интеллекта - Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов
ПК-6	Способен руководить проектами по созданию комплексных систем на основе аналитики больших данных в различных отраслях	<p>ИД-1 ПК-6 Осуществляет руководство проектом по построению комплексных систем на основе аналитики больших данных в различных отраслях:</p> <ul style="list-style-type: none"> - Знает методологию и принципы руководства проектом по созданию, поддержке и использованию комплексных систем на основе аналитики больших данных - Знает специфику сфер и отраслей, для которых реализуется проект по аналитике больших данных - Умеет решать задачи по руководству коллективной проектной деятельностью для создания, поддержки и использования комплексных систем на основе аналитики больших данных
ПК-7	Способен руководить проектами по созданию, внедрению и использованию одной или нескольких сквозных	<p>ИД-1 ПК-7 Руководит проектами в области сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений»:</p> <ul style="list-style-type: none"> - Знает фундаментальные правила построения рекомендательных систем и систем поддержки принятия решений, основанных на интеллектуальных принципах, методы и подходы к планированию и реализации проектов по

	цифровых субтехнологий искусственного интеллекта в прикладных областях	созданию систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений» - Умеет руководить проектами по созданию, внедрению и поддержке систем искусственного интеллекта на основе сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений»
ПК-8	Способен разрабатывать и модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности в различных предметных областях	<p>ИД-1 ПК-8 Разрабатывает программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях: - Знает новые научные принципы и методы разработки программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач в различных предметных областях - Умеет разрабатывать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности для решения профессиональных задач в различных предметных областях</p> <p>ИД-2 ПК-8 Модернизирует программное и аппаратное обеспечение технологий и систем искусственного интеллекта для решения профессиональных задач с учетом требований информационной безопасности в различных предметных областях: - Знает особенности модернизации программного и аппаратного обеспечения технологий и систем искусственного интеллекта для решения профессиональных задач в различных предметных областях - Умеет модернизировать программное и аппаратное обеспечение технологий и систем искусственного интеллекта с учетом требований информационной безопасности для решения профессиональных задач в различных предметных областях</p>

2.3. Содержание практики

Процесс прохождения практики состоит из нескольких этапов:

Этап 1. Подготовительный. Включает в себя: проведение организационного собрания; прохождение инструктажа по технике безопасности; организацию рабочего места.

Этап 2. Аналитический. Включает в себя: выполнение запланированной исследовательской и/или производственной работы, в т. ч. сбор и анализ информации о предмете исследования, статистическую и математическую обработку информации, анализ состояния предмета исследования (в соответствие с тематикой выпускной квалификационной работы), анализ научной литературы с использованием различных методик доступа к информации: посещение библиотек, работа с внутренней документацией организации-базы практики, работа в Интернете.

Этап 3. Оформление отчетных документов по практике (отчет и дневник), подготовка к защите отчета о прохождении практики. Включает в себя: сбор и обобщение материалов, изученных во время прохождения практики, анализ и в необходимых случаях проведение соответствующих расчетов по позициям практики с выводами и предложениями, подготовку форм отчетности: отчета и дневника о прохождении практики.

Этап 4. Отчетный. Включает в себя: получение отзыва (характеристики) руководителя от организации-базы практики и руководителя практики от Университета, защиту результатов преддипломной практики.

Этапы прохождения практики детализируются в таблице.

Этапы и виды работ производственной (преддипломной) практики

№ п/п	Этапы практики	Виды выполняемой работы, на практике включая самостоятельную работу обучающихся			Формы текущего контроля/ промежуточной аттестации
		Дата начала выполнения работы	Дата конца выполнения работы	Виды работ	
1	Подготовительный	Первый день	Первый день	Инструктаж по технике безопасности; выдача индивидуальных заданий	Оформление дневника практики
		Первый день	Первый-второй день	Знакомство со структурными подразделениями	Краткий обзор бизнес-

№ п/п	Этапы практики	Виды выполняемой работы, на практике включая самостоятельную работу обучающихся			Формы текущего контроля/ промежуточ- ной аттестации
		Дата начала выполнения работы	Дата конца выполнения работы	Виды работ	
2	Аналитический			организации	процессов организации
		Первый день	Первый- второй день	Составление плана работы	Оформление дневника практики (план практики)
		Второй день	Последний день	Выполнение заданий в соответствии с планом практики. Выполнение индивидуальных заданий. Сбор, обработка и систематизация фактического и литературного материала и другие виды работ, предусмотренные планом прохождения производственной практики.	Ведение дневника с практики
3.	Оформление отчетных документов	Двадцать первый день	Последний день	Обработка полученных результатов и написание отчета	Отчет
4.	Отчетный	Последний день	В течение 3- х дней после окончания практики	Защита отчета по практике	Дифференци- рованный зачет

2.4. Формирование отчетных документов по практике

Дневник практики является отчетным документом, характеризующим и подтверждающим прохождение обучающимся производственной практики, в котором должно быть оформлено направление на практику с отметкой о

прибытии и убытии обучающегося по месту прохождения практики, заверенные печатью организации, а также отражается его текущая работа в процессе практики:

- выданное обучающемуся индивидуальное задание на производственную (преддипломную) практику (приложение А);
- календарный план выполнения обучающимся программы практики с отметками о полноте и уровне его выполнения (план составляется совместно с куратором практики от кафедры и научным руководителем практики);
- анализ состава и содержания выполненной обучающимся практической работы с указанием структуры, объемов, сроков выполнения и ее оценки руководителем практики;
- перечень и обзор использованной обучающимся научной литературы (монографии, научные сборники и статьи, реферативные издания) и нормативных материалов (стандарты, отраслевые руководящие и методические материалы);
- выводы и предложения обучающегося по практике;
- краткая характеристика и оценка работы обучающегося в период практики руководителем практики от организации и руководителем от Университета.

В отчете магистрант описывает свою деятельность в период прохождения практики, оформляет выполнение программы практики и индивидуального задания, анализирует наиболее сложные вопросы, встретившиеся на практике, трудности при ее прохождении, вносит предложения по совершенствованию ее организации.

С дневником и отчетом о прохождении практики обязательно должен ознакомиться руководитель практики от организации, после чего он дает письменный отзыв (характеристику) о выполнении магистрантом программы практики, приобретенных им умениях и навыках. В характеристике также отражается отношение магистранта к трудовой деятельности, наличие особенных деловых качеств. Отзыв руководителя практики от организации должен быть оформлен в дневнике за подписью руководителя практики от организации или руководителя подразделения организации или руководителя организации и заверен печатью организации.

Оформление отчета о прохождении практики

Изложение текста и оформление отчета выполняют в соответствии с требованиями ГОСТ Р 2.105-2019. Отчет о прохождении практики оформляется на стандартной бумаге формата А4, общим объемом 20-30 страниц.

Оформление текстовой и иллюстративной части отчета осуществляется с учетом следующих требований.

- поля: левое – 30 мм, правое – 10 мм, верхнее и нижнее – 20 мм;
- шрифт основного текста Times New Roman, размер 14 пунктов, обычный, междустрочный интервал – одинарный, при форматировании следует устанавливать выравнивание текста по ширине, отступ первой строки абзаца – 1,25 см (абзацный отступ должен быть одинаковым по всему тексту документа.);
- шрифт заголовков – Times New Roman, 14 пунктов, полужирный;
- шрифт подзаголовков Times New Roman, 14 пунктов, обычный;
- номер страницы проставляют вверху страницы на расстоянии не менее 10 мм от верхнего края листа. На титульном листе и на листах, соответствующих началу разделов, номера страниц не ставят, но подразумевают, что отражается в последующей нумерации листов текстового документа;
- каждая структурная часть отчета начинается с нового листа; точка в конце заголовка структурной части не ставится;
- заголовки отчета (введение, заголовки разделов, заключение) выравниваются по левому краю, остальные заголовки выравниваются по центру;
- расстояние между заголовком и текстом при выполнении документа печатным способом должно быть равно двум одинарным интервалам;
- расстояние между заголовками раздела и подраздела – по одному печатному интервалу;
- при представлении табличного материала над таблицей с правой стороны помещают заголовок «Таблица» с указанием ее порядкового номера (сквозная нумерация) и тематического заголовка;
- приводимые в отчете иллюстрации (схема, диаграмма, график, технический рисунок, фотография) должны иметь порядковый номер (сквозная нумерация) и подрисуночную подпись;
- отчет представляется на провепку в сброшюрованном виде (листы должны быть скреплены по левому краю).

Форма титульного листа отчета по преддипломной практике приведена в приложении (Приложение Б).

Отчет проверяется и подписывается непосредственным руководителем практики от университета и руководителем практики от профильной организации.

Содержание и оформление отчета должны соответствовать требованиям, разработанным выпускающей кафедрой. Информационные блоки отчета должны быть представлены в следующем порядке:

1. Титульный лист.
2. Содержание.
3. Введение (цели и задачи практики, краткая характеристика базы и места практики, описание основных видов деятельности, выполняемых практикантом. Введение должно обобщить собранные материалы и раскрыть основные вопросы и направления, которыми занимался обучающийся на практике).
4. Основная часть - разделы и подразделы (сведения о конкретно выполненной обучающимся работе в период практики в соответствии с индивидуальным заданием, достигнутые результаты).
5. Заключение (выводы о результатах практики, общие результаты проделанной работы, анализ возникших проблем, практические рекомендации).
6. Список литературы.
7. Приложения (при наличии).

2.5. Защита отчета по практике, промежуточная аттестация

Преддипломная практика оценивается руководителем от университета. Защита оформленного письменного отчета обучающегося по практике проходит в ходе сдачи дифференцированного зачета по практике. Оценки по практике приравниваются к оценкам (зачетам) по теоретическому обучению и учитываются при подведении итогов общей успеваемости обучающегося. Оценка по практике проставляется одновременно в экзаменационную ведомость и зачетную книжку руководителями практики.

К защите допускаются обучающиеся, прошедшие практику в установленные сроки и выполнившие ее программу. На защиту представляются:

- заполненный по всем разделам дневник практики, подписанный руководителем практики от профильной организации, в которой обучающийся проходил преддипломную практику;
- отчет по практике, включающий текстовые, табличные и графические материалы, отражающие решение предусмотренных программой практики задач;
- отзыв руководителя практики о работе обучающегося в период практики с оценкой уровня и оперативности выполнения им задания по практике, отношения к выполнению программы практики и т.п.

Критериями результатов прохождения практики магистрантом являются:

- отзыв-характеристика руководителя преддипломной практики от профильной организации;
- объем выполнения индивидуального задания на период преддипломной практики;
- качество предоставленных отчетных материалов по практике;
- уровень знаний нормативно-правовой базы информационного законодательства, правовых норм и стандартов в области искусственного интеллекта, информационной безопасности и смежных областей, регулирующей деятельность учреждения или организации, в которых обучающийся проходил практику;
- умение обобщать и анализировать научную литературу и практику создания и использования систем искусственного интеллекта по теме магистерского исследования.

Обучающиеся, не прошедшие практику по уважительной причине, не выполнившие программу практики либо не приступившие к прохождению практики, направляются на практику повторно в свободное от аудиторных занятий время. Если повторное прохождение практики осуществляется в свободное от аудиторных занятий время (параллельно с учебным процессом), календарная продолжительность практики увеличивается в два раза. Повторное направление на практику осуществляется оформлением нового приказа.

Обучающиеся, не прошедшие практику и/или не выполнившие программу и/или получившие неудовлетворительную оценку по итогам практики, считаются имеющими академическую задолженность и обязаны ликвидировать ее в соответствии со сроками, установленными локальными актами Университета.

3. ПЕРЕЧЕНЬ УЧЕБНОЙ ЛИТЕРАТУРЫ И РЕСУРСОВ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫХ ДЛЯ ПРОВЕДЕНИЯ ПРАКТИКИ

Литература:

1. Т.В. Афанасьева. Информационное общество и проблемы прикладной информатики: учебное пособие / Т.В. Афанасьева. – Ульяновск : УлГТУ, 2018. – 116 с. <http://venec.ulstu.ru/lib/disk/2017/344.pdf>
2. Советов, Б. Я. Моделирование систем: учебник для обучающихся вузов, обучающихся по инженерно-техническим направлениям и специальностям / Советов Б. Я., Яковлев С. А.; С. - Петерб. гос. электротехн. ун-т "ПЭТИ им. В. И. Ульянова-Ленина". - 7-е изд. - Москва: Юрайт, 2014. - 343 с.: ил. Гриф: УМО РФ.

3. Т.В. Афанасьева. Моделирование в задачах анализа свойств систем: учебное пособие / Т.В. Афанасьева. – Ульяновск : УлГТУ, 2019. – 114 с.
4. Афанасьева, Т. В., Афанасьев А.Н. Введение в проектирование систем интеллектуального анализа данных: учебное пособие. – Ульяновск : УлГТУ, 2017. 64 с. <http://venec.ulstu.ru/lib/disk/2017/231.pdf>
5. Воронина В. В. Теория и практика машинного обучения: учебное пособие /В. В. Воронина и др.. – Ульяновск: УлГТУ, 2017. – 290 с. <http://venec.ulstu.ru/lib/disk/2017/191.pdf>
6. Кувайскова, Ю.Е.. Алгоритмы дискретной математики : учебное пособие / Ю. Е. Кувайскова. – Ульяновск : УлГТУ, 2017. – 99 с. <http://venec.ulstu.ru/lib/disk/2017/28.pdf>
7. Афанасьева, Т.В. Онтологический и нечеткий анализ слабоструктурированных информационных ресурсов: научное издание/ Т.В. Афанасьева, В.С. Мошкин, А.М. Наместников, И.А. Тимина, Н.Г. Ярушкина; под ред. Н.Г. Ярушкиной.- Ульяновск: УлГТУ, 2016.- 130 с. <http://venec.ulstu.ru/lib/disk/2017/11.pdf>
8. Коваленко, Владимир Васильевич. Проектирование информационных систем: учебное пособие для вузов / Коваленко В. В. - Москва: Форум, 2012. - (Высшее образование). - 319 с.: ил. - ISBN 978-5-91134-549-5.
9. Клименко, И. С. Теория систем и системный анализ : учебное пособие / И. С. Клименко. — Сочи : РосНОУ, 2018. — 264 с. // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/162178>
- 10.Кориков, Анатолий Михайлович. Теория систем и системный анализ: учебное пособие для вузов / Кориков А. М., Павлов С. Н. - Москва: ИНФРА-М, 2016.
- 11.Т.В. Афанасьева. Основы управления качеством программных средств: учебное пособие / Т.В. Афанасьева, А.Н. Афанасьев. – Ульяновск : УлГТУ, 2017. <http://venec.ulstu.ru/lib/disk/2017/232.pdf>
- 12.Гагарина Л.Г., Кокорева Е.В., Виснадул Б.Д.. Технология разработки программного обеспечения: учебное пособие / под. Ред. Л.Г. Гагариной. – М. : ИД «ФОРУМ» ИН-ФРА-М, 2011. – 399 с.
- 13.Похилько, Александр Федорович. Моделирование процессов и данных с использованием CASE-технологий [Текст]: учебное пособие / Похилько А. Ф., Горбачев И. В., Рябов С. В.; М-во образования и науки Рос. Федерации, Ульян. гос. техн. ун-т. - Ульяновск: УлГТУ, 2014. - 163 с. <http://venec.ulstu.ru/lib/disk/2014/179.pdf>

14. Исследование ИТ-кластера Ульяновской области / Н.Г. Ярушкина, Т.В. Афанасьева, О.В. Шиняева и др. – Ульяновск : УлГТУ, 2013. – 137 с. <http://venec.ulstu.ru/lib/disk/2014/35.pdf>
15. Чернышев, Илья Васильевич. Информационные системы в экономике [Текст]: учебное пособие / Чернышев И. В.; М-во образования и науки Рос. Федерации, Ульян. гос. техн. ун-т. - Ульяновск: УлГТУ, 2014. - 113 с.: табл. - Библиогр.: с. 108 <http://venec.ulstu.ru/lib/disk/2015/14.pdf>
16. В. Г. Тронин. Планирование и управление научными проектами с применением современных ИКТ: учебное пособие: УлГТУ, 2019. <http://venec.ulstu.ru/lib/disk/2019/145.pdf>
17. Семушин, И. В. Письменная и устная научная коммуникация [Электронный ресурс]: учебное пособие / Семушин И. В.; М-во образования и науки Рос. Федерации, Ульян. гос. техн. ун-т. - Электрон. текст. данные (Файл pdf). - Ульяновск: УлГТУ, 2014. - Доступен в Интернете. - Библиогр.: с. 140-142 <http://venec.ulstu.ru/lib/disk/2015/11.pdf>
18. Написание и презентация научной работы. Существенные навыки для обучающихся, магистрантов и аспирантов : электронное учебное пособие / составитель и разработчик макета И. В. Семушин. – Ульяновск, 2013. – 1148 слайдов (312 фреймов). <http://venec.ulstu.ru/lib/disk/2014/Semushin-root/Semushin.pdf>

Учебно-методическое обеспечение:

1. Организация магистерских научно-исследовательских работ : методические рекомендации / Т. В. Афанасьева. – Ульяновск : УлГТУ, 2015. – 37 с. <http://venec.ulstu.ru/lib/disk/2015/198.pdf>
2. Гуськов Г.Ю. Расширение функциональности средств демонстрации презентаций // Информатика, моделирование, автоматизация проектирования: сборник научных трудов / под ред. Н. Н. Войта. – Ульяновск : УлГТУ, 2011.– 416 с. <http://venec.ulstu.ru/lib/disk/2012/Voit1.pdf>
3. Афанасьева Т. В. Преддипломная практика [Электронный ресурс]: методические указания. – Ульяновск : УлГТУ, 2017. virtual.ulstu.ru
4. Новиков, Ю. Н. Подготовка и защита бакалаврской работы, магистерской диссертации, дипломного проекта : учебное пособие для вузов / Ю. Н. Новиков. — 5-е изд. испр. и доп. — Санкт-Петербург : Лань, 2021. — 36 с. — ISBN 978-5-8114-4727-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/174283>

Ресурсы сети «Интернет»:

1. Интернет-портал образовательных ресурсов по ИТ <http://www.intuit.ru>
2. Википедия – свободная энциклопедия <https://ru.wikipedia.org>

3. Федеральный портал Единое окно доступа к образовательным ресурсам
<http://window.edu.ru/library>
4. Научная электронная библиотека <http://elibrary.ru/defaultx.asp>
5. РГБ фонд диссертаций <http://diss.rsl.ru/>

**Типовое индивидуальное задание
по производственной (преддипломной) практике**

1. Сформировать обзор предметной области по теме магистерской диссертации, включая актуальность проблематики, обзор существующих методов и инструментальных средств искусственного интеллекта для решения задач в зависимости от особенностей предметной области, провести анализ эффективности их использования для решения задач по теме магистерской диссертации.

2. Исследовать архитектуру информационных систем базы практики.

3. Применяя методологии и технологии реинжиниринга, проектирования и аудита информационных систем различных классов, разработать архитектуру системы искусственного интеллекта для предметной области по теме магистерской диссертации на основе комплексов методов и инструментальных средств систем искусственного интеллекта.

4. Разработать и провести экспериментальную проверку работоспособности программных компонентов системы искусственного интеллекта по обеспечению требуемых критериев эффективности и качества функционирования, включая автоматизированные тесты (например, нагрузочное тестирование — написание кода для автоматического генерирования большого объема данных в базе данных и ручной проверке поведения программы при работе с разными объемами данных)

5. Оформить протокол тестирования: цель, объект, метод, результат, включая пошаговое исполнение.

6. Оформить отчет по практике, который должен структурно содержать следующие компоненты (разделы):

1. Титульный лист.

2. Содержание.

3. Введение (цели и задачи практики, краткая характеристика базы и места практики, описание основных видов деятельности, выполняемых практикантом. Введение должно обобщить собранные материалы и раскрыть основные вопросы и направления, которыми занимался обучающийся на практике).

4. Основная часть - разделы и подразделы (сведения о конкретно выполненной обучающимся работе в период практики в соответствии с индивидуальным заданием, достигнутые результаты).

5. Заключение (выводы о результатах практики, общие результаты проделанной работы, анализ возникших проблем практические рекомендации)
6. Список литературы.
7. Приложения (при наличии).

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных систем и технологий
Кафедра Информационные системы

ОТЧЕТ
по производственной (преддипломной) практике

(вид практики)

Студента _____
(ФИО)

Направление (специальность,
профиль) подготовки _____

Группа _____

Место прохождения практики _____

(наименование профильной организации, подразделение университета)

Отчет по практике
защищен с оценкой

(оценка прописью)

Руководитель практики
от университета

(должность)

(подпись)

(ФИО)

Ульяновск 20__ г.



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования

«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

ОФОРМЛЕНИЕ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ МАГИСТРА

Методические указания

Составители: Н.В. Корунова
Е.Н. Эгов

Ульяновск
УлГТУ
2021

УДК
ББК
М

Одобрено секцией методических пособий научно-методического совета университета.

Оформление выпускной квалификационной работы магистра: методические указания / сост. Н. В. Корунова., Е.Н. Эгов – Ульяновск: УлГТУ, 2021. – 49 с.

М

Изложены основные требования и даны рекомендации студентам по оформлению выпускных квалификационных работ. Правила оформления учебной документации приведены в соответствии со стандартами. Предназначены для магистров, обучающихся по направлениям 09.04.03 «Прикладная информатика», 09.04.04 «Программная инженерия».

Работа подготовлена на кафедре «Информационные системы».

УДК
ББК

© Корунова Н.В., составление, 2021
© Эгов Е.Н., составление, 2021

© Оформление. УлГТУ, 2021

Содержание

Введение	4
1 Общие положения	5
2 Выпускная квалификационная работа в учебной программе	7
3 Структура, объем пояснительной записки	8
4 Правила оформления пояснительной записки	12
4.1 Общие требования	12
4.2 Рубрикация и заголовки	15
4.3 Содержание	16
4.4 Перечисления, знаки и числа в тексте	16
4.5 Сокращения и условные обозначения	18
4.6 Единицы измерения и размерности	19
4.7 Индексы буквенных обозначений	20
4.8 Математические формулы	20
4.9 Таблицы и выводы	21
4.10 Иллюстрации	24
4.11 Список литературы	26
4.12 Приложения	31
4.13 Исходный код программы (листинг)	32
5 Нормоконтроль	33
Список использованных источников	36
Приложение А	40
Приложение Б	41
Приложение В	43
Приложение Г	45
Приложение Д	47
Приложение Е	48

Введение

Данные методические указания предоставляют изложение требований по оформлению учебной документации на выпускную квалификационную работу на кафедре «Информационные системы» Ульяновского государственного технического университета и подробно освещают правила оформления пояснительной записки в соответствии с требованиями стандартов.

Учебно-методические указания, определяющие требования к содержанию выпускной квалификационной работы с учетом специфики предметной области конкретных областей знаний, разрабатывают преподаватели соответствующих направлений подготовки.

При подготовке методических указаний были использованы следующие основные стандарты:

1. ГОСТ 7.32–2017. Отчет о научно-исследовательской работе. Структура и правила оформления.
2. ГОСТ 2.105–2019. Общие требования к текстовым документам.
3. ГОСТ Р 7.0.97–2016. Требования к оформлению документов.

1 Общие положения

В соответствии с требованиями «Положения о государственной итоговой аттестации» [3] защита выпускной квалификационной работы является составляющей государственной итоговой аттестации выпускников вузов, целью которой является установление уровня подготовки выпускника высшего учебного заведения к выполнению профессиональных задач и соответствия его подготовки требованиям федерального государственного образовательного стандарта высшего образования.

Термины и определения

В данном методическом указании применены следующие термины с соответствующими определениями, приведенными ниже.

Выпускная квалификационная работа (ВКР) магистра (магистерская диссертация) — законченное исследование на заданную тему по профессиональной образовательной программе ВПО, написанное лично автором под руководством научного руководителя, содержащее элементы научного исследования и свидетельствующее об умении автора работать с литературой, обобщать и анализировать фактический материал, демонстрирующее владение общекультурными и профессиональными компетенциями, приобретенными при освоении профессиональной образовательной программы.

ВКР магистра обозначает подготовленность к самостоятельной научной деятельности и практической работе в соответствии с полученной квалификацией.

Научный руководитель — специалист в научно-производственной области, в рамках которой определена тема ВКР, обладающий высокой квалификацией и надлежащей педагогической компетенцией (наличие специального образования или документа о повышении квалификации в психолого-педагогической сфере).

Рецензент — дипломированный специалист в научно-производственной области, в рамках которой определена тема ВКР,

обладающий высокой квалификацией, позволяющей ему оценить выполненную работу на соответствие темы, содержания, актуальности, новизны, полноты решения направлению подготовки.

Научный консультант — специалист в узкой научно-производственной области, использующейся при написании ВКР, по которой компетенции научного руководителя недостаточно.

Обозначения и сокращения

В данном методическом указании применяются следующие обозначения и сокращения:

ВКР – выпускная квалификационная работа

МД – магистерская диссертация

ВО – высшее образование

ГИА – государственная итоговая аттестация

ФГОС – федеральный государственный образовательный стандарт

ПЗ – пояснительная записка.

2 Выпускная квалификационная работа в учебной программе

Выполнение и защита ВКР являются видом учебной деятельности, который завершает процесс освоения студентом образовательной программы ВПО и выполняется на последнем году обучения. В соответствии с положением [3] содержание ВКР и уровень ее защиты рассматриваются как основной критерий при оценке уровня профессиональной подготовки выпускника и качества реализации образовательной программы.

Цели защиты ВКР – установление уровня подготовки выпускника государственной аттестационной комиссией к выполнению профессиональных задач и соответствия его подготовки требованиям ФГОС ВО [1, 2].

МД выполняется студентом самостоятельно на базе теоретических знаний и практических навыков, полученных обучающимся в период обучения. При этом она преимущественно ориентирована на знания, полученные в процессе изучения общепрофессиональных и специальных дисциплин. ВКР магистра может быть, прикладного, аналитического или научно-исследовательского характера.

В рамках учебной программы к ВКР предъявляются следующие общие требования по оформлению:

- соответствие названия работы ее содержанию, четкая целевая направленность, актуальность, наличие научной новизны;
- логическая последовательность изложения материала, базирующаяся на прочных теоретических знаниях по избранной теме и убедительных аргументах;
- корректное изложение материала с учетом принятой научной терминологии;
- достоверность полученных результатов и обоснованность выводов;
- научный стиль написания;
- оформление работы в соответствии с требованиями ГОСТ.

3 Структура, объем пояснительной записки

Материал магистерской диссертации должен быть систематизирован и оформлен надлежащим образом. Для аргументации своих предложений, решений и выводов необходимо оформлять ссылки на соответствующие источники сведения, напрямую без дополнительной переработки заимствованные из литературных источников и сети Internet.

Общими требованиями к ВКР являются:

- логическая последовательность и преемственность изложения материала;
- убедительность аргументации выбранных методов анализа, расчетов и предложений;
- краткость и четкость формулировок;
- конкретность изложения результатов работы;
- доказательность выводов и обоснованность рекомендаций.

По результатам выполнения выпускной квалификационной работы оформляется документация, в случае магистерской диссертации – это пояснительная записка, структура и объем данной документации устанавливаются кафедрой, исходя из характера проекта и направления подготовки.

Материал ВКР должен быть систематизирован и оформлен надлежащим образом.

Рекомендуемый объем магистерской диссертации без учета приложений должен составлять от 60 до 80 листов. Объем приложений не ограничен. Список литературы должен начинаться минимум с 60 страницы, не ранее.

Рекомендуемый состав и порядок расположения материала в ВКР:

- титульный лист (прил. А) (блок «УТВЕРЖДАЮ» («Руководитель предприятия») заполнять в случае наличия акта о внедрении, в противном случае можно полностью удалить его, блок «К ЗАЩИТЕ ДОПУСТИТЬ» заполняется по указаниям кафедры);

- задание на магистерскую диссертацию работу (прил. Б) (блок «УТВЕРЖДАЮ» заполняется по указаниям кафедры);
- реферат (аннотация);
- пояснительная записка:
 - а) содержание;
 - б) перечень используемых условных обозначений, сокращений, терминов;
 - в) введение (указывается тема работы, обосновывается ее выбор, раскрывается актуальность темы, определяются объект и предмет исследования, определяется цель работы и основные задачи, приводятся обоснования научной новизны работы, указываются используемые методики);
 - г) основная часть (структура основной части должна включать в себя научный обзор проблемной области, существующих методов, алгоритмов и решений заявленной цели в описываемой предметной области; детальное представление собственных методов, алгоритмов или решений заявленной цели, с выделением научной новизны предлагаемых методов, алгоритмов по сравнению с существующими; описание программной реализации предлагаемых методов, алгоритмов и решений; качественные эксперименты, подтверждающие достижение заявленных целей и задач);
 - д) заключение (включает выводы и рекомендации);
 - ж) список использованных источников, в т. ч. нормативных, проектных и справочных материалов;
- з) приложения:
 - демонстрационно-графическая часть (при необходимости);
 - протоколы тестирования;
 - листинг;
- и) последний лист;
- к) носитель информации с пояснительной запиской, исходным кодом программы и прочими документами.

В пояснительную записку вкладываются, но не подшиваются:

- отзыв руководителя (прил. В);
- отзыв рецензента (прил. Г);
- первый лист антиплагиата (где указан общий процент оригинальности работы) с указанием фамилии и росписью;
- заявление о самостоятельности работы (прил. Д);
- отзыв от специалиста-практика (при наличии);
- акт о внедрении (при наличии);

Структурный элемент «Титульный лист» является обязательным для любого текстового документа. Образец титульного листа приведен в приложении А. На титульном листе в каждом конкретном случае уточняются тема, данные о студенте, руководителе и рецензенте.

Структурный элемент «Задание» является обязательным для ВКР. Задание должно содержать наименование темы ВКР и предусматривать по возможности комплексное решение исследовательских задач. Шаблон задания приведен в приложении Б.

Структурные элементы «Содержание», «Основная часть», «Список использованных источников» являются обязательными для пояснительной записки любого вида ВКР.

В содержании приводятся заголовки разделов, граф, параграфов и т. д. с указанием страниц всех частей работы. При этом заголовки и их рубрикационные индексы должны быть приведены в строгом соответствии с текстом.

Список использованных источников должен содержать сведения об источниках, использованных при выполнении ВКР. Первоначальный список литературы для изучения указывается в задании, окончательный – в тексте пояснительной записки магистерской диссертации. Он оформляется в соответствии с требованиями ГОСТ Р 7.0.12–2011 [5] и его наличие в пояснительной записке обязательно. На более чем 80% источников в работе должны быть даны ссылки. Необходимо минимум 30 ссылок на различные источники литературы.

Структурный элемент «Приложения» заполняется демонстрационно-графической частью, в том числе иллюстрациями, таблицами, листингом программ и диаграммами.

Последний лист содержит в себе заявление о самостоятельности выполненной работы (прил. Е)

Остальные структурные элементы включают в конкретный текстовый документ, исходя из его требований к содержанию.

ВКР оформляются в соответствии с требованиями государственных стандартов, а также соответствующих требований УлГТУ.

4 Правила оформления пояснительной записки

4.1 Общие требования

Выпускная квалификационная работа оформляется в соответствии с требованиями государственных стандартов, действующих на территории Российской Федерации, а также соответствующих требований УЛГТУ.

Текст выпускной квалификационной работы оформляется в виде пояснительной записки.

Пояснительная записка представляет собой текстовый документ, содержащий описания проблем, решаемых в ходе работы над ВКР, расчеты и описание проектируемого объекта, принцип его действия, научную новизну, обоснование принятых технических, технологических и технико-экономических решений.

Текст пояснительной записки оформляется в текстовом редакторе (MS Word или любой свободно распространяемый офисный пакет с возможностью создания и редактирования текстового документа). Материал пояснительной записки излагается грамотно, четко, сжато. Расчеты иллюстрируются эскизами, схемами, эпюрами, графиками, диаграммами, выполненными соответствующими программными средствами.

Каждый лист пояснительной записки ВКР для технических направлений/специальностей заключается в рамку.

По ГОСТ 7.32–2017 [6] текст печатается на одной стороне листа белой бумаги формата А4, при этом размеры полей: правое – не менее 10 мм, верхнее и нижнее – не менее 20 мм, левое – не менее 30 мм.

При оформлении в текстовом редакторе следует соблюдать следующие параметры: выбранный шрифт должен быть четким и разборчивым (рекомендуется «times new roman», размер шрифта – 14 или «arial» размер – 12), печать через 1,5 интервала. Цвет шрифта – черный.

Названия глав, параграфов, пунктов, подпунктов следует начинать с абзаца, их можно писать более крупным кеглем, чем текст. Допускается выделение интенсивностью (полужирный шрифт).

Каждая глава должна начинаться с новой страницы. Названия глав (разделов), параграфов (подразделов) должны соответствовать оглавлению (содержанию) и быть оформлены единообразно во всем документе.

Все страницы работы (за исключением титульного листа, задания и последнего листа) должны быть оформлены рамками с основными надписями по форме 2 (первая страница раздела «Содержание») и 2а (рисунки 1,2). Страницы ПЗ следует нумеровать арабскими цифрами, соблюдая сквозную нумерацию по всему тексту ПЗ. Номер страницы проставляют в центре нижней части листа без точки, черточек и скобок. Титульный лист включают в общую нумерацию страниц ПЗ. Номер страницы на титульном листе не проставляют, но учитывают при нумерации. Иллюстрации и таблицы, расположенные на отдельных листах, включают в общую нумерацию страниц ПЗ.

					(2)			
Изм	Лист	№ документа	Подпись	Дата	(1)	Лит.	Лист	Листов
Разраб.		(7)				(3)	(4)	(5)
Руков.		(7)				(6)		
Рецензент		(7)						
Н. контр.		(7)						
Утвержд.		(7)						

Рисунок 1 – Структура основной надписи по форме 2

					(2)	Лист
						(8)
Изм	Лист	№ документа	Подпись	Дата		

Рисунок 2 – Структура основной надписи по форме 2а

Заполнение граф основной надписи:

- В графе 1 – тема магистерской диссертации.
- В графе 2 – обозначение документа. Для обозначения документов, входящих в дипломный проект, рекомендуется придерживаться структуры, указанной на рисунке 3.

<i>Аббревиатура ВУЗа</i>	<i>Номер зачетки</i>
МД – УлГТУ – 09.04.03 – 19/0180 – 2021	
<i>Магистерская диссертация</i>	<i>Шифр направления подготовки</i>
	<i>Год выпуска</i>

Рисунок 3 – Обозначение документа

Шифр направления:

- Прикладная информатика – 09.04.03;
- Программная инженерия – 09.04.04.
- В графе 3 – литера «М» (магистерская диссертация).
- В графе 4 – порядковый номер листа.
- В графе 5 – количество листов в пояснительной записке (считается по список литературы, приложения уже не учитываются).
- В графе 6– помещают: кафедра ____, учебная группа ____.
- В графе 7 указывают фамилии:
 - в строке «Разраб.» – исполнителя проекта;
 - в строке «Руков.» – руководителя проекта;
 - в строке «Рецензент» –рецензент работы;
 - в строке «Н.контр.» – ответственного за соблюдение требований ЕСКД от кафедры Информационные системы;
 - в строке «Утвержд.» – заполняется по указанию кафедры;
- в графе 8 – Порядковый номер страницы записки.

Пример заполнения основной надписи приведен на рисунках 4, 5.

					МД–УлГТУ– 09.04.03–19/0180–2021			
Изм.	Лист	№ докум.	Подп.	Дата				
Разраб.		<i>Иванов И.И.</i>			<i>Разработка информационной системы документооборота для подразделений областной администрации</i>	Лит.	Лист	Листов
Руков.		<i>Ярушкина Н.Г.</i>				<i>М</i>	<i>4</i>	<i>70</i>
Рецензент		<i>Петров П.П.</i>				<i>кафедра ИС группа ИСЭмд-21</i>		
Н.контр.		<i>Корунова Н.В.</i>						
Утвержд.		<i>Романов А.А.</i>						

Рисунок 4 – Основная надпись по форме 2

					МД–УлГТУ– 09.04.03–19/0180–2021	Лист
Изм.	Лист	№ докум.	Подп.	Дата		6

Рисунок 5 – Основная надпись по форме 2а

4.2 Рубрикация и заголовки

По ГОСТ 7.32–2017 [6] главы основной части работы не являются структурными элементами. Таким элементом (наряду с содержанием, введением, заключением, списком использованных источников, приложением и др.) является только вся основная часть в целом. Каждый структурный элемент следует начинать с новой страницы.

Разделы (главы) пояснительной записки могут делиться на подразделы (параграфы), которые в свою очередь могут делиться на пункты и подпункты (и более мелкие разделы). При делении текста на пункты и подпункты необходимо, чтобы каждый пункт содержал законченную информацию.



Номер подраздела состоит из номеров раздела и подраздела в разделе, разделенных точкой. В конце номера точка не ставится. Аналогичным образом нумеруются и пункты в подразделе (например, 1.4.3 Вывод анализа). В принципе, допускается наличие в разделе всего одного подраздела, а в подразделе – одного пункта. В этом случае подраздел и пункт все равно нумеруются. Заголовки подразделов, пунктов и подпунктов следует печатать с абзацного отступа с прописной буквы без точки в конце, не подчеркивая. Если заголовок состоит из двух предложений, их разделяют точкой. Переносы слов в заголовках не допускаются. Заголовки в содержании должны точно соответствовать заголовкам в тексте.

Если основная часть пояснительной записки не имеет подразделов, то нумерация пунктов в нем должна быть в пределах каждого раздела, и номер пункта должен состоять из номеров раздела и пункта, разделенных точкой.

Пример нумерации и заголовков подразделов и пунктов:

3 Методы тестирования

3.1 Структурное тестирование

- 3.1.1
- 3.1.2  Нумерация пунктов первого параграфа третьей главы
- 3.1.3
- 3.2 Функциональное тестирование
- 3.2.1
- 3.2.2  Нумерация пунктов второго параграфа третьей главы
- 3.2.3

Размер абзацного отступа регламентируется ГОСТ Р 2.105-2019 «Общие требования к текстовым документам» [7], по которому абзацный отступ равен пяти ударам пишущей машинки (или 15–17 мм). Расстояние между заголовком и текстом должно быть равно 3 или 4 интервалам (15 мм). Если основной текст напечатан интервалом 1,5, то это значит, что расстояние между заголовком и текстом равно одной пустой строке. Расстояние между заголовками главы и параграфа – 2 интервала (8 мм).

4.3 Содержание

По ГОСТ 7.32–2017 [6] заголовок СОДЕРЖАНИЕ пишется заглавными буквами посередине строки.

Содержание включает введение, наименование всех разделов, подразделов, пунктов, заключение, список использованных источников и наименование приложений с указанием номеров страниц, с которых начинаются эти элементы диссертации.

По ГОСТ Р 2.105-2019 [7] наименования, включенные в содержание, записывают строчными буквами, начиная с прописной буквы.

4.4 Перечисления, знаки и числа в тексте

Внутри пунктов или подпунктов могут быть приведены перечисления. По ГОСТ 7.32–2017 [6] перед каждым перечислением следует ставить дефис или, при необходимости ссылки в тексте документа на одно из перечислений, строчную букву (за исключением ё, з, о, г, ь, и, ы, ъ).

Для дальнейшей детализации перечислений необходимо использовать арабские цифры, после которых ставится скобка, а запись производится с абзацного отступа, как показано в примере.

Пример:

а) _____

1) _____

2) _____

б) _____

в) _____

Перечисления, состоящие из отдельных слов и небольших словосочетаний (без знаков препинания), пишутся в подбор с текстом со строчных букв и отделяются запятыми.

Пример: ...спиральная модель эволюционной стратегии проектирования программных продуктов определяет четыре действия: 1 – планирование, 2 – анализ риска, 3 – конструирование, 4 – оценивание.

Если перечисление состоит из отдельных фраз или развернутых сочетаний со знаками препинания, то каждый элемент пишут с новой строки и отделяют фразы точкой с запятой.

Пример: ...спиральная модель эволюционной стратегии конструирования определяет четыре последовательных действия:

определение целей, вариантов и ограничений (планирование);

анализ вариантов и распознавание или выбор рисков (анализ риска);

разработка продукта следующего уровня (конструирование);

оценка заказчиком текущих результатов конструирования (оценивание).

Нельзя обрывать основную фразу перед нумерованными перечислениями на предлогах и союзах: из, на, от, что, как и т. д.

Математические знаки применяются только в формулах. В тексте их пишут словами.

Пример: ...количество объектов равно 30.

Исключение составляют знаки (+) и (–) в сопровождении цифр.

Например, температура изменяется от – 5°C до + 25°C.

Знаки: °, №, %, >, ln и т.д. применяются только при цифровых или буквенных величинах. Знаки №, % для обозначения множественного числа удваивать не следует.

Пример: Рисунки № 3,4 и 8.

Числа с размерностью пишутся только числами. *Например,* Диаметр 25 миллиметров. Числа до десяти без размерностей или единиц измерения пишутся в тексте словами, свыше десяти – цифрами. Дроби пишутся всегда цифрами, например, 1/2; 3,25.

Количественные числительные, обозначаемые цифрами, пишутся в буквенно-цифровой форме, *например,* 25 млн; 150 тыс.; 3 млрд.

При указании пределов измерения значений величин их приводят один раз, *например,* 35–40 мм; от 1 до 5 м; 7,2 × 3,4 мм (а не 7,2 мм × 3,4 мм).

4.5 Сокращения и условные обозначения

В тексте пояснительной записки все слова, как правило, должны быть написаны полностью. Правила сокращений слов и словосочетаний устанавливаются ГОСТ Р 7.0.12–2011. Библиографическая запись. Сокращение слов и словосочетаний на русском языке. Общие требования и правила [5]. В таблице 1 отражены основные сокращения, применяемые при написании пояснительной записки.

Не допускается применять индексы стандартов (т. е. ГОСТ), технических условий (т. е. ТУ) и других документов без регистрационного номера.

Допускается употребление без расшифровки только сокращений, понятных читателю: ЭВМ, UML, GDI, ЭДС, КПД и т. п.

Другие сокращения должны быть расшифрованы при первом упоминании текста (в последующем тексте принятое сокращение пишется без скобок) или приводится в отдельном списке условных сокращений.

Пример: Создание автоматизированного рабочего места (АРМ) призвано...

Форма сокращений по всей работе должна быть одинакова.

Таблица 1 – Перечень допускаемых и не допускаемых сокращений

Допускается сокращать	Не допускаются сокращения
т. е. – то есть и т. д. – и так далее и т. п. – и тому подобное (после перечисления) и др. – и другие и пр. – и прочие см. – смотри (при повторной ссылке) напр. – например в., вв., гг. – при датах г., д., обл., с. – при географических названиях гл., п., подп., разд., рис., с., см., ср., табл. – при ссылках млн, млрд, тыс., экз. – при числах в цифровой форме	т. о. – таким образом т. н. – так называемый т. к. – так как

Сокращенные названия учреждений, предприятий, марки изделий, аппаратов и материалов, состоящие из начальных букв слов, входящих в название, пишут прописными буквами без точек и кавычек. *Например*, УлГТУ – Ульяновский государственный технический университет.

4.6 Единицы измерения и размерности

В тексте пояснительной записки единицы измерения, размерности и обозначения должны соответствовать ГОСТ 8.417–2002 «Единицы величин» [8], технологическим стандартам и рекомендациям международных организаций: ИСО, МЭК, МОЗМ и др.

Для каждой физической величины применяется одно (основное) условное буквенное обозначение. При большом количестве физических величин можно использовать запасные обозначения.

Единицы измерения и размерности, употребляемые без числовых величин, пишут в тексте полностью словами. В таблицах, выводах, на чертежах и графиках, в расшифровке буквенных формул размерности – с сокращениями.

После условных буквенных обозначений единицы измерения пишутся полностью без сокращений, *например*, t микросекунд. Сложные

размерности пишут сокращенно при условных буквенных обозначениях.
Пример: а см/с².

4.7 Индексы буквенных обозначений

По ГОСТ Р 7.0.97-2016 [9] нижними (подстрочными) индексами могут быть при буквенных обозначениях:

- а) цифры, например, U_1 , P_3 ;
- б) строчные буквы русского, латинского и греческого алфавитов: R_a , L_k , $C_{вх}$, V_x , V_y , $U_{нач}$, $U_{вых}$.

Индексы, представляющие собой сокращение одного русского слова, пишутся без точки на конце как знака сокращения.

Если в состав индекса входит несколько цифр или букв, то они отделяются запятой. *Например:* $J_{k,a}$; $a_{1,2,3}$.

4.8 Математические формулы

По ГОСТ 7.32–2017 [6] формулы и уравнения следует выделять из текста в отдельную строку. Над и под каждой формулой или уравнением нужно оставить по пустой строке. Если уравнение не умещается в одну строку, то оно должно быть перенесено после знаков равенства, умножения, сложения, вычитания и знаков соотношения ($<$, $>$ и т. п.), причем этот знак в начале следующей строки повторяют. При переносе формулы на знаке, символизирующем операцию умножения, применяют знак «Х».

Переносить на другую строку допускается только самостоятельные члены формулы. Не допускается при переносе разделение показателей степени, выражений в скобках, дробей, а также выражений, относящихся к знакам корня, интеграла, суммы, логарифма, тригонометрических функций и т. п.

Все формулы нумеруются. Обычно нумерация сквозная. Номер проставляется арабскими цифрами в круглых скобках в крайнем правом положении на строке, *например*, (1). В многострочной формуле номер формулы ставят против последней строки.

Допускается нумерация формул в пределах раздела. В этом случае номер формулы состоит из номера раздела и порядкового номера внутри раздела, разделенных точкой, *например*, (3.1).

Формулы в приложениях имеют отдельную нумерацию в пределах каждого приложения с добавлением впереди обозначения приложения, *например*, (B.2).

Если нужны пояснения к символам и коэффициентам с расшифровкой их размерностей, то они приводятся сразу под формулой в той же последовательности, в которой они идут в формуле. Перечень располагают с новой строки после слова «где» в виде колонки; символ отделяют от его расшифровки знаком тире. После расшифровки каждого символа ставят точку с запятой, размерность буквенного обозначения отделяют от текста запятой.

Примеры:

1) Система есть множество вещей, свойств и отношений:

$$S = (\{m\}, \{n\}, \{r\}), \quad (1)$$

m – вещи; n – свойства; r – отношения.

2) В формулах точка или знак умножения не ставится перед буквенным символом, после скобки и перед скобкой. *Например*,

$$2n \left(\frac{m+n}{r} \right) \left(\frac{r+u}{n-m} \right).$$

3) Перед числом, выраженным цифрами, а также между дробями ставится точка или знак умножения. *Например*,

$$x * 2.5; \frac{a+b}{c} * 30; 5.2 * \frac{c+d}{a}; \frac{7v * 3a}{3b * 5d}$$

В пределах текста пояснительной записки нельзя обозначать одинаковыми буквенными символами разные понятия и разными символами одинаковые понятия.

4.9 Таблицы и выводы

Материал может быть оформлен в виде таблиц и выводов, помещаемых в тексте пояснительной записки. *Таблицей* называют цифровой и текстовый материал, сгруппированный в определенном

порядке в горизонтальные строки и вертикальные графы (столбцы), разделенные линиями. Небольшой и несложный цифровой материал дается текстом, цифровые данные располагаются в виде колонок, называемых *выводами*.

По ГОСТ 7.32–2017 [6] на все таблицы в тексте должны быть ссылки. Таблица должна располагаться непосредственно после текста, в котором она упоминается впервые, или на следующей странице.

Все таблицы нумеруются (нумерация сквозная, либо в пределах раздела – в последнем случае номер таблицы состоит из номера раздела и порядкового номера внутри раздела, разделенных точкой (*например*, Таблица 1.2)). Таблицы каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением впереди обозначения приложения (*например*, Таблица В.2).

Слово «Таблица» пишется полностью. Наличие у таблицы собственного названия по рассматриваемому стандарту не обязательно, но в учебной документации название таблиц требуется всегда. Название таблицы следует помещать над таблицей слева, без абзацного отступа в одну строку с ее номером через тире (*например*, Таблица 3 – Доходы фирмы). Точка в конце названия не ставится.

Форма таблицы и все линии в ней выполняются тонкими линиями одинаковой толщины. Заголовок таблицы отделяется линией от остальной части таблицы.

Заголовки граф таблицы начинают с прописных букв, а подзаголовки со строчных, если они составляют одно предложение с заголовком. В конце заголовков и подзаголовков знаки препинания не ставятся. Заголовки указываются в единственном числе. Графа № п/п без необходимости в таблицу не включается. Повторяющийся в графе таблицы текст, состоящий из одного слова, допускается заменять кавычками, если строки в таблице не разделены линиями. Если повторяющийся текст состоит из двух и более слов, то при первом повторении его заменяют словами «то же», а далее кавычками. Графы диагональными линиями не разделяются. Ставить кавычки вместо повторяющихся цифр, математических знаков, марок

материала и других символов не допускается. Если цифровые данные в таблице не приводятся, то в графе ставят прочерк (тире).

Таблица 2 – Показатели уровня квалификации разработчиков подсистемы

Показатель	Описание	Вес, от –1 до 1	Значение, от 0 до 5	Значение с учетом веса
F1	Знакомство с технологией	1,5	4	6
F2	Опыт разработки приложений	0,5	3	1,5
F3	Опыт использования объектно-ориентированного подхода	1	4	4
F4	Наличие ведущего аналитика	0,5	0	0
F5	Мотивация	1	5	5
F6	Стабильность требований	2	4	8
F7	Частичная занятость	–1	3	–3
F8	Сложные языки программирования	–1	3	–3
			Сумма	18,5

При переносе таблицы на следующую страницу название помещают только над первой частью, при этом верхнюю горизонтальную черту, ограничивающую последующие части таблицы, не проводят.

Таблица 3 иллюстрирует *пример* переноса на следующую страницу.

Таблица 3 – Реквизиты документа «Путевой лист»

Реквизит	Тип значения (длина)
Номер	Строка (9)
Дата	Дата
Машина	СправочникСсылка.МашиныАвтопарка
МаркаГСМ	СправочникСсылка.МаркиГСМ

Продолжение таблицы 3

Водитель	СправочникСсылка.СотрудникиОрганизаций
Пробег	Число(10)
НормаРасхода	Число(10)
КоличествоИзрасходованногоГСМ	Число(10)

Шапка таблицы также не повторяется при переносе. Над другими частями также слева пишут слово «Продолжение» и указывают номер таблицы (*например*, Продолжение таблицы 1).

Если цифровые данные в графах таблицы имеют различную размерность, то ее указывают в наименованиях каждой графы или строки. Если параметры имеют одну размерность, то сокращенное обозначение единиц измерения помещают под заголовком таблицы.

Вывод приводят без заголовка, если он является непосредственным продолжением излагаемого материала и грамматически связан с вводной фразой текста, и с заголовком, если вывод имеет самостоятельное значение. Выводы не нумеруются.

Примечания и сноски к таблицам и выводам пишутся непосредственно под ними. Сноски к цифрам и в таблицах, и выводах обозначают только звездочками, до четырех. Нумерация сносок отдельная для каждой страницы текста.

Пример

Основные технические характеристики монитора.....	CPD-200GST
Максимальное разрешение, пикселов.....	1280*1024
Частота горизонтальной развертки, кГц.....	31,5–64,0
Частота вертикальной развертки, Гц.....	60–85

4.10 Иллюстрации

По ГОСТ 7.32–2017 [6] на все иллюстрации (рисунки, чертежи, графики, схемы, компьютерные распечатки, диаграммы и т. п.) в тексте должны быть даны ссылки. Рисунки должны располагаться

непосредственно после текста, в котором они упоминаются впервые, или на следующей странице.

Графики с результатами экспериментов следует выполнять с сеткой, но без стрелок. Сетка графика определяется масштабом шкал (равномерных или логарифмических) осей координат. Сетка не приводится на графиках, поясняющих только характер изменения функции. На осях графиков указывают наименование и единицу величины, числовые значения которых помещены у делений шкалы на осях. Если на рисунке имеется несколько графиков, то они вычерчиваются разными линиями (непрерывной, штриховой и т. д.), или разными цветами, или около линий ставят порядковые номера с последующей расшифровкой.

Иллюстрации, за исключением иллюстраций приложений, нумеруются арабскими цифрами, при этом нумерация сквозная, но допускается нумеровать и в пределах раздела (главы). В последнем случае номер рисунка состоит из номера раздела и порядкового номера иллюстрации, разделенных точкой (*например*, Рисунок 1.1). Подпись к рисунку располагается под ним посередине строки. Слово «Рисунок» пишется полностью. По стандарту можно ограничиться только номером (т. е. оставить, *например*, подпись: Рисунок 2), но в учебной документации практически всегда требуется еще и название. В этом случае подпись должна выглядеть так: Рисунок 1 – Схема базы данных. Точка в конце названия не ставится.

Если в работе есть приложения, то рисунки каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением впереди обозначения приложения (*например*, Рисунок А.3).

При ссылках на иллюстрации следует писать «... в соответствии с рисунком 2» при сквозной нумерации и «... в соответствии с рисунком 1.2» при нумерации в пределах раздела.

Примеры иллюстраций приведены на рисунках 6, 7.

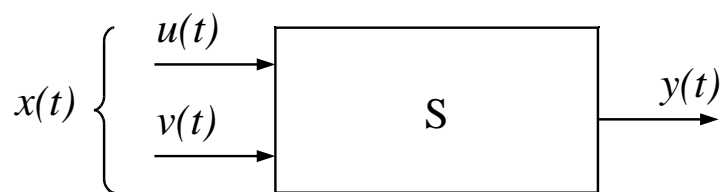


Рисунок 6 – Модель системы

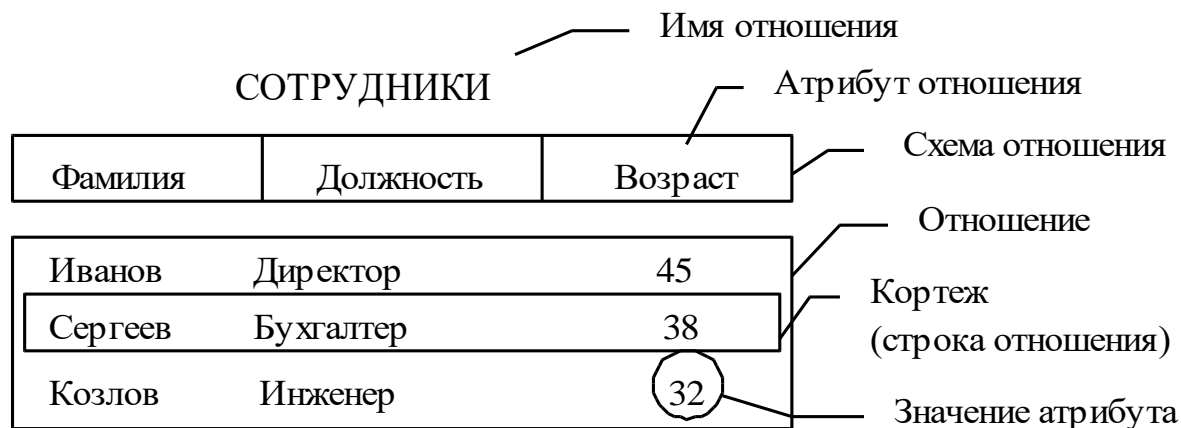


Рисунок 7 – Описание отношения в реляционной модели данных

4.11 Список литературы

Использованные в процессе работы специальные литературные источники указываются в конце пояснительной записки перед приложением. Список использованной литературы входит в основной объем работы. На более чем 80% литературных источников в тексте работы обязательно должна быть хотя бы одна ссылка.

Государственного стандарта по оформлению списка литературы нет, но существует общепринятая практика. *Например*, источники в списке литературы принято располагать в алфавитном порядке (относительно заголовка соответствующей источнику библиографической записи). При этом независимо от алфавитного порядка вначале обычно идут нормативные акты. Исходя из этого при составлении списка литературы следует придерживаться следующего порядка:

- 1) нормативные акты;
- 2) книги;
- 3) печатная периодика;
- 4) источники на электронных носителях локального доступа;

5) источники на электронных носителях удаленного доступа (т. е. Интернет-источники).

В каждом разделе сначала располагаются источники на русском языке, а потом – на иностранных языках (так же в алфавитном порядке).

Библиографический список необходимо оформлять в соответствии с требованиями составления ГОСТ Р 7.0.12–2011 «Библиографическая запись. Библиографическое описание. Общие требования и правила составления» [4]. По данному стандарту описание документа содержит ряд областей, в учебных целях используют следующие области:

- 1) область заглавия и сведений об ответственности (название и ФИО автора или редактора);
- 2) область издания (особенности данного издания по отношению к предыдущему изданию того же произведения);
- 3) область выходных данных (место издания, издательство, дата издания);
- 4) область физической характеристики (объем материала, размеры и пр.).

Области описания отделяются друг от друга точкой и тире (точка, пробел, тире, пробел). В конце библиографического описания ставится точка.

Библиографические сведения указывают в описании в том виде, в каком они даны в описываемом источнике информации (приводится в начале источника на 2–3 странице). Недостающие уточняющие сведения, а также полностью отсутствующие необходимые данные формулируют на основе анализа документа. При этом сведения, сформулированные на основе анализа документа, а также заимствованные из источников вне документа, во всех областях библиографического описания, кроме области примечания, приводят в квадратных скобках.

Для многотиражной литературы при составлении списка указываются: полное название источника, фамилия и инициалы автора, издательство и год выпуска (для статьи – название издания и его номер).

Для законодательных актов необходимо указывать их полное название, принявший орган и дату принятия.

Под электронными ресурсами подразумеваются как собственно данные из Интернета, так и данные на конкретном «винчестере», CD, дискетах и т. п. Все такого рода данные считаются опубликованными. При указании адресов серверов сначала указывается название организации, которой принадлежит сервер, а затем его полный адрес (см. п. 8, 9, 10, 11 из примера списка литературы).

Описание книги одного автора

Ярушкина, Н. Г. Основы теории нечетких и гибридных систем : учебное пособие / Н. Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.: ил.

Описание книги трех авторов

Ярушкина, Н. Г. Интеллектуальный анализ временных рядов: учебное пособие / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева – Ульяновск : УлГТУ, 2010. – 320 с.: ил.

Описание книги пяти и более авторов

Актуальные проблемы управления образования в регионе / Г. Н. Сериков, В. К. Кузнецов, И. Н. Розанов и др. – Челябинск, 2003.

Описание книги под редакцией

Нечеткие гибридные системы. Теория и практика / И. З. Батыршин, А. О. Недосекин, А. А. Стецко и др.; под ред. Н. Г. Ярушкиной. – М. : Физматлит, 2007. – (Информационные и компьютерные технологии). – 207 с.: ил.

Описание методических указаний

Операционные оболочки и системы Windows X. XX : метод. указания / сост. Ярушкина Н. Г. – Ульяновск : УлГТУ, 1996. - 36с.

Описание учебного пособия

Меркулова, Т. А. Программирование на языках высокого уровня с использованием прерываний MS-DOS : учебное пособие для студентов специальности 071900 «Информационные системы в экономике» /

Т. А. Меркулова, Н. Г. Ярушкина; Ульянов. гос. техн. ун-т, Каф. «Информ. Системы». – 2-е изд. – Ульяновск : УлГТУ, 1998. – 135 с.

Описание статьи из сборника, книги

Ярушкина, Н. Г. Интегральный метод нечеткого моделирования и анализа нечетких тенденций / Н. Г. Ярушкина, Т. В. Афанасьева, И. Г. Перфильева // Интеллектуальный анализ временных рядов : сб. науч. тр. семинара с междунар. участием «Интеллектуальный анализ временных рядов», 15 июня. – Ульяновск : УлГТУ, 2010. – С. 110-120.

Описание статьи из журнала

Ярушкина, Н. Г. Структура компонентно-ориентированной системы для анализа экономического состояния предприятия / Н. Г. Ярушкина // Прикладная информатика. – 2009. – № 2 (20). – С. 18-24.

Описание стандартов

ГОСТ Р 6.30–2003. Требования к оформлению документов. – Взамен ГОСТ 6.30–97 : введ. 2003–07–01. – М. : Изд-во стандартов, 2003. – 16 с.

Описание электронного ресурса

Российская государственная библиотека [Электронный ресурс] / Центр информ. технологий РГБ ; ред. Власенко Т.В. ; Web-мастер Козлова Н.В. – Электрон. дан. – М. : Рос. гос. б-ка, 1997- . – Режим доступа: <http://www.rsl.ru>, свободный. – Загл. с экрана. – Яз. рус., англ.

Пример списка литературы:

.....

3. Приказ от 26.12.94 №170 Положение о бухгалтерском учете и отчетности в Российской Федерации, приказ минфина рф №170 от 26.12.94.

4. Гиппиус, З. Н. Сочинения [Текст] : в 2 т. / Зинаида Гиппиус ; [вступ. ст., подгот. текста и коммент. Т. Г. Юрченко ; Рос. акад. наук, Ин-т науч. информ. по обществ. наукам]. – М. : Лаком-книга : Габестро, 2001.

5. Разумовский, В. А. Управление маркетинговыми исследованиями в регионе [Текст] / В. А. Разумовский, Д. А. Андреев ; Ин-т экономики города. – М., 2002. – 210 с. – Деп. в ИНИОН Рос. акад. наук 15.02.02, №139876.

6. Джапарова, Р. Н. К вопросу о сущности и содержании маркетинга / Р. Н. Джапарова // Вестник КРСУ. – 2002.– №3.

7. Информационные системы в экономике : учебник/под ред. проф. В.В. Дика. – М.: Финансы и статистика, 1996. – 272 с.

8. Oxford interactive encyclopedia [Электронный ресурс]. – Электрон, дан. и прогр. – [Б. м.] : The Learning Company, 1997. – 1 электрон, опт. диск (CD-ROM) : зв., цв. ; 12 см. – Систем, требования: ПК с процессором 486 + ; Windows 95 или Windows 3.1 ; дисковод CD-ROM ; зв. карта. – Загл. с этикетки диска.

9. Российский сводный каталог по НТЛ [Электронный ресурс] : база данных содержит сведения о зарубеж. и отечеств. кн. и зарубеж. период. изд. по естеств. наукам, технике, сел. хоз-ву и медицине, поступившие в организации-участницы Автоматизированной системы Рос. свод. кат. по науч.-техн. лит. : ежегод. пополнение ок. 30 тыс. записей по всем видам изд.—Электрон. дан. (3 файла).—М., [199—].—Режим доступа: <http://www.gpntb.ru/win/search/help/rsk.html>. — Загл. с экрана.

10. Электронный каталог ГПНТБ России [Электронный ресурс]: база данных содержит сведения о всех видах лит., поступающей в фонд ГПНТБ России.—Электрон. дан. (5 файлов, 178 тыс. записей).—М., [199—].—Режим доступа: <http://www.gpntb.ru/win/search/help/el-cat.html>. — Загл. с экрана.

11. Электронный каталог Большая советская энциклопедия [Электронный ресурс]: 30 томов (24-й том в двух книгах). —Электрон. дан. (Всего записей — 95280).—М., [1969-1978].—Режим доступа: <http://slovari.yandex.ru/~книги/БСЭ/>. — Загл. с экрана.

12.

При ссылке на литературу в тексте приводится порядковый номер источника, заключенный в квадратные скобки, и номер страницы, на которой содержится используемый из данного источника материал. В случае дословного цитирования цитата заключается в кавычки (*например*: «программное обеспечение – это совокупность программ

системы обработки данных и программных документов, необходимых для эксплуатации этих программ» [9, с.18]).

При использовании или описании мнений, суждений других авторов в своей работе необходимо также указывать номер литературного источника и номер страницы, где излагаются используемые материалы, (*например*, Иванов И.И. под программным обеспечением понимает совокупность программ системы обработки данных и программных документов, необходимых для использования данных программ [9, с.18]).

4.12 Приложения

Приложения оформляются как продолжение пояснительной записки на последующих страницах, но в основной листаж не включаются. Содержание приложений определяется студентом по согласованию с научным руководителем. При этом в основном тексте работы целесообразно оставить только тот иллюстративный материал, который позволяет непосредственно раскрыть содержание излагаемой темы. Вспомогательный же материал выносится в приложения. Объем приложений не ограничивается, поэтому основной листаж можно регулировать за счет переноса иллюстративного материала в приложения или из приложений.

По ГОСТ 7.32–2017 [6] в тексте работы на все приложения должны быть даны ссылки. Приложения располагают в порядке ссылок на них в тексте.

Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова «ПРИЛОЖЕНИЕ» и его обозначения. Приложение должно иметь заголовок, который записывают симметрично относительно текста с прописной буквы отдельной строкой.

Приложения обозначают заглавными буквами русского алфавита, начиная с А, за исключением букв Ё, З, Й, О, Ч, Ъ, Ы, Ь. После слова «Приложение» следует буква, обозначающая его последовательность (*например*: ПРИЛОЖЕНИЕ Б). Допускается обозначение приложений буквами латинского алфавита, за исключением букв I и O. В случае полного

использования букв русского и латинского алфавитов допускается обозначать приложения арабскими цифрами. Если в документе одно приложение, оно обозначается «ПРИЛОЖЕНИЕ А».

Текст каждого приложения может быть разделен на разделы, подразделы и т. д., которые нумеруют в пределах каждого приложения. Перед номером ставится обозначение этого приложения.

Нумерация страниц приложений и основного текста должна быть сквозная.

Все приложения должны быть перечислены в содержании документа (при наличии) с указанием их номеров и заголовков.

4.13 Исходный код программы (листинг)

Действующих государственных стандартов на оформление исходного кода программ нет. Но при этом существует огромное количество рекомендаций, стандартов кодирования для каждого языка программирования.

Таким образом, правила оформления исходного кода (листинга) на бумажном носителе необходимо смотреть в нотациях, стандартах конкретного языка программирования.

5 Нормоконтроль

Каждая магистерская диссертация проходит нормоконтроль, проводимый с целью проверки выполнения установленных норм, требований и правил оформления в выпускной квалификационной работы.

На нормоконтроль сдается выверенный с точки зрения содержания, стиля, грамотности окончательный вариант магистерской диссертации с подписью самого обучающегося, руководителя и рецензента (опционально).

Содержание нормоконтроля выпускной квалификационной работы, проводимого на кафедре Информационные системы приведено в таблице 4.
Таблица 4 – Содержание нормоконтроля ВКР

Наименование	Требования
Пояснительная записка	<ul style="list-style-type: none">• соответствие темы работы теме, утвержденной приказом;• правильное обозначение направления обучения;• правильное оформление обозначения документа;• правильное оформление даты и номера приказа о закреплении темы ВКР;• соответствие структуры пояснительной записки ВКР структуре, заданной в данных методических указаниях (наличие титульных листов, листа задания, аннотации, списка литературы и т.д.), наименование разделов должны быть определены;• наличие рамок в пояснительной записке, правильность заполнения основной рамки (большая рамка ТОЛЬКО на первой странице содержания);• наличие последнего листа ПЗ (без рамки и нумерации), в общую нумерацию не входящего;• соблюдение требований стандартов на текстовые документы («Times New Roman», размер шрифта – 14, печать через 1,5 интервала, цвет шрифта – черный); При проверке распечатанной ПЗ: <ul style="list-style-type: none">• комплектность документации в соответствии с требованиями, установленными на выпускную квалификационную работу (пояснительная записка, отзыв руководителя, отзыв рецензента (опционально) антиплагиат, заявление о самостоятельной работе, диск);• заполнение всех обязательных строк на титульных листах (подписи и даты от студента, руководителя, рецензента)

<p>Главы</p>	<ul style="list-style-type: none"> • только главы начинаются с новой страницы; • номер подраздела состоит из номеров раздела и подраздела в разделе, разделенных точкой; • в конце номера точка не ставится; • переносы слов в заголовках не допускаются.
<p>Формулы</p>	<ul style="list-style-type: none"> • формулы и уравнения следует выделять из текста в отдельную строку; • над и под каждой формулой или уравнением нужно оставить по пустой строке; • все формулы нумеруются; • номер проставляется арабскими цифрами в круглых скобках в крайнем правом положении на строке, • в многострочной формуле номер формулы ставят против последней строки; • формулы в приложениях имеют отдельную нумерацию в пределах каждого приложения с добавлением впереди обозначения приложения;
<p>Таблицы</p>	<ul style="list-style-type: none"> • на все таблицы в тексте должны быть ссылки; • таблица должна располагаться непосредственно после текста, в котором она упоминается впервые, или на следующей странице; • таблицы каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением впереди обозначения приложения; • название таблицы следует помещать над таблицей слева, слово «Таблица» пишется полностью, без абзацного отступа в одну строку с ее номером через тире, точку в конце названия не ставить; • при переносе таблицы на следующую страницу название помещают только над первой частью, при этом верхнюю горизонтальную черту, ограничивающую последующие части таблицы, не проводят; • при переносе таблицы шапка таблицы также не повторяется; • при переносе над другими частями также слева пишут слово «Продолжение» и указывают номер таблицы.
<p>Рисунки</p>	<ul style="list-style-type: none"> • на все иллюстрации в тексте должны быть даны ссылки; • рисунки должны располагаться непосредственно после текста, в котором они упоминаются впервые, или на следующей странице; • подпись к рисунку располагается под ним посередине строки, слово «Рисунок» пишется полностью в одну строку с ее номером через тире, точку в конце названия не ставить; • рисунки каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением впереди обозначения приложения.

Источники литературы	<ul style="list-style-type: none"> • наличие списка литературы; • количество источников не менее 30; • ссылки в тексте обязательны на более 80% источников, правильность оформления списка литературы и ссылок.
Приложения	<ul style="list-style-type: none"> • наличие приложения (листинг обязателен); • в тексте обязательны ссылки на каждое приложение; • Каждое приложение следует начинать с новой страницы с указанием сверху посередине страницы слова «ПРИЛОЖЕНИЕ» и его обозначения; • приложения обозначают заглавными буквами русского алфавита, начиная с А, за исключением букв Ё, З, Й, О, Ч, Ъ, Ы, Ь (либо буквами латинского алфавита, за исключением букв I и O); • если в документе одно приложение, оно обозначается «ПРИЛОЖЕНИЕ А».

Нормоконтролер проверяет работу по всем показателям и подписывает в месте, отведенном для подписи нормоконтролера.

Пояснительные записки, не исправленные в срок, не допускаются до защиты выпускной квалификационной работы, либо может быть допущена с понижением итоговой оценки на 1 балл.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры). Утвержден приказом Министерства образования и науки Российской Федерации от 30 октября 2014 г. № 1404. Москва, 2014.

2. Федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.04.04 Программная инженерия (уровень магистратуры). Утвержден приказом Министерства образования и науки Российской Федерации 30 октября 2014 г. № 1406. Москва, 2014.

3. СМК-ДП-7.5-03 Организация итоговой государственной аттестации выпускников, УлГТУ. – 2006 г.

4. ГОСТ Р 7.0.100-2018. Библиографическая запись. Библиографическое описание. Общие требования и правила составления = System of standards on information, librarianship and publishing. Bibliographic record. Bibliographic description. General requirements and rules : национальный стандарт Российской Федерации : издание официальное : утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 3 декабря 2018 г. N 1050-ст : введен впервые : дата введения 2019-07-01 / разработан Федеральным государственным унитарным предприятием "Информационное телеграфное агентство России (ИТАР-ТАСС)", филиал "Российская книжная палата", Федеральным государственным бюджетным учреждением "Российская государственная библиотека", Федеральным государственным бюджетным учреждением "Российская национальная библиотека". – Москва : Стандартинформ, 2018. – 73 с. – Текст : непосредственный.

5. ГОСТ Р 7.0.12–2011. Библиографическая запись. Сокращение слов и словосочетаний на русском языке. Общие требования и правила = System of standards on information, librarianship and publishing. Bibliographic record. Abbreviation of words and word combinations in Russian. General requirements and rules rules : национальный стандарт Российской Федерации : издание

официальное : утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 13 декабря 2011 г. N 813-ст : введен впервые : дата введения 2012-09-01 / подготовлен Федеральным государственным бюджетным учреждением науки "Российская книжная палата" (РКП). – Москва : Стандартинформ, 2020. – 32 с. – Текст : непосредственный.

6. ГОСТ 7.32–2017. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления = System of standards on information, librarianship and publishing. The research report. Structure and rules of presentation : межгосударственный стандарт : издание официальное : утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 24 октября 2017 г. No 1494-ст : взамен ГОСТ 7.32—2001 : дата введения 2018-07-01 / разработан Федеральным государственным бюджетным учреждением науки "Всероссийский институт научной и технической информации Российской академии наук" в рамках Технического комитета по стандартизации ТК 191 "Научно-техническая информация, библиотечное и издательское дело" – Москва : Стандартинформ, 2018. – 35 с. – Текст : непосредственный.

8. ГОСТ Р 2.105-2019. Единая система конструкторской документации. Общие требования к текстовым документам = Unified system for design documentation. General requirements for textual documents : национальный стандарт Российской Федерации : издание официальное : утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 29 апреля 2019 г. N 175-ст : введен впервые : дата введения 2020-02-01 / разработан Федеральным государственным унитарным предприятием "Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия" (ФГУП "СТАНДАРТИНФОРМ") – Москва : Стандартинформ, 2021. – 44 с. – Текст : непосредственный.

9. ГОСТ 8.417–2002. Государственная система обеспечения единства измерений. Единицы величин = State system for ensuring the uniformity of

measurements. Units of quantities : межгосударственный стандарт : издание официальное : утвержден и введен в действие Постановлением Государственного комитета Российской Федерации по стандартизации и метрологии от 4 февраля 2003 г. N 38-ст : взамен ГОСТ 8.417-81 : дата введения 2003-09-01 / разработан Федеральным государственным унитарным предприятием "Всероссийский научно-исследовательский институт метрологии им.Д.И.Менделеева" (ФГУП "ВНИИМ им.Д.И.Менделеева"), Техническим комитетом по стандартизации ТК 206 "Эталоны и поверочные схемы" – Москва : Стандартинформ, 2018. – 34 с. – Текст : непосредственный.

10. ГОСТ Р 7.0.97-2016. Система стандартов по информации, библиотечному и издательскому делу. Организационно-распорядительная документация. Требования к оформлению документов = System of standards on information, librarianship and publishing. Organizational and administrative documentation. Requirements for presentation of records : национальный стандарт Российской Федерации : издание официальное : утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 8 декабря 2016 г. N 2004-ст : взамен ГОСТ Р 6.30-2003 : дата введения 2018-07-01 / разработан Федеральным бюджетным учреждением "Всероссийский научно-исследовательский институт документоведения и архивного дела" (ВНИИДАД) Федерального архивного агентства – Москва : Стандартинформ, 2019. – 32 с. – Текст : непосредственный.

11. Методические указания по выполнению дипломного проектирования по специальности 08080165 «Прикладная информатика (в экономике)» / сост.: О. Н. Евсеева, А. М. Наместников, Е. В. Суркова. – Ульяновск : УлГТУ, 2008. – 29 с.: ил.

12. Правила оформления рукописей для издания в УлГТУ. Основные положения / Федер. агентство по образованию, Гос. образовательное учреждение высш. проф. образования Ульян. гос. техн. ун-т ; сост. М. В. Теленкова. – 3-е изд., испр. и доп. – Ульяновск: УлГТУ, 2009. – 46 с.

13. Электронный каталог Большая советская энциклопедия [Электронный ресурс]: 30 томов (24-й том в двух книгах). —Электрон. дан. (Всего записей — 95280).—М., [1969-1978].—Режим доступа: <http://slovari.yandex.ru/~книги/БСЭ/>. — Загл. с экрана.

Перечень графического материала (с точным указанием обязательных чертежей) _____

Календарный график работы над ВКР на весь период (с указанием сроков выполнения и содержания отдельных этапов)

№ этапа	Содержание этапа	Срок выполнения

Дата выдачи задания « ___ » _____ 20__ г.

Руководитель _____ / _____ /
должность, учёная степень, ученое звание подпись инициалы, фамилия

Задание принял к исполнению _____ / _____ /
подпись обучающегося инициалы, фамилия

Замечания к ВКР _____

Рекомендации по внедрению _____

Рекомендуемая оценка ВКР _____

Дополнительная информация для ГЭК _____

вид ВКР: комплексный; характер проекта: фундаментальный, прикладной, методологический, поисковый;

получение по проблеме грантов; публикации; выступления на конференциях и полученные награды;

рекомендация на конкурс или выставку ВКР, к продолжению исследований, дальнейшему обучению в

аспирантуре или работе на кафедре и др.

Руководитель _____ / _____ /

должность, учёная степень, ученое звание

подпись

инициалы, фамилия

« _____ » _____ 20 _____ г.

ПРИЛОЖЕНИЕ Д

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, _____,
Ф.И.О.

Обучающийся группы _____, направления подготовки/специальности

_____,
заявляю, что моя выпускная квалификационная работа на тему _____

_____,
представленная в государственную экзаменационную комиссию для публичной защиты, не имеет элементов плагиата.

Все прямые заимствования из печатных и электронных источников имеют соответствующие ссылки.

С Положением о проверке самостоятельности выполнения выпускных квалификационных работ с использованием системы «Антиплагиат», в соответствии с которым обнаружение неправомерных заимствований является основанием для недопуска выпускной квалификационной работы до защиты, ознакомлен.

«__» _____ 20__ г. _____ / _____ /
дата подпись Ф.И.О.

ПРИЛОЖЕНИЕ Е

Магистерская диссертация выполнена мной совершенно самостоятельно. Все использованные в работе материалы и концепции из опубликованной научной литературы и других источников имеют ссылки на них.

Объем работы ___ листов.

Библиография ___ наименования.

Объем приложений ___ листов.

Отпечатано в ___ экземпляре.

Один экземпляр сдан на кафедру.

«__» _____ 20__ г.

_____/_____/_____
(подпись)

Учебное издание

ОФОРМЛЕНИЕ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ
МАГИСТРА
Методические указания

Составители КОРУНОВА Надежда Владимировна,
ЭГОВ Евгений Николаевич

Редактор Н.А. Евдокимова
Подписано в печать _____.
Усл. печ. л. _____.

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.

ПЕДАГОГИКА И ПСИХОЛОГИЯ ВЫСШЕЙ ШКОЛЫ

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ к проведению лекционных занятий

в соответствии с учебным
планом по направлению
подготовки (специальности)

09.04.03 Прикладная информатика

профиль
(программа / специализация)

Искусственный интеллект и бизнес-аналитика

Основные вопросы, освещаемые на лекциях

Раздел, тема учебной дисциплины (модуля), содержание темы
Раздел 1. Психология высшей школы
Тема 1. Предмет и задачи курса «психология высшей школы». Предмет, цели, задачи, функции психологии высшей школы. Место дисциплины в системе наук. Становление и перспективы развития.
Тема 2. Особенности развития личности студента. Психолого-педагогические особенности одаренных студентов. Социализация личности студента. Адаптация личности студента, ее трудности и последствия. Адаптация к учебной деятельности в вузе.
Тема 3. Профессиональное становление. Факторы профессионального становления. Противоречия профессионального становления. Стадии и кризисы профессионального становления.
Тема 4. Лидерство в организации. Феномен лидерства. Психологическое содержание понятия «лидерство». Стили лидерства. Лидерство и руководство. Гендерные аспекты организационного руководства и лидерства. Методика формирования команды. Организация межличностных, групповых и организационных коммуникаций.
Раздел 2. Педагогика высшей школы
Тема 5. Педагогика высшей школы. Предмет, задачи, категории педагогики высшей школы. Принципы и методы педагогического исследования.
Тема 6. Приоритетные стратегии и тенденции развития высшего образования. Современные стратегии модернизации высшего образования в России и за рубежом.
Тема 7, 8. Формы организации обучения в вузе: традиции и инновации. Трехмерная модель систематики форм организации обучения. Вузовская лекция. Игры. Семинары и конференции. Самостоятельная работа студентов. Проектно-творческая деятельность. Дистанционное обучение. Авторские технологии обучения. Научно-исследовательская работа студентов. УИР как часть профессиональной подготовки студентов. Формы организации НИР в вузе. Защита интеллектуальной собственности.

Раздел 1. Психология высшей школы

Тема 1. Предмет и задачи курса «Психология высшей школы».

1.1. Предмет, цели, задачи, функции психологии высшей школы.

1.2. Место дисциплины в системе наук.

1.3. Становление и перспективы развития.

1.1. Предмет, цели, задачи, функции психологии высшей школы.

Концепция высшего образования в современной России предполагает, что человек, гражданин должен стремиться повышать свой образовательный уровень постоянно, хорошо разбираться в различных профессиональных сферах, людях, их отношениях. Самообразование и мотивация на достижение успеха - вот главные ценности профессионала нынешнего времени. Трудно выстраивать комфортные взаимоотношения с людьми в повседневной жизни, а также налаживать профессиональные коммуникативные каналы с коллегами, клиентами или партнерами, не обладая минимумом научно достоверных сведений по психологии, а руководствуясь лишь житейскими, обывательскими представлениями.

Предметом психологии высшей школы выступают психологические закономерности и условия эффективности процессов обучения и воспитания в высшей школе. В этом совокупном предмете можно выделить ряд частных предметов изучения и отдельных проблем: психологическая структура учебной деятельности в высшей школе; формирование и функционирование познавательных процессов учащихся (профессионального восприятия, мышления, памяти, внимания), профессиональных способностей, черт личности, умений и навыков; возрастные психические и психофизиологические особенности студентов; дифференциально-психологические характеристики учащихся, которые необходимо учитывать в процессе обучения и воспитания, и методы их диагностики; социально-психологические закономерности формирования студенческого и преподавательского коллектива; психологические аспекты педагогического общения; психологические конфликты в студенческой среде и между студентом и преподавателем (виды и способы их разрешения) и др.

Психология высшей школы так же изучает место и роль психических процессов, состояний, свойств, опыта, их проявление, развитие и функционирование в деятельности студентов, преподавателей и руководителей вузов.

Психологию высшей школы интересуют прежде всего функционирование, изменение, развитие и формирование психики студентов, психологические особенности деятельности преподавателей.

Цель изучения «Психология высшей школы»: овладение методологическими основами и инструментариями психологии высшей школы, повышение образованности молодых специалистов в вопросах психологии, а также формирование умений анализировать основные достижения и тенденции развития психологии высшей школы и использовать их в своей практической и профессиональной деятельности.

Основные задачи:

- ознакомление с основами психологической науки, ее возможностями в успешном решении проблем жизни и профессиональной деятельности, возникающих перед каждым человеком и человеческими общностями;
- достижение научного понимания студентами основ психологической и педагогической реальностей, их проявлений и влияний в жизни и деятельности людей;
- раскрытие роли и возможностей психологии в самореализации и самоутверждении человека;
- ознакомление с психологическими и педагогическими основами деятельности в условиях современного российского образовательного и профессионального пространства,

способствование развитию у них элементов государственного мышления и активной гражданской позиции;

- психологическая подготовка студентов к предстоящей профессиональной деятельности;
- содействие гуманитарному развитию студентов, их психологического мышления, наблюдательности, культуры их отношения к людям, общения и поведения;
- выработка навыков по использованию методов психологии в повышении студентами личной образованности, воспитанности, в освоении учебных программ, повышении профессионального мастерства, овладении различными психологическими техниками;
- формирование устойчивого интереса к продолжению работы по повышению своей психологической подготовленности в рамках повседневной и профессиональной деятельности.

Функции «Психологии высшей школы» как учебной дисциплины:

– Образовательно-мировоззренческая – постоянное расширение знаний студентов о человеке, об обществе, о себе, собственных мотивах, особенностях и ориентирах для будущей деятельности. Благодаря реализации этой функции общежитийский опыт дополняется и расширяется научно достоверными, систематизированными знаниями, которые помогают более четко сформировать взгляды на жизнь, убеждения, надежные опоры на жизненном пути;

– Воспитательно-мобилизующая – выражается в мощном вкладе, который вносят психология высшей школы в процесс гуманизации студента и преподавателя ВУЗа. Т.е. благодаря этому люди по-другому, глубже и обстоятельнее, начинают воспринимать других людей, студентов и преподавателей как своего, так и других ВУЗов. Также происходит более достоверная оценка себя, своих достоинств и недостатков, понимание новых возможностей и побуждения к самосовершенствованию. Формируется постоянное стремление к самовоспитанию, самообразованию и в конечном счете к самосовершенствованию.

– Жизненно-практическая – заключается в определённой установке на использование многих конкретных знаний, принципов, методик и рекомендаций психологии в своей жизни, в образовательном учреждении, в семье, среди людей, на досуге, в трудных ситуациях и пр.; в стремлении преподавателя не просто четко реализовывать свои профессиональные обязанности, но выстроить эффективные коммуникативные каналы со студенческим обществом и воспринимать каждого студента как значимую личность. Все это позволит значительно уменьшить число промахов, неудач, конфликтов, связанных с профессиональной деятельностью и повысить успешность при самоутверждении и самореализации;

– Профессионально-прикладная – связана с наполнением специалиста психологическими знаниями, навыками и умениями, нужными для профессиональной деятельности, так как любая такая деятельность связана с контактами в социальной сфере на разных уровнях;

– Развивающая – всё вышесказанное о функциях, уже свидетельствует о широком развивающем влиянии изучения психологии высшей школы, как на личность студента, так и на профессиональные навыки преподавателя. Усвоение знаний, выработка практики их применения в образовательном процессе неразрывно связано с элементами такого специфического качества, как психологическое и педагогическое мышление. Содержание, формы и методы преподавания учебной дисциплины предусматривают также целенаправленное профессиональное развитие отдельных качеств: наблюдательности, памяти, внимания и др.

Изучение психологии высшей школы строится эффективно, если реализуются все функции.

1.2. Место дисциплины в системе наук.

Связь психологии высшей школы с психологией очевидна. Психология - наука о закономерностях, механизме и фактах психической жизни человека. Психологи изучают устойчивое и повторяющееся в индивидуальном поведении. В центре внимания - проблемы восприятия, памяти, мышления, обучения и развития человеческой личности, усвоения нового знания на различных этапах возрастного развития. И если на начальных этапах взросления достаточно было возрастной психологии, то на этапе вузовского и поствузовского обучения центральную роль начинает играть психология ВУЗа. Именно изучение проблем обучения, воспитания, усвоения нового знания выработка новых навыков на основе этого знания и применение их в своей профессиональной деятельности в рамках ВУЗа позволило выделиться психологии высшей школы как самостоятельной дисциплине.

Так же психология высшей школы связана с социологией, как с учением о человеческом поведении, оформленном в групповую жизнь, включающую, с одной стороны, коллективные действия, мотивы и поступки, а с другой - способы, какими люди придают значение своему жизненному опыту (саморефлексия). Если говорить о процессе обучения в ВУЗе, то здесь важен не столько студент как отдельный индивид, но вся совокупность студенческих коллективов, их взаимосвязь как друг с другом, так и преподавательским коллективом. Именно студенческий коллектив и его взаимосвязь с преподавателями и является предметом рассмотрения Психологии высшей школы.

Трудно переоценить для психологии высшей школы значение тесных связей с философией и естествознанием. Философия обосновывает общие подходы к пониманию человека, раскрывает понятие личности студента, причины и цели его активности, уровни детерминации поведения. Философский анализ взаимосвязи человека и общества, лежащий в основе этических концепций, раскрывается психологией в правилах и нормах воспитания. Естествознание показывает, какие объективные параметры лежат в основе психолого-педагогических закономерностей, например, как связана динамика работы нервной системы с индивидуальными особенностями.

Психология высшей школы и педагогика - две науки, изучающие процесс обучения в целом, включая и Вузовское обучение, в неразрывном единстве. Такое тесное сотрудничество происходит в рамках действия такой перспективной отрасли научного знания как педагогическая психология. Она изучает условия и закономерности формирования психических процессов под воздействием образования и обучения. Педагогическая психология стала сферой совместного изучения взаимосвязей между воспитанием, обучением и развитием подрастающих поколений на различных этапах взросления (Б.Г. Ананьев). Например, одной из многих важнейших педагогических проблем является проблема несоответствия ожиданий преподавателей усвоения учениками (студентами) нового материала с теми профессиональными навыками, которыми они оперируют в своей профессиональной деятельности. Проще говоря, почему учебный материал не был усвоен или был усвоен неправильно. В связи с этой проблемой складывается предмет педагогической психологии, исследующей закономерности усвоения, учения. На основе сложившихся научных представлений формируются техника, практика учебно-педагогической деятельности, обоснованные со стороны психологии закономерностей процессов усвоения. Вторая педагогическая проблема возникает, когда осознается различие обучения и развития в образовательной системе. Предмет исследования в этом случае - закономерности развития интеллекта, личности, способностей, вообще человека. Данное направление педагогической психологии разрабатывает практику не обучения, а организации развития. Еще одной проблемой становится проблема взаимоотношения преподавателя и студента, а также роль педагогического коллектива в воспитании студента. Здесь на первый план выходят вопросы воспитательной деятельности ВУЗа, государственной политики образования, а также личная мотивация педагога и студента в рамках самовоспитания и саморазвития.

1.3. Становление и перспективы развития

Становление «Психологии высшей школы» как научной дисциплины неразрывно связано с генезисом таких наук как психология и педагогика. Становление педагогической научной мысли, неразрывно связана с деятельностью таких ученых как Яна Амоса Коменского, Жан-Жака Руссо, Иоганна Песталоцци, Иоганна Гербарта, Адольфа Дистервега, К.Д. Ушинского, П.Ф. Каптерева, чьи труды положили начало развитию педагогической теории и целенаправленной организации обучения как процесса. Сама же Психология высшей школы начала формироваться как самостоятельная наука в рамках педагогической психологии только в конце XIX в.

Первоначальный этап становления науки был посвящен таким вопросам как: связь развития, обучения и воспитания; творческая активность обучаемого, способности ученика и их развитие, роль личности преподавателя, организация обучения и многие другие. Однако это были только первые попытки научного осмысления этого процесса. Хотя уже и здесь в достаточной мере раскрываются такие проблемы как соотношение творчества, обучения и саморазвития (Песталоцци).

Так же анализируются вопросы методологического единства обучения и воспитания, роли преподавателя в этом процессе. Так Дистервегу принадлежит тезис о главенствующей роли педагога, учителя в образовательном процессе. Он рассматривает учебный процесс как единство ученика - обучаемого субъекта, учителя, изучаемого предмета и условий обучения.

Огромный вклад в разработку основ психологии высшей школы П.Ф. Каптерева. В его трудах были рассмотрены педагогические проблемы учительского труда и подготовки учителя, проблемы эстетического развития и воспитания и многие другие. Образовательный процесс, по мнению автора, представляет собой «выражение внутренней самодеятельности человеческого организма», развитие способностей и др.

Существенен вклад в становление педагогической психологии и психологии высшей школы С.Т. Шацкого (1878-1934), разработавшего целостную концепцию гуманизации и демократизации воспитания в процессе социализации человека. Ему принадлежит авторство модели педагога, в которой соединены обобщенные требования к его личности и профессиональной компетентности как к субъекту социально-педагогической деятельности.

Следующий этап (конец XIX в. - середина XX в.) проходит на фоне интенсивного развития экспериментальной психологии, создания и разработки конкретных педагогических систем.

Начало этого неразрывно связано с идеями, представленными в работах А.П. Нечаева, А. Бине и Б. Анри, М. Оффнера, Э. Меймана, В.А. Лайя, в исследованиях Г. Эббингауза, Ж. Пиаже, А. Валлона, Дж. Дьюи, С. Фрэнэ, Э. Клапереда. Экспериментальное изучение особенностей поведения научения (Дж. Уотсон, Э. Толмен, Э. Газри, К. Халл, Б. Скиннер), развития детской речи (Ж. Пиаже, Л.С. Выготский, П.П. Блонский, Ш. и К. Бюлер, В. Штерн и др.), развитие специальных педагогических систем Вальдорфской школы, школы М. Монтессори.

Особое значение имеет развитие, начиная с работ Ф. Гальтона, тестовой психологии, психодиагностики. Благодаря исследованиям А. Бине, Б. Анри, Т. Симона во Франции и Дж. Кэттелла в Америке это позволило найти действенный механизм (при взаимодействии тестов достижения и тестов способностей) не только контроля знаний и умений обучающихся, но и управления подготовкой учебных программ, учебным процессом в целом. Как отмечает М.В. Гамезо, в этот период в Европе образовался ряд лабораторий при школах. Так, в Германии возникла лаборатория Э. Меймана, в которой для решения учебных и воспитательных задач использовались приборы и методики, созданные в лабораториях университетов. В 1907 г. Мейман публикует книгу «Лекции по экспериментальной психологии», где дает обзор работ по экспериментальной дидактике. В Англии вопросами экспериментального изучения типологических особенностей

школьников занимался известный детский психолог Дж. Селли, который в 1898 г. опубликовал работу «Очерки по психологии детства». Во Франции А. Бине основал при одной из школ Парижа экспериментальную детскую лабораторию. В лаборатории изучались физические и душевные способности ребенка, а также методы преподавания учебных дисциплин.

О самостоятельности психологии высшей школы как науки, свидетельствует не только использование тестовой психодиагностики, исследовательских программ на основе ВУЗов, экспериментально-педагогических систем и программ, но и попытки научной рефлексии образовательного процесса, его строгого теоретического осмысления.

Основанием для выделения третьего этапа развития педагогической психологии и психологии высшей школы служит создание психологических теорий обучения, т.е. разработка теоретических основ педагогической психологии. Так, в 1954 г. Б. Скиннер выдвинул идею программированного обучения, а в 60-х годах Л.Н. Ланда сформулировал теорию его алгоритмизации. Затем В. Оконь, М.И. Махмутов построили целостную систему проблемного обучения. Это, с одной стороны, продолжило разработку системы Дж. Дьюи, полагавшего, что обучение должно идти через решение проблем, а с другой - соотносилось с положениями О. Зельца, К. Дункера, С.Л. Рубинштейна, А.М. Матюшкина и др. о проблемном характере мышления, его фазности, о природе возникновения каждой мысли в проблемной ситуации (П.Г. Блонский, С.Л. Рубинштейн). В 50-е годы появились первые публикации П.Я. Гальперина и затем Н.Ф. Талызиной, в которых излагались исходные позиции теории поэтапного формирования умственных действий, впитавшей в себя основные достижения и перспективы педагогической психологии. В это же время разрабатывается теория развивающего обучения, описанная в работах Д.Б. Эльконина, В.В. Давыдова на основе общей теории учебной деятельности (сформулированной этими же учеными и развиваемой А.К. Марковой, И.И. Ильясовым, Л.И. Айдаровой, В.В. Рубцовым и др.). Развивающее обучение нашло свое отражение и в экспериментальной системе Л.В. Занкова.

В этот же период С.Л. Рубинштейн в «Основах психологии» дал развернутую характеристику учения как усвоения знаний. Усвоение с разных позиций детально разрабатывалось далее Л.Б. Ительсоном, Е.Н. Кабановой-Меллер и др., а также в работах Н.А. Менчинской и Д.Н. Богоявленского (в рамках концепции экстерниоризации знаний). Появившаяся в 1970 г. книга И. Лингарта «Процесс и структура человеческого учения» и в 1986 г. книга И.И. Ильясова «Структура процесса учения» позволили сделать широкие теоретические обобщения в этой области.

Заслуживает внимания возникновение принципиально нового направления - суггестопедии, суггестологии Г.К. Лозанова (60-70-е годы). Его основой является управление педагогом неосознаваемыми обучающимися психическими процессами восприятия, памяти с использованием эффекта гипермнезии и суггестии. В дальнейшем был разработан метод активизации резервных возможностей личности (Г.А. Китайгородская), группового сплочения, групповой динамики в процессе такого обучения (А.В. Петровский, Л.А. Карпенко).

В настоящее время психология высшей школы стала неотъемлемой частью новой концепции непрерывного обучения, существующей в рамках Болонского процесса и активно внедряющегося в Российское образование. Именно изучение психологических основ поведения студентов, их мотивации и особенности взаимодействия с преподавателем способствует достижению основных целей непрерывного образования, стремления к постоянному самообучению и самосовершенствованию, а также усвоению основных компетенций и использованию навыков, полученных как в ВУЗе, так в процессе поствузовского образования.

Вопросы для повторения:

1. Каковы предмет, цели и задачи «Психологии высшей школы»?

2. С какими науками психология высшей школы связана наиболее тесно?
3. Какое влияние на становление и развитие психологии высшей школы оказало становление и развитие педагогической психологии?
4. Каковы перспективы развития психологии высшей школы в современном образовательном пространстве?
5. Почему необходимо изучать психологию высшей школы в ВУЗе?

Тема 2. Особенности развития личности студента.

- 2.1. Социализация личности студента.
- 2.2. Психолого-педагогические особенности одаренных студентов.
- 2.3. Адаптация личности студента, ее трудности и последствия. Адаптация к учебной деятельности в вузе.

2.1. Социализация личности студента.

Понятие «студент» с латинского языка на русский язык переводится как усердно работающий, занимающийся, т.е. овладевающий знаниями. При этом в самом понятии «студент» заключен смысл личностных качеств человека, его индивидуальность, желание трудиться и работать. Студент как индивид определенного возраста, прошедший пубертатный период, и как личность еще формирующаяся характеризуется с трех аспектов:

1) психологического, который помогает охарактеризовать совокупность психологических процессов, состояний и свойств личности индивида;

2) социального, где можно заметить анализ межличностных отношений, качества, появляющиеся вследствие принадлежности студента к определенной социальной группе, национальности и т.д.;

3) биологического, где находит свое применение учет типа высшей нервной деятельности, строения анализаторов, безусловных рефлексов, инстинктов, физической силы, телосложения, черт лица, цвета кожи, глаз, роста и т.д.

Изучение трех аспектов в совокупности раскрывает качества и возможности студента как индивида, его возрастные и личностные особенности. Все этапы личного развития в процессе появления статуса студента можно подразделить на 3 этапа, в каждом из которых существуют разные личностные установки и цели (таблица 1).

Таблица 1. Этапы развития личности студента

Этапы саморазвития	Критерии личностного развития в обучении	Принципы и личностные установки	Личностная цель	Зоны развития мыслительной деятельности
Первый: самопознание, самоутверждение	Активность, обучаемость	1. Что я умею и смогу делать? 2. Я должен понять и выучить программы	Выявление своих возможностей при активной учебной деятельности	Зона актуального развития
Второй: самосовершенствование, самовоспитание	Активность, обучаемость и рефлексивность	1. Приобрести уверенность в себе	Активное и уверенное проявление себя и	Зона ближайшего развития

		2. Я должен стать лучше	своих возможностей в учебной деятельности	
Третий: самоактуализация, саморазвитие	Рефлексивность, творчество и креативность	Я должен сделать отличную работу, чтобы показать свою способность и уважать себя	Умение реализовать свои возможности и способности в создании своего продукта	

Процесс поступления в высшее учебное заведение, процесс становления школьника студентом укрепляет его веру в личностные силы и личные способности, помогает возродить надежду на полноценную и интересную жизнь в купе с тесным взаимодействием с остальными. Вместе с тем на 2 и 3 курсах чаще всего появляется вопрос о правильности решения выбранного учебного заведения, а также специальности, профессии. К концу 3 курса полностью решается вопрос о профессиональном личностном определении индивида. Зачастую замечаются сдвиги в настроении студентов - от восторженного в первые полгода учебы в вузе до скептического в процессе узнавания режима внутри вуза, его системы преподавания, отдельных преподавателей, общественной атмосферы, коллектива и т.п.

2.2. Психолого-педагогические особенности одаренных студентов.

Одаренность – это системное, развивающееся в течение жизни качество психики, которое определяет возможность достижения человеком более высоких, незаурядных результатов в одном или нескольких видах деятельности по сравнению с другими людьми.

Одаренный студент – это студент, который выделяется яркими, очевидными, иногда выдающимися достижениями (или имеет внутренние предпосылки для таких достижений) в том или ином виде деятельности.

Уровень, качественное своеобразие и характер развития одаренности – это всегда результат сложного взаимодействия наследственности (природных задатков) и социокультурной среды.

Признаки одаренности проявляются в реальной деятельности обучающегося и могут быть выявлены на уровне наблюдения за характером его действий. Признаки явной (проявленной) одаренности зафиксированы в ее определении и связаны с высоким уровнем выполнения деятельности.

В качестве примера признаков одаренности можно привести:

Особый тип организации знаний одаренного студента:

- высокая структурированность;
- способность видеть изучаемый предмет в системе разнообразных связей;
- свернутость знаний в соответствующей предметной области при одновременной их готовности развернуться в качестве контекста поиска решения в нужный момент времени;

Признаки мотивационного аспекта поведения одаренного обучающегося:

1. Повышенная избирательная чувствительность к определенным сторонам предметной действительности (знакам, звукам, цвету, техническим устройствам, растениям и т. д.) либо определенным формам собственной активности (физической, познавательной, художественно-выразительной и т. д.), сопровождающаяся, как правило, переживанием чувства удовольствия.

2. Повышенная познавательная потребность, которая проявляется в ненасытной любознательности, а также готовности по собственной инициативе выходить за пределы исходных требований деятельности.

3. Ярко выраженный интерес к тем или иным занятиям или сферам деятельности, чрезвычайно высокая увлеченность каким-либо предметом, погруженность в то или иное дело. Наличие столь интенсивной склонности к определенному виду деятельности имеет своим следствием поразительное упорство и трудолюбие.

4. Высокая требовательность к результатам собственного труда, склонность ставить сверхтрудные цели и настойчивость в их достижении, стремление к совершенству.

Психологические особенности студентов, демонстрирующих одаренность, могут рассматриваться лишь как признаки, сопровождающие одаренность, но не обязательно как факторы, ее порождающие.

Блестящая память, феноменальная наблюдательность, способность к мгновенным вычислениям и т. п. сами по себе далеко не всегда свидетельствуют о наличии одаренности.

Критерии одаренности:

– «степень сформированности одаренности»: актуальная; потенциальная.

Актуальная одаренность — это психологическая характеристика обучающегося с такими наличными (уже достигнутыми) показателями психического развития, которые проявляются в более высоком уровне выполнения деятельности в конкретной предметной области по сравнению с возрастной и социальной нормами.

Потенциальная одаренность — это психологическая характеристика обучающегося, который имеет лишь определенные психические возможности (потенциал) для высоких достижений в том или ином виде деятельности, но не может реализовать свои возможности в данный момент времени в силу их функциональной недостаточности. Развитие этого потенциала может сдерживаться рядом неблагоприятных причин (трудными семейными обстоятельствами, недостаточной мотивацией, низким уровнем саморегуляции, отсутствием необходимой образовательной среды и т. д.). Потенциальная одаренность проявляется при благоприятных условиях, которые обязан создать педагог.

– «форма проявления»: явная; скрытая.

Явная одаренность проявляется в деятельности обучающегося достаточно ярко и отчетливо (как бы «сама по себе»), в том числе и при неблагоприятных условиях.

Скрытая одаренность проявляется в атипичной, замаскированной форме, она не замечается окружающими. В результате возрастает опасность ошибочных заключений об отсутствии одаренности такого студента.

2.3. Адаптация личности студента, ее трудности и последствия. Адаптация к учебной деятельности в вузе.

Адаптация - означает приспособление. Адаптация студента к вузу – это не только «голое» приспособление к новой вузовской образовательной деятельности, сколько приобщение его к новой вузовской среде, которая дает новые возможности для реализации своих функций. Процесс адаптации неизбежно проходит через череду противоречий требований предъявляемой вузовской средой и возможностями (способностями) личности соответствовать им. Преодоление этих противоречий, сопровождается изменением поведения, деятельности личности, переосмыслением ранее полученного опыта, приобретением новых ценностей, навыков и умений. Социологические и психолого-педагогические подходы к определению вузовской адаптации различаются. В частности, социальная адаптация личности студента к вузу, как представителя определенной социальной группы, предполагает его активное взаимодействие в социуме. Средствами успешного достижения активности служат общее образование и профессиональная подготовка. Психолого-педагогическая адаптация студента, как объекта (субъекта) учебно-воспитательного процесса вуза и прежде всего конкретного курса, понимает

приспособление личности студента к новым условиям учебной деятельности, выработку оптимальной модели поведения для целенаправленного функционирования.

Показателями успешной адаптации являются высокий статус студента в коллективе и психологическая удовлетворенность этим коллективом в целом. Речь идет о приспособлении студента к условиям учебного процесса и окружения «без ощущения внутреннего дискомфорта и без конфликта со средой». Показателями низкой адаптации - неудовлетворенность учебным трудом, его организацией, коллективом, неформальными связями, бытом и др., что приводит к низким показателям учебы, а в некоторых случаях и к отчислению из вуза.

Важным моментом процесса адаптации студентов-первокурсников является выработка активно-положительного отношения к будущей профессии и скорейшее приспособление к ритму учебной и учебно-профессиональной деятельности.

Структура вузовской адаптации зависит от конкретного исследования. Наиболее часто рассматривают: социальную или социально-психологическую (отражает изменение социальной роли студента, усвоение норм и традиций, сложившихся в вузе), психофизиологическую (ломка прежнего динамического стереотипа, формирование новых установок, навыков и привычек), психологическую (отражает перестройку мышления, речи, внимания, памяти, восприятия, воли, способностей), профессиональную (вхождение в профессиональную среду, усвоение норм и ценностей) и учебную (деятельностную) (отражает приспособление к учебному ритму, методам и формам работы, приобщение к учебному труду, то есть является дидактической адаптацией) - виды адаптации.

Деление адаптации на структурные компоненты является достаточно условным, так как «человек адаптируется как целостная структура, как организм (физиологическая адаптация); как индивид (психологическая адаптация); как личность (социальная адаптация); как субъект труда (профессиональная адаптация); как субъект учебной деятельности (учебная адаптация)».

Таким образом, условно выделяемые компоненты представляют собой единую систему - вузовскую адаптацию.

Выделяют первичную адаптацию - период начального включения студента в учебный процесс, в среду вуза и в коллектив, и вторичную адаптацию последующий период профессионального становления.

Начальный этап вузовской адаптации характеризуется освоением студентами новой информации (об условиях среды, деятельности) претерпевают изменения их представления о предстоящей деятельности и об особенностях новой социальной среды. Как следствие, изменяется самооценка и уровень притязаний личности. На следующем этапе новые представления о профессиональной деятельности и среде приводят к коррекции: норм, ценностей, структуры опыта, направленности на себя, на свою деятельность и свое социальное окружение.

Выделяют пять этапов адаптации:

1) Подготовительный этап адаптации. Он состоит в аккумулировании релевантной информации о предметных и социальных условиях предстоящей деятельности. Он может протекать в одной из двух форм (активно-целенаправленной или пассивной) в зависимости от индивидуально-психологических свойств и мотивационной сферы личности.

2) Этап стартового психического напряжения. Он связан с состоянием нервно-психического переживания подготовительных действий (событий) и первоначального вхождения в новые условия профессиональной деятельности. Это время внутренней мобилизации психических и психофизиологических ресурсов человека, обеспечивающей необходимые предпосылки для функционирования в новых условиях.

3) Этап острых психических реакций входа, на котором студент начинает ощущать воздействие изменившихся факторов предметной и социальной среды на себе (например, впервые столкнувшись с необходимостью принятия решения в новых условиях). Он

характеризуется переживанием состояния фрустрации, вызывающим конструктивные или деструктивные реакции.

4) Этап завершающего психического напряжения. Переход к этому этапу происходит в случае благоприятного развития адаптационного процесса. Этап, своеобразной подготовки психики человека к актуализации прежних режимов функционирования, привычных способов поведения. Такая подготовка связана с предстоящим возвращением к привычной жизни.

5) Этап острых психических реакций выхода. Состоит из комплекса эмоциональных и поведенческих реакций, связанных с вхождением в уже знакомую среду обитания и профессиональной деятельности.

1-3 этапы приходятся на 1-ый и 2-ой курсы обучения. В этот период происходит формирование определенных способов поведения (конструктивных реакций) личности, позволяющих ей справляться с адаптационными трудностями и успешно овладевать новой деятельностью. К конструктивным реакциям можно отнести переоценку ситуации и адекватное замещение способа удовлетворения потребностей, к деструктивным - агрессию и избегание решения возникшей проблемы.

Как известно, процесс адаптации на первом курсе обучения у студентов связан со следующими факторами, например:

- недостаточная психологическая готовность к будущей профессии;
- отсутствие повседневного контроля за выполнением заданий (ослабевает саморегуляция и самоконтроль учащегося);
- неумение работать с первоисточниками, словарями.
- неумение вести конспект.
- налаживание быта при условии жизни в общежитии;
- отрицательные переживания, связанные с уходом старого школьного коллектива.

Процесс адаптации у студентов проходит по следующим формам:

- формальная (приспособление студентов к структуре высшей школы);
- общественная (приспособление студентов к студенческому коллективу);
- дидактическая (приспособление студентов к новым формам обучения).

Никитиной К.А. выделяет низкий, средний и высокий уровни адаптированности в которых отражены показатели физиологического, психологического и социально-психологического видов адаптации.

Низкий уровень адаптированности характеризуется повышенной утомляемостью, низким и ниже среднего уровнями соматического здоровья, несогласованностью взаимной ответственности при выполнении коллективной работы, наличием отрицательного эмоционального состояния тревожности, студенты несамостоятельны, не умеют выстраивать межличностные отношения, не готовы к обучению в вузе, испытывают сложности в усвоении информации.

Для среднего уровня адаптированности характерны: средний уровень соматического здоровья, наличие мотивов самостоятельной работы и построения межличностных отношений, но при этом недостаточная способность включаться в жизнь новых социальных групп, готовность преодолевать трудности, наличие волевых качеств.

Высокий уровень адаптированности предполагает наличие выше среднего и высокого уровня соматического здоровья, сформированность умений и навыков практической работы, умения самостоятельно учиться, удовлетворенность общением с сокурсниками, преподавателями и низким уровнем тревожности.

Адаптация студентов к новым условиям обучения и общения на разных курсах имеет свои отличительные черты. На 1 курсе, поведение студента отличается высокой степенью конформизма, отсутствует дифференцированный подход к своей роли. На 2 - период напряженной учебы, процесс адаптации к среде завершается. На 3 - происходит углубление профессиональных знаний и умений, сужается сфера интересов в связи с конкретизацией своей будущей профессии. На 4 - переоценка ценностей, связанная с практической

деятельностью в профессии, студенты отходят от коллективной формы общения, идет перестройка на будущие профессиональные и семейные установки.

Важным является вопрос о типологии студентов в процессе вузовской адаптации. Здесь рассматривая адаптационные типы людей на основании способности стратегического или комплексного решения проблем жизнедеятельности, выделяют: прогрессивно-творческий, адаптивно-репродуктивный, адаптивно-активационный и адаптивно-деформирующий типы.

Прогрессивно-творческий тип - для него характерна постановка широкого спектра перспективных, обобщенных задач оптимизации труда и собственного саморазвития.

Адаптивно-репродуктивный тип - для него характерно упрощение актуальной ситуации и сведение их к известному из прошлого опыту, здесь в основе проблематизации лежит стереотипизация.

Адаптивно-активационный тип - для него характерно вместе с использованием стереотипизации использовать проблематизацию, искусственно усложняющая ситуацию, создавая квазипроблему.

Адаптивно-деформирующий тип - для него характерно невозможность комфортного существования «вне точки» своей профессии, что является следствием неправомерной аксеологии субъектов средств, условий и т.п. своей профессиональной деятельности.

Важным вопросом в процессе успешного обучения студента считается освоение специфики учебы в вузе, исключая ощущение внутреннего дискомфорта и ограничивающее возможность появления конфликта с окружающим миром. В начале процесса обучения формируется студенческий коллектив, определяются имеющиеся навыки и возможности рациональной организации интеллектуального труда, осознается призвание к выбранной профессии, формируется оптимальный режим трудовой деятельности, создается структура работы по личному образованию и воспитанию профессионально значимых качеств индивида.

Можно заметить, что система обучения в вузе имеет массу отличий от школьной системы, потому как в школе учебный процесс построен таким образом, что у ученика постоянно и регулярно прививается желание к занятиям, в противном случае появится масса двоек. В вузовской же системе все обстоит иначе: лекции, лекции, лекции. Во время семинаров тоже, как оказалось, можно не постоянно готовиться, не надо на каждое занятие учить, решать и запоминать. Как результат часто формируется мнение о кажущейся легкости обучения в высшем учебном заведении в первом семестре, создается перспектива все наверстать и освоить информацию перед сессией, появляется беспечное отношение к учебе. Однако сессия ставит все мнения на правильные места и меняет беспечность индивида на озадаченность.

Адаптация студентов к учебному процессу в вузе заканчивается в конце 2-го - начале 3-го учебного семестра.

Главной задачей работы с первокурсниками считается процесс формирования и применения на практике методов рационализации и оптимизации свободного времени и самостоятельной работы. Уже сформированная система контроля за самостоятельной работой студентов с помощью семинарских, практических и лабораторных занятий включает в себя и пассивность, и уклонение студентов от выполнения соответствующих требований. Для формирования тактики и стратегии, позволяющей создать оптимальную адаптацию индивида к системе вузовского обучения, необходимо знать жизненные цели и интересы индивида, структуру доминирующих мотивов, уровень притязаний, личную самооценку, возможность индивидуальной и личной регуляции поведения и т.д.

Адаптация и ее протекание для каждого индивида происходит по-разному. Индивиды, обладающие уже трудовым стажем, проще и быстрее адаптируются к новым условиям студенческой жизни. Основная цель студенческой группы заключается в формировании подходящих условий для совокупной оптимальной деятельности всех индивидов в отдельности.

Еще одним важным аспектом в понимании важности адаптации у индивида в вузе состоит в объяснении основных черт индивида на каждом курсе.

На первом курсе индивид только учится приобщаться к студенческим формам коллективной жизни. Его поведение характеризуется чрезмерно высокой степенью конформизма; у индивида на первом курсе отсутствует дифференцированный подход к своим ролям как личности.

На втором курсе обычно происходит самая напряженная учебная деятельность индивида. У него как у второкурсника с особой силой включены все формы и обучения, и воспитания. К данному периоду процесс адаптации к внутривузовской среде чаще всего завершен.

Третий и четвертый курс характеризуется углубленным изучением специализации, увеличением интереса к выбранному направлению и научной работе как отражение дальнейшего развития. Индивид уже вполне сформирован как личность, к этому моменту пересмотрены уже все жизненные ценности и приоритеты, что помогает установить правильные установки на последующий род его деятельности.

Процесс нахождения спутника жизни играет значительную роль, оказывая непосредственное влияние и на успеваемость, и на общую направленность личности студента. Однако ошибочно считать, что в нахождении второй половинки есть негативное влияние. Интимные отношения часто вовлекают индивида в увеличение желания учиться, рабочему настроению, а также и в увеличение творческой деятельности.

Вопросы для повторения:

1. Что включает в себя адаптация индивида?
2. Какие виды адаптации выделяют?
3. Какие три аспекта характеристики личности студента существуют?
4. Как можно охарактеризовать юность с точки зрения развития личности?
5. Какие основные трудности появляются в процессе адаптации индивида к вузу?
6. Дайте определение понятия «адаптация» с точки зрения физиологии и медицины, педагогики, социологии и психологии.
7. Вузовская адаптация. Каковы показатели высокой и низкой вузовской адаптации?
8. Перечислите компоненты, выделяемые различными авторами в структуре вузовской адаптации.
9. Какие адаптационные типы людей вам известны?
10. Дайте описание уровней адаптированности студентов.

Тема 3. Профессиональное становление.

3.1. Профессиональное становление. Факторы профессионального становления.

3.2. Стадии и кризисы профессионального становления. Противоречия профессионального становления.

3.1. Профессиональное становление. Факторы профессионального становления.

Изменения, происходящие с личностью в процессе подготовки, овладения профессиональной деятельностью и самостоятельного выполнения, приводят к становлению ее как специалиста и профессионала.

Методологической основой профессиональной психологии является концепция профессионального становления личности (ПСЛ). Ее суть заключается в том, что, выбирая и осваивая профессию, личность изменяется: обогащается ее направленность, расширяются опыт и компетентность. Профессиональное становление личности сопровождается кризисами, конфликтами и деструктивными изменениями. Темп и траектория этого процесса детерминируется биологическими и социальными факторами, собственной активностью личности, а также случайными обстоятельствами, жизненно важными событиями и профессионально обусловленными инцидентами.

Изменение личности в процессе освоения мира профессий отражается в понятии «*профессиональное становление*», которое характеризует индивидуально своеобразный путь (траекторию) личности, большую часть онтогенеза человека: с начала формирования профессиональных намерений до завершения профессиональной жизни.

Для решения проблем профессионального становления личности, обусловленных необходимостью профессионального самоопределения и выбора профессии, профессионального образования и повышения квалификации, профессионального роста и карьеры, профессиональной адаптации и достижения вершин профессионализма, прежде всего, требуется определиться с ведущими смыслообразующими понятиями.

К ним относятся, прежде всего: «профессиональное становление», «профессиональное развитие» и «профессионализация» личности. Понятия эти не тождественны.

Профессиональное становление личности - движение личности в профессионально-образовательном пространстве и времени профессиональной жизни. Это индивидуально своеобразный путь (траектория) личности с начала формирования профессиональных интересов и склонностей до окончания активной профессиональной деятельности.

Профессиональное развитие - это изменение психики в процессе освоения и выполнения профессионально-образовательной, трудовой и профессиональной деятельности. Объектом развития выступает субъект деятельности. Факторами, детерминирующими его развитие, являются социально-экономическая ситуация и ведущая деятельность. Профессиональное развитие человека происходит при его взаимодействии с миром профессий.

Профессионализация - «формообразование» субъекта, адекватного содержанию и требованиям профессиональной деятельности.

Существует множество внешних и внутренних факторов, в течение длительного времени оказывающих влияние на профессионализацию личности.

В качестве основных внешних факторов, оказывающих влияние на профессиональное развитие личности, ученый указывает 1) семью, школу, церковь; 2) социально-классовую позицию; 3) этнос, расу, национальные особенности индивида; 4) общую культуру (т.е. те нормы, традиции, которые сложились в обществе). Профессиональные интересы развиваются в результате взаимодействия между родителями и их детьми. Выбор карьеры индивидом отражает желание удовлетворить потребности, которые не были удовлетворены родителями в детстве. Формирование жизненных стереотипов происходит прежде всего в течение нескольких первых лет детства.

1. Существует три различных стереотипа отношений между родителями и детьми:

1 - характеризуется эмоциональной концентрацией на детях. Существуют две формы этого стереотипа: гиперопека, при которой родители слишком много делают для детей и поддерживают их зависимость; сверх требовательность, родители концентрируются на достижениях детей. В этом случае у детей вырабатывается потребность в постоянной обратной связи и поощрении. Они часто выбирают карьеры, которые обеспечивают общественное признание.

2 – отстраненность от детей. Существуют два экстремальных проявления данного стереотипа: пренебрежение родительскими обязанностями (для удовлетворения потребностей детей прилагается чрезвычайно мало усилий); отказ от воспитания (не предпринимается никаких усилий, чтобы удовлетворить потребности ребенка). Дети, в таких условиях, во взрослой жизни концентрируются на карьере, представляющей научный и технический интерес, находя в этом удовлетворение. Они более склонны иметь дело с предметами и идеями.

3 – принятие детей. Может быть непреднамеренным или более активным проявлением любви. Здесь поощряется независимость детей. Дети, выросшие в таких семьях, обычно выбирают карьеру, уравнивающую личностные и социальные аспекты жизни, например, карьеру учителя или консультанта.

Существуют три типа косвенного влияния семьи на профессиональное становление детей: 1) социальный уровень семьи; 2) усвоение ценностей, предпочтений семейного окружения; 3) ретрансляционная функция семьи. В этом случае семейное окружение осуществляет оценку информации, исходящей из СМИ, и транслирует ее ребенку

Школа оказывает целенаправленное воздействие на профессиональное определение личности, поскольку это ее основная цель. Для выполнения данной функции школа имеет специальные средства (способы воздействия) и специалистов (учителей). Здесь закладываются интересы, формируются и развиваются способности. Если в семье влияние на профессиональное становление ребенка носит стихийный характер, то школа обязана заниматься этим системно.

Церковь с точки зрения профессионализации личности осуществляет в основном воспитательную задачу, формируя в личности понимание необходимости трудиться.

II. В качестве следующего внешнего фактора, оказывающего влияние на профессиональное развитие личности, является социально-классовая позиция. Ребенок, выросший в семье рабочих, вряд ли сможет осуществить свою профессиональную деятельность в банковской системе.

III. Этнос, раса, национальные особенности индивида. В связи с этим идет дифференциация сфер деятельности (неквалифицированный труд) в зависимости от цвета кожи.

IV. Общая культура, нормы и традиции, сложившиеся в обществе. В данном случае можно привести пример норм и традиций, сложившихся на Востоке, когда профессия обязательно передается по мужской линии от отца к сыну

Достаточно большое количество разнообразных факторов оказывает воздействие на профессиональное становление личности извне. Однако еще большее количество факторов, связанных с индивидуально-психологическими, личностными особенностями индивида, влияют на профессионализацию субъекта деятельности «изнутри». Данную категорию механизмов воздействия принято обозначать как «внутренние факторы».

3.2. Стадии и кризисы профессионального становления. Противоречия профессионального становления.

Согласно эвристической модели профессионально-образовательного пространства, предложенной Э.Ф. Зеером, профессиональное становление личности представляется в нем траекторией, которая формируется как результат совместного взаимодействия трех факторов:

- возрастные изменения, обуславливающие периодизацию развития личности;
- система непрерывного образования;
- ведущая профессионально ориентированная деятельность.

Данные факторы являются координатами профессионально-образовательного пространства (рис. 1).

Для визуального представления взаимодействия трех факторов, определяющих профессиональное становление личности, можно допустить их направленное, последовательное изменение, выделив уровни их выраженности. Такое допущение позволяет представить возможные траектории профессионального становления личности (рис. 1). Кратко остановимся на характеристике данных факторов.



Рисунок 1. Модель профессионально-образовательного пространства

Возрастные изменения человека в течение длительного периода онтогенеза являются важным фактором профессионального становления личности. Особое значение в обеспечении и поддержке профессионального становления имеет система непрерывного профессионально ориентированного образования. Общее образование является ведущим фактором развития личности и предпосылкой успешного профессионального становления.

Также базовым, ключевым фактором профессионального становления личности является ведущая деятельность, которая формирует ее отношения с социально-экономической средой, общение с окружающими, институализирует социальную ситуацию развития.

Для каждой стадии профессионального становления характерна практически одна ведущая деятельность. Многоаспектность выполняемой личностью деятельности обогащают процесс профессионального становления. Считается, что для периода зрелости дифференциация ведущей деятельности отсутствует.

В качестве основания дифференциации онтогенеза может использоваться активность личности. В зависимости от уровня психической (социально и профессионально обусловленной) активности личности можно выделить следующие уровни социально-профессиональной активности: нормативно заданная; адаптивная; надситуативная; сверхнормативная.

Эти уровни активности имеют место на всех возрастных стадиях, но для каждого возрастного периода можно выделить преобладающий уровень психической активности, который и определяет характер ведущей деятельности. Нормативно заданная активность определяет учебно-профессиональную и профессионально-образовательную ведущую деятельность, адаптивная - нормативно одобряемую профессиональную деятельность (или воспроизводящую), надситуативная - продуктивную (высококвалифицированную), сверхнормативная - творческую самодетельность. Последняя ведущая деятельность не имеет предела совершенствования.

Согласно модели (рис.1) прямолинейные участки индивидуальной траектории профессионального становления личности представляют собой усредненный вектор ее эволюционного развития. В силу нелинейного характера совместного действия факторов профессионально-образовательного пространства и других воздействий среды, оказывающих влияние на субъект деятельности, могут наблюдаться случайные отклонения вокруг вектора профессионального становления личности. Их существование обусловлено случайными внешними обстоятельствами, личностными кризисами, неудовлетворенностью собой, пресыщением рутинной деятельностью и т.п. Они постоянно нарушают эволюционный характер профессионального становления личности.

Ограниченность и предопределенность траекторий профессионального становления личности обусловлена социальной структурированностью общественного воспроизводства. При определенных внешних обстоятельствах и уровне развития структуры личности отклонения могут служить «спусковым механизмом» для перехода или переключения человека на новую траекторию профессионального становления. Области, где происходит изменение вектора развития, являются критическими точками или точками перестройки. В них происходит «излом» траектории, появляются несколько вариантов новых траекторий профессионального становления личности. Периоды между критическими точками называются стадиями профессионального становления.

Развитие профессионального становления является неустойчивым, неупорядоченным. Не все стадии периодизации поочередно сменяют друг друга, некоторые стадии могут даже отсутствовать. При переходе от одной стадии профессионального становления личности к другой происходит смена социальной ситуации развития, изменяется содержание ведущей деятельности, возникает новая социальная роль, профессиональное поведение и, конечно, трансформируется личность. В критических точках траектории профессионального становления поведение личности под воздействием внешних факторов становится неустойчивым и может продолжиться по одной из нескольких альтернативных новых стадий развития субъекта деятельности. Нарушение эволюционного развития может инициироваться одним из факторов профессионально-образовательного пространства: возрастными изменениями, социально-экономической ситуацией, ростом уровня профессионального образования и квалификации, перестройкой способов выполнения деятельности, а также случайным стечением обстоятельств.

Профессионально обусловленные изменения порождают субъективные и объективные трудности, межличностные и внутри-личностные конфликты. Развертывание этих психологических проблем приводит к кризисам профессионального становления.

Субъективное переживание кризисов способствует возникновению критических моментов, так называемых точек «раздвоения», которые побуждают личность к поиску новых путей реализации профессиональной биографии.

Профессиональное становление охватывает длительный период жизни человека - 35-40 лет. Поэтому возникает необходимость разделения данного процесса на периоды, или стадии. Цикличность процесса профессионального становления человека отражается в его периодизации, которая позволяет систематизировать, обобщить и согласовать многочисленные эмпирические факты и частные закономерности. Периодизацию процесса профессионального становления проводят по двум группам содержательных и формальных классификационных признаков, соответственно:

- качественные изменения (новообразования), характерные для каждого периода;
- длительность периодов в определенных временных единицах измерения.

В настоящее время наибольшую известность в России приобрела периодизация жизненного пути профессионала, предложенная Е.А. Климовым, который предлагает в профессиональном становлении личности вычленять следующие стадии, или фазы:

- оптация - период выбора профессии и пути ее приобретения;
- адаптация - вхождение в профессию и привыкание к ней;
- фаза интернала - приобретение профессионального опыта;
- мастерство - квалифицированное выполнение трудовой деятельности;
- фаза авторитета-достижение наивысшей квалификации;
- наставничество - передача опыта молодому поколению.

А.К. Маркова в качестве критерия для выделения этапов становления профессионала берет уровни профессионализма личности. Она различает пять уровней и девять этапов:

- допрофессионализм - этап первичного ознакомления с профессией;
- профессионализм состоит из трех этапов: адаптации к профессии, самоактуализации в ней и свободного владения профессией в форме мастерства;

- суперпрофессионализм также состоит из трех этапов: свободное владение профессией на уровне творчества, овладение рядом смежных профессий, самопроектирование себя как личности;

- непрофессионализм - выполнение труда по профессионально искаженным нормам на фоне деформации личности;

- слепопрофессионализм - завершение профессиональной деятельности.

За рубежом широкое признание получила периодизация Дж. Сьюпера, выделившего пять основных этапов профессиональной зрелости (в некоторых изданиях - от 4 до 6 этапов):

- 1) рост - развитие интересов и способностей (до 14 лет);

- 2) исследование - апробация своих сил (14-25 лет);

- 3) утверждение — профессиональное образование и упрочение своих позиций в обществе (25-44 года);

- 4) поддержание - создание устойчивого профессионального положения (45-64 года);

- 5) спад - уменьшение профессиональной активности (после 65).

В онтогенетических моделях профессионального становления и реализации субъекта деятельности в основном используется следующая временная структура периодизации:

- 1) стадии или уровни. *Стадия* - определенная ступень в развитии. Это самые длительные временные интервалы между критическими точками индивидуальной траектории профессионального. Для выделения стадий профессионального становления личности используют социальную ситуацию и уровень реализации ведущей деятельности, а также факторы профессионально образовательного пространства;

- 2) периоды или этапы. Стадии по своему психологическому содержанию являются неоднородными и в свою очередь могут делиться на отдельные периоды или этапы. Основанием для выделения является уточнение ситуации профессионального развития и конкретизация задач профессионального становления. В общем случае можно выделить три нормативных периода в составе стадии: адаптация или завершение решения задач развития предыдущей стадии, период решения основной задачи профессионального становления данной стадии, подготовка к переходу на новую стадию развития.

- 3) фазы. Структурными элементами периода являются фазы. Данный элемент периодизации должен определяться основаниями, выделяющего его в рамках нормативных задач профессионального становления.

Обобщив различные подходы к периодизации профессионального становления личности, Э.Ф. Зеер предлагает выделить следующие стадии (табл. 2):

Таблица 2 - Стадии профессионального становления личности

Название стадии и возраст	Социальная ситуация	Ведущая деятельность	Основные психологические новообразования на данной стадии
1. Аморфная оптация (0-12 лет)	Влияние родителей, родственников и учителей	Сюжетно-ролевые игры и учеба в школе, занятия в кружках и секциях	Зарождение профессионально ориентированных интересов и склонностей
2. Оптация (12-16 лет)	Завершающий период детства. Поиск своего места в мире профессий и в жизни	Учебно-профессиональная. Развитие познавательных и	Профессиональные намерения. Выбор пути профессионального образования, учебно-

		профессиональных интересов	профессионального самоопределения
3. Профессиональная подготовка (16-23 года)	Поступление в профессиональное учебное заведение. Новые социальные роли, взаимоотношения, социальная независимость	Профессионально-познавательная, ориентированная на получение конкретной профессии	Профессиональная подготовленность, профессиональное самоопределение, готовность к самостоятельному труду
4. Профессиональная адаптация (18-25 лет)	Новая система отношений в разновозрастном производственном коллективе	Профессиональная деятельность на нормативно репродуктивном уровне	Освоение новой социальной роли, самостоятельная профессиональная деятельность
5. Первичная профессионализация (23-27 лет)	Новая система отношений к окружающей действительности	Стабильная профессиональная деятельность	Профессиональная позиция, интегративные профессионально значимые конstellляции
6. Вторичная профессионализация (27-33 года)	Стабилизация профессиональной активности, высокий уровень профессиональной деятельности	Дальнейшее повышение квалификации, выработка собственной профессиональной позиции, высокое качество и производительность труда	Профессиональный менталитет, идентификация с профессиональным сообществом, профессиональная мобильность, корпоративность, гибкий стиль и высококвалифицированная деятельность
7. Профессиональное мастерство (33-55 лет)	Этой стадии достигают не все, а только обладающие творческой потенцией, развитой потребностью в самоосуществлении и самореализации	Высокая профессиональная и социальная активность, поиск новых способов деятельности и взаимоотношений, стремление выйти за пределы себя	Творческая профессиональная деятельность, подвижные интегративные психологические новообразования, само проектирование своей деятельности и карьеры, вершина профессионального развития

Переход между стадиями сопровождается кризисами профессионального становления.

Кризис определяется как резкий, крутой перелом, тяжелое переходное состояние. Каждый кризис свидетельствует о завершении одного этапа жизни, развитии человека и о начале перехода к следующему или о начале поиска такового. Любой психологический кризис сопровождается трудными психическими состояниями и тяжелыми эмоциональными переживаниями.

Анализируя различные подходы и детерминанты, вызывающие кризисы, Э.Ф. Зеер классифицирует их на шесть типов: 1) нормативные, 2) психического развития и 3) профессионального становления, 4) ненормативные, 5) критические, 6) невротические). Первые три типа психологических кризисов, имеющих относительно выраженный хронологический характер, объединяются в группу возрастных, или нормативных. А вторые три типа - в группу жизненных, имеющих ненормативный, вероятностный характер.

Нормативный - значит изменяющий траекторию развития в соответствии с нормами профессионального становления личности, а ненормативный - имеющий случайный характер, не связанный с поступательным движением к овладению профессией.

Возрастные кризисы переживают все люди, но уровень их выраженности не всегда приобретает характер конфликта. Преобладающая тенденция возрастных кризисов конструктивная, развивающая личность.

Жизненные кризисы во многом случайны. Они возникают вследствие стечения обстоятельств. Выход из таких кризисов проблематичен. Иногда он бывает деструктивным, и тогда общество получает циников, маргиналов, бомжей, алкоголиков, самоубийц.

Типологию этих групп кризисов можно представить в виде следующей схемы (рис. 2).



Рисунок 2 - Типология кризисов

Под кризисами профессионального становления понимаются непродолжительные по времени периоды (до года) кардинальной перестройки профессионального сознания, деятельности и поведения личности, изменения вектора ее профессионального развития. Кризисы приводят к переориентации на новые цели, коррекции и ревизии социально-профессиональной позиции, подготавливают смену способов выполнения деятельности, ведут к изменению взаимоотношений с окружающими людьми, а в отдельных случаях - к смене профессии.

Основываясь на концепции профессионального становления личности, кризисы можно определить как резкие изменения вектора ее профессионального развития. Непродолжительные по времени, они наиболее ярко проявляются при переходе от одной стадии профессионального становления к другой. Кризисы протекают, как правило, без ярко выраженных изменений профессионального поведения. Однако происходящая перестройка смысловых структур профессионального сознания могут привести и к более кардинальным изменениям, вплоть до смены места работы и профессии.

К факторам, детерминирующим кризисы профессионального развития, относятся:

- сверхнормативная профессиональная активность, которая может выразиться в переходе на новый образовательно-квалификационный либо творческий уровень выполнения деятельности;

- возросшая социально-профессиональная активность личности вследствие ее неудовлетворенности своим социальным и профессионально-образовательным статусом;
- социально-экономические условия жизнедеятельности человека: ликвидация предприятия, сокращение рабочих мест, неудовлетворительная зарплата, переезд на новое местожительство и др.;
- возрастные психофизиологические изменения: ухудшение здоровья, снижение работоспособности, ослабление психических процессов, профессиональная усталость и т.д.;
- полная поглощенность профессиональной деятельностью;
- изменения жизнедеятельности (смена местожительства; перерыв в работе, связанный с уходом за малолетними детьми; «служебный роман» и т.п.).

Кризисные явления нередко сопровождаются нечетким осознанием недостаточного уровня своей компетентности и профессиональной беспомощностью. Иногда наблюдаются кризисные явления при уровне профессиональной компетентности, более высоком, чем требуется для выполнения нормативной работы. Как следствие возникает состояние профессиональной апатии и пассивности.

Проанализируем кризисы профессионального развития личности.

- Предкризисная фаза обнаруживается в неудовлетворенности существующим профессиональным статусом, содержанием деятельности, способами ее реализации, межличностными отношениями. Эта неудовлетворенность не всегда отчетливо осознается, но проявляется в психологическом дискомфорте на работе, раздражительности, недовольстве организацией, оплатой труда, руководителями и т.п.

- Кризисная фаза отличается осознанной неудовлетворенностью реальной профессиональной ситуацией. Намечаются варианты ее изменения, проигрываются сценарии дальнейшей профессиональной жизни, усиливается психическая напряженность. Противоречия усугубляются, и возникает конфликт, который становится ядром кризисных явлений. Конфликт сопровождается рефлексией, ревизией учебно-профессиональной ситуации, анализом своих возможностей и способностей.

- Разрешение конфликта приводит кризис в посткризисную фазу. Способы разрешения конфликтов могут иметь конструктивный, профессионально-нейтральный и деструктивный характер.

Конструктивный выход из конфликта предполагает повышение профессиональной квалификации, поиск новых способов выполнения деятельности, изменение профессионального статуса, смену места работы и переквалификацию. Такой путь преодоления кризисов требует от личности проявления сверхнормативной профессиональной активности, совершения поступков, которые прокладывают новое русло для ее профессионального развития.

Профессионально-нейтральное отношение личности к кризисам приводит к профессиональной стагнации, равнодушию и пассивности. Личность стремится реализовать себя вне профессиональной деятельности: в быту, различного рода хобби, садоводстве и т.п.

Деструктивные последствия кризисов выражаются в нравственном разложении, профессиональной апатии, пьянстве, безделье.

В рассматриваемой нами концепции профессионального становления личности выделены следующие стадии этого процесса: оптация, профессиональное образование и подготовка, профессиональная адаптация, первичная и вторичная профессионализация и мастерство. Согласно определению кризисов, переход от одной стадии к другой порождает нормативные кризисные явления. Рассмотрим их психологические особенности, следуя логике профессионального становления.

1. Профессиональное становление личности начинается со стадии оптации, когда происходит смена ведущей деятельности с учебно-познавательной на учебно-профессиональную. Кардинально изменяется социальная ситуация развития, порождающая

неизбежное столкновение желаемого будущего и реального настоящего, которое приобретает характер кризиса учебно-профессиональной ориентации.

Старшеклассники, продолжившие учебу в 10-11 классах, переживают этот кризис в 16-17 лет, перед завершением школьного образования. Ядром кризиса является необходимость выбора способа получения профессионального образования или профессиональной подготовки. Деструктивное разрешение кризиса приводит к ситуативному выбору профессиональной подготовки или профессии, выпадению из нормальной социальной сферы.

2. На стадии профессиональной подготовки многие учащиеся и студенты переживают разочарование в получаемой профессии. Возникает недовольство отдельными учебными предметами, появляются сомнения в правильности профессионального выбора, падает интерес к учебе. Наблюдается кризис профессионального выбора. Как правило, он отчетливо проявляется в первый и последний годы профессионального обучения. За редким исключением этот кризис преодолевается сменой учебной мотивации на социально-профессиональную.

3. После завершения профессионального образования наступает стадия профессиональной адаптации. Первые недели и месяцы самостоятельной работы вызывают большие трудности. Но они не становятся фактором возникновения кризисных явлений. Основная причина психологическая - несовпадение реальной профессиональной жизни со сформировавшимися представлениями и ожиданиями. Это несоответствие вызывает кризис профессиональных экспектаций (ожиданий).

Переживание этого кризиса выражается в неудовлетворенности организацией труда, его содержанием, должностными обязанностями, производственными отношениями, условиями работы и зарплатой.

Возможны два варианта разрешения кризиса:

- конструктивный: активизация профессиональных усилий по скорейшей адаптации и приобретению опыта работы;
- деструктивный: увольнение, смена специальности; неадекватное, некачественное, непродуктивное выполнение профессиональных функций.

4. Возникает на завершающей стадии первичной профессионализации, после 3-5 лет работы. При отсутствии перспектив профессионального роста личность испытывает дискомфорт, психическую напряженность, появляются мысли о возможном увольнении, смене профессии. Кризис профессионального роста может временно компенсироваться разного рода непрофессиональными, досуговыми видами деятельности, бытовыми заботами или же кардинально решаться путем ухода из профессии. Стабилизация же всех сторон профессиональной жизни способствует профессиональной стагнации личности: смирению и профессиональной апатии. Стагнация может длиться годами, иногда до ухода на пенсию.

5. Вторичная профессионализация. Кардинально перестраиваются социально-профессиональные ценности и отношения. Ведущая деятельность на этой стадии характеризуется индивидуальным стилем и элементами творчества. Во многих случаях качественное и высокопродуктивное выполнение деятельности приводит к тому, что личность перерастает свою профессию. Сформировавшееся к этому времени профессиональное самосознание подсказывает альтернативные сценарии дальнейшей карьеры, и не обязательно в рамках данной профессии. Противоречия между желаемой карьерой и ее реальными перспективами приводят к развитию кризиса профессиональной карьеры. При этом серьезной ревизии подвергается «Я-концепция», вносятся коррективы в сложившиеся производственные отношения. Возможные сценарии выхода из кризиса: увольнение, освоение новой специальности в рамках той же профессии, переход на более высокую должность. Одним из продуктивных вариантов является переход на следующую стадию профессионального становления стадию мастерства.

6. Стадия мастерства характеризуется творческим и инновационным уровнем выполнения профессиональной деятельности. Движущим фактором дальнейшего профессионального развития личности становится потребность в самореализации, которая нередко приводит к неудовлетворенности собой, окружающими людьми. Кризис социально-профессиональной самоактуализации, - это душевная смута, бунт против себя. Продуктивный выход из него - новаторство, изобретательство, стремительная карьера, социальная и профессиональная сверхнормативная активность. Деструктивные варианты разрешения кризиса - увольнение, конфликты, профессиональный цинизм, алкоголизм, создание новой семьи, депрессия.

7. Кризисный характер для многих работников приобретает и предпенсионный период. Уход на пенсию означает сужение социально-профессионального поля и контактов, снижение финансовых возможностей. Острота протекания кризиса утраты профессиональной деятельности зависит от характера трудовой деятельности (работники физического труда переживают его легче), семейного положения и здоровья.

8. Социально-психологическое старение. Проявляется в ослаблении интеллектуальных процессов, повышении или снижении эмоциональных переживаний. Отмечается пристрастие к морализированию и осуждению поведения молодежи, четко прослеживается противопоставление своего поколения поколению, идущему на смену. Неудовлетворенность современной жизнью обусловлена укорочением жизненной перспективы. Данное беспокойное и тревожное состояние личности определяется как кризис социально-психологической адекватности.

Ненормативные кризисы - это кризисы, обусловленные ненормативными, случайными событиями. Предсказать эти кризисы невозможно, они сугубо индивидуальны и возникают вследствие стечения обстоятельств. К ненормативным кризисам относятся:

1) Кризисы невротического характера, которые появляются вследствие внутриличностных изменений, а именно перестройки сознания, различных бессознательных впечатлений, иррациональных тенденций;

2) Жизненные кризисы, которые связаны с изменением индивидуальной биографии человека. В большинстве случаев следствием данных событий является перестройка сознания и поведения у человека;

3) Критические кризисы, которые связаны с трагическими событиями в жизни человека. Эти кризисы инициируют сильнейшие эмоциональные переживания, перестройку ценностей, потерю прежнего смысла жизни и поиск нового, изменения поведения и сознания.

Таким образом, ненормативные кризисы - это сугубо индивидуальный спектр событий в жизни каждого конкретного человека, спрогнозировать который крайне затруднительно.

Условием преодоления развивающегося кризиса может стать определение областей приложения накопленного опыта, в которых можно получить подтверждение своей полезности, нужности. Главное - насытить жизнь активной деятельностью.

Решающее значение в возникновении кризисов на первых стадиях профессионального становления имеют *объективные факторы*: смена ведущей деятельности, кардинальное изменение социальной ситуации. На последующих стадиях все большую роль играют *субъективные факторы*: изменение «Я-концепции», перестройка профессионального сознания, возрастание уровня притязаний и самооценки, проявление потребности в самоутверждении и самоосуществлении. Продуктивное выполнение деятельности приводит к тому, что профессионализм личности перерастает саму деятельность.

Общие моменты в определении кризиса. Во-первых, кризис обязательный этап развития, относительно непродолжительный во времени по сравнению со стабильными этапами развития личности. Во-вторых, исход кризиса, может быть либо «положительным» для личности, т. е. происходит переход на качественно новый уровень развития; либо «отрицательным», т.е. деструктивная, неадаптивная линия развития личности. В-третьих,

кризис является следствием «конфликта» между личностью и наличными условиями развития социальной ситуации, возникшее противоречие порождает кризис.

Вопросы для повторения:

1. Что понимают под профессиональным становлением личности?
2. Какие факторы влияют на профессиональное становление личности?
3. Что представляет собой профессиональное становление личности в модели профессионально-образовательного пространства?
4. Как определяются стадии профессионального становления личности по Зееру Э.Ф.?
5. Приблизительно какой период жизни человека охватывает его профессиональное становление?
6. Что такое периодизация профессионального становления?
7. Сколько стадий периодизации профессионального становления и какие предложены Климовым Е.А.?
8. Какие типы кризисов личности выделяет Зеер Э.Ф.?
9. Какой тип кризиса лежит в основе кризиса профессионального становления?
10. Что понимается под кризисом профессионального становления?
11. Какие выделяют виды нормативных кризисов профессионального становления?

Тема 4. Лидерство в организации.

- 4.1. Феномен лидерства. Психологическое содержание понятия «лидерство».
- 4.2. Стили лидерства.
- 4.3. Лидерство и руководство.
- 4.4. Гендерные аспекты организационного руководства и лидерства.
- 4.5. Методика формирования команды. Организация межличностных, групповых и организационных коммуникаций.

4.1. Феномен лидерства. Психологическое содержание понятия «лидерство».

Английское leader имеет ряд сходных значений: руководитель, вождь, глава, командир, которые с известной степенью условности можно обобщить как «некто главный» или даже «самый главный». В то же время этимология этого английского слова восходит к lead - путь, дорога. Исходя из этого понятие «лидерство» часто трактуется как «идущий впереди», «указывающий дорогу». Лидерство – процесс межличностного влияния, обусловленный реализацией ценностей, присущих членам группы и направленный на решение стоящих перед группой целей.

Понятие «лидер» в психологии определяют следующим образом:

1) Лидер — член группы, обладающий наибольшим ценностным потенциалом, который и обеспечивает ему ведущее влияние в группе. В малой группе лидер может выступать как организатор, мотиватор деятельности, он может быть наиболее отзывчивым или влиятельным.

2) Лидер – член группы, за которым она признает право принимать ответственные решения в значимых для нее ситуациях, т.е. авторитетная личность, реально играющая центральную роль в организации совместной деятельности и регулировании взаимоотношений в группе.

Классификация лидеров:

- по содержанию деятельности: вдохновитель; исполнитель;
- по характеру деятельности: универсальный; ситуативный;
- по направленности деятельности: эмоциональный; деловой.

4.2. Стили лидерства. Лидерство и руководство. Гендерные аспекты организационного руководства и лидерства.

Модель лидерства как научного управления, при котором руководителя интересует не сам работник, а наиболее оптимально устроенная среда, заменились в середине двадцатых годов на модель «человеческих взаимоотношений». В модели «человеческих взаимоотношений» роль руководителя интерпретируется по типу «лидер – ведомый».

В соответствии с различными теориями лидерства проводили исследования на определение стилей лидерства, которые разделили на X-теории или Теории Y.

Теория X, в соответствии которой лидер должен обладать чертами диктатора, основана на следующих представлениях:

- люди обычно не любят работать и стараются уклониться от своих обязанностей;
- поэтому работников нужно заставлять трудиться, манипулировать ими, угрожать и наказывать, чтобы добиться стоящих перед организацией целей;

- люди хотят быть направляемыми, стремясь к защите и избегая ответственности.

Теория Y исходит из прямо противоположных представлений:

- людям нравится работать, и для многих в труде скрыт источник удовлетворения;
- большинство работников руководствуются самодисциплиной и нуждаются в угрозах, они также заинтересованы в выполнении общих целей;

- многие из них не только избегают ответственности, но и стремятся к ней;

- способности к творчеству в решении организационных проблем присущи не только лидерам или руководителям;

- поощрение является лучшим способом для вдохновения людей к выполнению задач, стоящих перед организацией.

В соответствии с теорией Y и лидер Y должен быть чувствителен к запросам и нуждам работников, прислушиваться к их предложениям по поводу улучшения работы в организации.

Теории X и Y обозначают крайние полюса стиля лидерства, характеризующиеся в терминах авторитарного или демократического стиля.

Наиболее известные исследования в данном ключе были выполнены под руководством К.Левина и Липпета. Авторы выделили 3 основных стиля лидерства:

- 1) авторитарный стиль;
- 2) демократический стиль;
- 3) попустительский стиль.

При авторитарном (автократическом) стиле лидер принимает решения единолично, определяя и регламентируя всю деятельность подчиненных, не давая им возможность проявить инициативу.

При попустительском стиле лидер вообще избегает принимать какие-либо решения, не участвуя в этом процессе и предоставляя подчиненным полную свободу действий.

При демократическом стиле лидер вовлекает сотрудников в процесс принятия решений, используя групповую дискуссию, стимулируя их активность и разделяя вместе с ним ответственность за принятие решений.

В настоящее время предпринимаются попытки переименовать их соответственно: директивный стиль, коллегиальный стиль, разрешительный стиль.

Описанное лидерство содержит в себе две стороны: содержательную и техническую. Содержательная сторона включает в себя решения, предлагаемые лидером, а техническая – способы этих решений. Г.М. Андреева предлагает рассматривать лидерские стили, анализируя их с 2 указанных сторон, на основе следующей таблицы (табл. 3).

Таблица 3 - Содержательные и технические характеристики стилей лидерства по Левину и Липпету

Стиль	Формальная (техническая)	Содержательная сторона
-------	--------------------------	------------------------

Авторитарный	Деловые, краткие распоряжения. Запреты без снисхождения, с угрозой. Четкий язык, неприветливый тон. Похвала и порицание субъективны. Эмоции не принимаются в расчет. Показ приемов – не система. Позиция лидера – вне группы.	Дела в группе планируются заранее во всем объеме. Определяются лишь непосредственные цели, дальние – неизвестны. Голос руководителя – решающий.
Демократический	Инструкция в форме предложений. Не сухая речь, товарищеский тон. Похвала и порицание – с советами. Распоряжение и запреты с дискуссиями. Позиция лидера – внутри группы.	Мероприятия планируются не заранее, а в группе. За реализацию предложений отвечают все. Все разделы работы не только предлагаются, но и обсуждаются.
Попустительский	Тон – конвенциональный. Отсутствие похвалы и порицаний. Никакого сотрудничества. Позиция лидера – незаметно в стороне от группы.	Дела в группе идут сами собой. Лидер не дает указаний. Разделы работы складываются их отдельных интересов или исходят от нового лидера.

Г.М. Андреева подчеркивает, что любая схема не может охватить все стороны и проявления стиля лидерства, однако исследователи пытаются усложнить ее или ввести новые понятия, например, новые типы лидеров: лидер-организатор, лидер-эрудит, лидер-инициатор, лидер – генератор эмоционального настроения, лидер эмоционального притяжения, лидер-умелец. При этом проблема заключается в том, что феномен лидерства в настоящий момент описан еще неопределенно, в том числе нет четкого разделения между понятиями «лидер» и «руководитель».

С точки зрения Г.М. Андреевой, самым большим упрощением является мнение, согласно которому руководитель и лидер – обязательно одно лицо.

Что касается эффективности того или иного стиля лидерства: в группе руководимой лидером с демократическим стилем руководства, уровень общей удовлетворенности работой наивысший, также, как и стремление к творчеству, в группе лидера с авторитарным стилем продуктивность самая высокая, по сравнению с другими группами. Попустительский стиль приводит к беспорядкам и конфликтам, что сказывается на снижении объема и качества выполняемой работы. Выбор стиля лидерства во многом зависит от содержания той или иной деятельности.

В целом, предпочтительным является демократический стиль, однако, когда речь идет о выполнении «простой» работы за короткий срок силами группы со средним уровнем образования более эффективным является авторитарное лидерство. Попустительский стиль продуктивен, когда речь заходит о выполнении сложной работы с неопределенными целями, которую выполняет креативная и образованная группа.

Основу теорий лидерства Ф.Фидлера составляют следующие понятия: руководитель, «руководитель, ориентированный на задачу», «руководитель, ориентированный на межличностные отношения». При этом особую роль в анализе продуктивности того или иного лидера играет «наименее предпочитаемый сотрудник» (он особо «мешает» руководителю, ориентированному на задачу).

В рамках данной модели была разработана Шкала Измерения Предпочтений Сотрудников. Лидеры с высокими оценками характеризуются в позитивно-окрашенных тонах (имеют более близкую дистанцию в отношении с работниками, активно обсуждают

ход выполнения заданий с сотрудниками). С низкими оценками характеризуются негативно и оцениваются как ориентированные на задание. Для личностно-ориентированных лидеров оптимальным является средний, сдержанный уровень контроля. Предметно-ориентированные лидеры были гораздо эффективнее в условиях или слишком высокого, или слишком низкого уровня контроля. Стиль личностно-ориентированного лидера следует назвать демократическим, а ориентированного на результат – авторитарным или практикующим либерально-попустительский стиль.

4.3. Лидерство и руководство.

Лидер может быть одновременно руководителем группы, но может и не быть им. Руководителя целенаправленно избирают, а чаще назначают, он отвечает за положение дел в возглавляемом коллективе, обладает официальным правом поощрять или наказывать участников совместной деятельности.

Лидер выдвигается стихийно, не обладает никакими властными полномочиями, на него не возложены никакие официальные обязанности (табл. 4). Лидерство и руководство имеет много общего в области использования инструментов воздействия на людей, в процессе управления группой, в стремлении к результату.

Таблица 4 - Сравнительный анализ лидерства и руководства

Основания сравнения	Лидерство	Руководство (по Парыгину)
Социальная роль	Регуляция межличностных отношений	Регуляция официальных отношений
Сферы влияния	Микросреда	Макросреда - связь со всей системой общественных отношений
Рождение	Стихийно возникает	Назначается, выбирается и т.п.
Устойчивость	Нестабильно	Стабильно
Наличие инструментов власти	Не может применить санкций	Может применить санкции: наказать или наградить
Сфера ответственности	Непосредственное принятие решения, личная ответственность	Сложный многоступенчатый процесс принятия решения, разделение ответственности
Сфера влияния	Малая сфера действия	Широкая сфера действия

4.4. Гендерные аспекты организационного руководства и лидерства.

Лидерство – процесс, присущий как мужчинам, так и женщинам, долгое время в социокультурном пространстве и социальных практиках считался «мужской сферой».

В современном мире участие женщин в управленческой деятельности имеет тенденцию ко все большему расширению: женщины занимают руководящие позиции во множестве организаций как в некоммерческом, так и в коммерческом секторах, в том числе в больших международных корпорациях, они становятся мэрами городов, занимают посты министров и глав правительств, возглавляют государства. Много девушек возглавляет и молодежные организации — профсоюзные организации студентов, волонтерские организации и политические.

Однако если мы сравним их количество с мужчинами, которые занимают руководящие должности, женщины окажутся в явном меньшинстве. Почему так происходит? Ответ прост: гендерные стереотипы отражают распределение мужчин и женщин в нашем обществе. Эти стереотипы все еще традиционно связывают женщин с домохозяйками, а мужчин с кормильцами и добытчиками. Женщины обычно описываются

ориентированными на поддержание хороших межличностных отношений, отзывчивыми и заботливыми, а мужчины — нацеленными на результат, уверенными в себе, напористыми и более независимыми. Именно поэтому традиционные стереотипы лидерства связаны с мужскими атрибутами и ассоциируются именно с мужчинами.

Действия же женщин, которые пытаются использовать активные, целенаправленные стили руководства (т. е. «вести себя как мужчины») воспринимаются негативно. Часто это приводит к тому, что одинаково квалифицированные женщины в одних и тех же руководящих позициях уступают по сравнению с мужчинами. «Вторичность» женщин в области лидерства объясняют «объективно». Так рождаются гендерные стереотипы относительно «естественности» подчинения и нелидерства женщин.

Рассмотрим данные стереотипы.

1. Стил ь поведения.

Гендерный стереотип:

- мужской стиль – инструментальный (добытчик, кормилец, глава семьи, ответственен за дисциплину детей);
- женский – экспрессивный (хранительница домашнего очага, ориентирована на отношения, поддерживает теплый эмоциональный климат).

Фактически: Мальчики являются более личностно-ориентированными в отношении сверстников, а в отношении взрослых оба пола не отличаются по указанной ориентации. Стил ь поведения представителя определенного пола не является врожденным, а задается обществом. Это соответствует исторически сложившемуся разделению ролей.

2. Эффективность деятельности, успешность.

Гендерный стереотип: Считается, что мужчины более продуктивны, чем женщины, в осуществлении деятельности. При этом имеются в виду либо все виды деятельности, либо те, что ценятся обществом – политика, наука, искусство, спорт и т.д. Отсутствие у женщин выдающихся успехов в этих областях объясняется обычно отсутствием у них соответствующих способностей к этим видам деятельности.

Фактически: Девочки успешнее мальчиков на протяжении всего школьного периода и больше интересуются успехами. Мужчины превосходят в деятельности, выполняемой индивидуально, и при решении задач, требующих ориентации на задачу, превосходство женщин - когда требуется интеракционный стиль.

3. Мотивация достижений и отношение к наградам за деятельность.

Гендерный стереотип:

Мужчины с детства имеют: большую потребность в достижениях; особенно в областях, связанных с неодушевленными предметами; мотивированы на успех ради успеха.

Женщины: имеют меньшую мотивацию достижений; эффективнее в областях взаимодействия людей; их усилия направлены не на успех, а обусловлены желанием нравиться другим или избегать осуждения.

Фактически: Исследования, проведенные на детях, обнаружили либо отсутствие половых различий по уровню мотивации достижения, либо преимущество девочек. Однако в период взрослости меняется соотношение мотивации достижения у мужчин и женщин. Мужчины стимулируются обществом на успех, женщины либо не поощряются, либо поощряются негативно. В результате женщины отказываются от социальных достижений.

4. Конкурентность и кооперативность.

Гендерный стереотип: Мужчины нацелены на конкуренцию, а женщины – на сотрудничество.

Фактически: И мальчики и девочки демонстрируют либо одинаковую конкурентность, либо мальчики превосходят девочек по этому качеству, более остро реагируя на конкурентную ситуацию и соперников. У взрослых женщин имеется феномен «боязни успеха», успех связывается с негативными последствиями: потерей женственности, осуждением близких и т.п. Также у взрослых нет различий по параметру

сотрудничество: большая конкурентность мужчин не означает большей кооперативности у женщин.

5. Стремление к лидерству и мотивация власти.

Гендерный стереотип: Мужчинам приписывается наличие стремления к лидерству и мотивация власти, женщинам – отсутствие подобной мотивации.

Фактически: Девушки показывают больше стремлений в получении управленческих должностей в будущем, чем юноши.

6. Характеристики личности, способствующие и препятствующие лидерству

Гендерный стереотип: Мужчины воспринимаются как агрессивные и доминантные, женщины – подчиненные и тревожные.

Фактически: Женщины более тревожны, а мужчины превосходят женщин по доминантности и агрессивности, но это не дает им преимуществ в лидерстве, т.к. в современном обществе агрессия неприемлема, а на высшем уровне управления наблюдается переход к неагрессивному лидерскому стилю.

7. С каким полом связывается лидерство.

Гендерный стереотип: Лидером должен быть мужчина.

Фактически: В детстве мальчики делают больше попыток лидировать, чем девочки. В организациях менеджерские должности чаще принадлежат мужчинам, но доля женщин растет.

Существующие гендерные стереотипы лидерства не всегда подтверждаются при изучении лидеров. Встречающаяся меньшая эффективность женщин-лидеров может объясняться: а) влиянием стереотипов; б) мотивацией; в) отношением общества; г) ролевым конфликтом у женщины; д) методологическими проблемами исследования.

4.3. Методика формирования команды. Организация межличностных, групповых и организационных коммуникаций.

Команда – это небольшая группа людей, стремящихся к достижению общей цели, постоянно взаимодействующих и координирующих свои усилия. Работа в командах является средством повышения эффективности деятельности организации. Огромное количество информации, быстро изменяющаяся внешняя среда, конкуренция и т.п. – все это затрудняет процесс управления в организации в целом. В этих условиях большое значение имеет использование команд, которые более гибки и мобильны, и способны быстро реагировать на сигналы, посылаемые окружением.

Чтобы деятельность команд была эффективной, необходимо обеспечить разнообразие ролей в команде:

1. Специалисты по решению задач. Их роль - достигать целей, стоящих перед командой. Черты, характерные для членов команды, играющих эту роль:

Инициатива: предлагают новый взгляд на проблемы, новые способы решения проблем.

Обмен мнениями: организуют дискуссию среди членов команд по поводу решения проблем, оценивают полученные идеи.

Поиск информации: отбирают материалы и факты, имеющие отношение к проблеме.

Подведение итогов: из возможных идей по решению проблем, возможностей и ограничений выстраивают целостную картину.

Энергия: побуждают членов команды к активности.

2. Члены команды, осуществляющие социально-эмоциональную поддержку. Их роль состоит в удовлетворении эмоциональных потребностей членов команды. Черты, характерные для членов команды, играющих эту роль:

воодушевляют: напоминают о прошлых успехах, высказывают комплименты и похвалы.

создают гармонию: сглаживают споры и конфликты.

- снижают напряженность: шутят, рассказывают анекдоты и т.п.
- готовы к компромиссам: способны поступиться собственным мнением ради поддержания гармонии в команде.

Если большинство членов команды склонны к исполнению этой роли, то члены команды получают высокое индивидуальное удовлетворение, но, как правило, за счет снижения эффективности действия.

Если же большинство членов команды склонны к исполнению роли «специалистов по решению задач», то такая команда оказывается очень эффективна, но только в течение короткого отрезка времени, однако в долгосрочной перспективе у членов таких команд снижается степень удовлетворения от работы, и, следовательно, снижается эффективность.

3. Члены команды, играющие двойную роль. Такие люди совмещают в себе две вышеописанные роли: выполняют поставленную перед командой задачу и удовлетворяют эмоциональные потребности членов команды. Обычно люди, способные играть двойную роль, становятся лидерами команд.

4. Члены команды, играющие роль стороннего наблюдателя. Такие люди обычно держатся отстраненно от повседневной жизни команды, активно не участвуют ни в решении задач, ни в создании положительного эмоционального климата. Однако такие люди очень полезны в критический момент, поскольку видят проблемы команды как бы «со стороны» и часто дают нетривиальную «обратную связь».

Руководители не должны забывать, что команда должна быть хорошо сбалансирована, в ней должен присутствовать весь «спектр» ролей.

Согласно другой классификации, помимо содержательных оснований (конкретные специальности, опыт, квалификация и т.п.), в команде должны быть представлены определенные типы людей:

- Доводящий до конца. Как правило, об успехе команды судят по окончательным результатам ее работы. Доводящие до конца завершают все, что начинают и неохотно предпринимают что-либо там, где есть сомнения по поводу того, что удастся довести дела до завершения. Они заботятся о завершении намеченного и настаивают на этом даже тогда, когда энтузиазм всех остальных членов команды уже исчерпан. Их присутствие не дает команде тратить время впустую на проекты, которые не могут быть доведены до конца.

- Возмутитель спокойствия. Возмутители спокойствия всегда выступают как побудители к действию, и если команда склонна к бездействию или самодовольству, то присутствие Возмутителя спокойствия выведет ее из этого состояния.

- Действующий. Основным качеством Действующих, отражающим их установки и характер, является дисциплинированность. Будучи дисциплинированными по своей сути, они упорядочено подходят к любой поручаемой им работе. Среди их отличительных качеств также следует назвать: организованность, сознательность, приверженность обязательствам, серьезное отношение к любому делу, надежность, практичность, терпимость к окружающим.

- Коллективист. Представители этой роли оказывают «смягчающее» воздействие на команду: их присутствие улучшает моральный климат и повышает степень сотрудничества между членами команды.

- Мыслитель. Основное назначение Мыслителя в команде – привнесение новых и оригинальных идей. Как правило, Мыслители действуют в одиночку, обдумывая различные варианты. Им свойственен самоуглубленный, аналитический подход к решению проблем.

- Оценивающий. Представители этой роли ярко не проявляют себя в команде до тех пор, пока не приходит время принятия важных решений. Представители этой роли в команде обладают высоким интеллектуальным уровнем, высокими показателями критичности мышления, особенно это касается их способности выдвигать контраргументы.

- Председатель. Основной залог успеха Председателя – его личностные качества. Прежде всего, по своей натуре он склонен доверять людям и принимать их такими, какие они есть, без проявлений ревности или подозрительности. В качестве противовеса этому

качеству он должен уметь доминировать в команде и быть приверженным целям и задачам команды, что усиливает и морально обосновывает его доминирующую позицию. Председатель – это хороший лидер для сбалансированной команды, перед которой стоят сложные и многогранные проблемы, требующие эффективного распределения ролей в команде.

- Исследователь ресурсов. Это еще один член команды, ориентированный на предложение новых идей. Однако, способ генерации идей Исследователями ресурсов и сам характер предлагаемых ими идей отличны от Мыслителей. Они склонны не столько сами предлагать оригинальные идеи, сколько «подбирать» фрагменты идей окружающих и развивать их. Исследователи ресурсов особо искусны в изучении ресурсов за пределами команды.

Этапы развития команды:

□ Формирование. Члены команды обмениваются информацией, узнают и принимают друг друга, формулируют задачи группы. Преобладает аура вежливости, взаимоотношения членов группы отличаются осторожностью. Все члены будущей команды «размахивают визитками», то есть стараются подчеркнуть свои прошлые мнимые и реальные заслуги перед человечеством.

□ Смятение. Наступает момент, когда большинству становится «ясно», кто есть кто, и они начинают самоопределяться в команде. Члены группы конкурируют за обладание более высоким статусом, за относительное влияние, дискутируют о направлениях развития. Группа испытывает внешнее давление, между ее участниками складываются достаточно напряженные отношения. На сцену выходят лидеры «первой волны». Они уверены в себе, опытные, настойчивы, громко говорят и всегда знают «точно», что надо делать.

□ Нормирование. Сформированная команда начинает совместное движение к общим целям, устанавливаются равновесие конкурирующих сил и групповые нормы, определяющие поведение ее членов, сотрудничество членов команды становится все более эффективным. В это время может произойти смена лидеров, на сцену могут выйти лидеры «второй волны». Они внешне менее эффективны, но люди чувствуют себя с ними более уверенными и раскованными.

□ Выполнение работы. Команда вступает в стадию зрелости, она способна решать самые сложные задачи, каждый ее член исполняет несколько функциональных ролей. На этом этапе команде присущи все те качества, которые мы сформулировали в виде списка тринадцати характеристик.

□ Расформирование. Рано или поздно расформируются (или трансформируются) самые успешные команды, интенсивные социальные отношения их участников постепенно сходят на нет. Но это скорее оптимистичное наблюдение, нежели пессимистичное. Люди устают друг от друга, поэтому возникающие новые крупные проекты и идеи (если это подлинно инновационная организация), собирают людей под новое знамя, предлагая иные конфигурации отношений и новые вызовы времени.

В организациях различают две крупные группы коммуникаций – формальные и неформальные коммуникации. Формальные коммуникации осуществляются между элементами формальной структуры организации – межуровневые коммуникации (нисходящие и восходящие), горизонтальные коммуникации (между подразделениями одного уровня в иерархии организации), коммуникации «руководитель – подчиненный», «руководитель – рабочая группа». Неформальные коммуникации связаны с неформальными группами и неслужебными вопросами, а также с распространением слухов о служебных вопросах. Обычным средством осуществления формальных коммуникаций является письменная и устная речь. При осуществлении коммуникаций следует принимать во внимание наличие «барьеров непонимания» (семантического, стилистического, логического, фонетического, барьера авторитета и др.) и овладевать методами их преодоления. В неформальных коммуникациях силен эмоциональный фактор.

Общее для всех типов коммуникаций внутри организаций – это то, что в организациях они имеют место между людьми, являющимися исполнителями разнообразных ролей, представителями разнообразных групп интересов, живыми сложными большими системами, обладающими всеми общесистемными свойствами и массой уникальных, индивидуальных свойств, качеств, особенностей. Отношения между организациями опосредованы людьми, представляющими организацию и выражающими ее интересы, поэтому и здесь важен человеческий фактор, который необходимо учитывать и проявлением которого следует управлять. Коммуникации в организациях имеют место между людьми, являющимися исполнителями разнообразных ролей, представителями разнообразных групп интересов, живыми сложными большими системами, обладающими всеми общесистемными свойствами и массой уникальных, индивидуальных свойств, качеств, особенностей.

Межличностные коммуникации - это коммуникации, которые осуществляются преимущественно на психическом уровне, под сильным влиянием эмоционального аспекта, в основе психология и социальная психология.

Групповые коммуникации - коммуникации внутри групп, межличностные и статусные, формальные, ролевые, и между группами, отражаемые в общении личностей, персонифицирующих или представляющих группы, а также личностей, идентифицирующих себя с группами. Это коммуникации, происходящие в переплетении формальных и неформальных, рациональных и эмоциональных, социальных и служебных и т.п. отношениях.

Коммуникации внутри групп связаны с достижением групповых результатов и трудовыми отношениями, но испытывают влияние межличностных отношений. Эффективные групповые коммуникации в решающей степени способствуют превращению группы в продуктивно работающую команду. Развал команды зачастую является следствием нарушения нормальных коммуникаций, усиления и преобладания личностных неприязненных отношений над деловыми.

Организационные коммуникации – это либо только формальные связи, определяемые административными актами, официальными организационно-распорядительными документами, т.е. связи, необходимые для исполнения организационных заданий, закрепления разделения и кооперации труда, его коммуникативного обеспечения, либо вся совокупность коммуникаций всех видов в организациях и между ними.

Вопросы для повторения:

1. В каких случаях уместно авторитарное лидерство? Какие плюсы оно имеет?
2. Выделите основные элементы трансформационного лидерства. Какие из них позволяют организации лучше развиваться и реагировать на изменения и подвижность внешней среды?
3. Дайте определение команде.
4. Перечислите и опишите командные роли.
5. Какова роль лидера в команде?
6. Раскройте содержание этапов развития команды.
7. Раскройте понятие коммуникация в организации.
8. Перечислите и охарактеризуйте этапы развития команды.

Раздел 2. Педагогика высшей школы

Тема 5. Педагогика высшей школы.

5.1. Предмет, задачи, категории педагогики высшей школы.

5.2. Принципы и методы педагогического исследования.

5.1. Предмет, задачи, категории педагогики высшей школы.

«Педагогика высшей школы – область знания, выражающая основные научные идеи, дающие целостное представление о закономерностях и существенных связях в учебно-познавательной, научной, воспитательной, профессиональной подготовке и всестороннем развитии студентов»

В первую очередь, нужно отметить, что педагогика высшей школы – это отрасль, раздел общей педагогики, а точнее будет сказать, профессиональной педагогики, изучающей закономерности, осуществляющей теоретическое обоснование, разрабатывающей принципы, технологии воспитания и образования человека, ориентированного на конкретно-профессиональную сферу действительности.

Предметом изучения педагогики высшей школы является лишь один этап в профессиональном становлении – процесс обучения и воспитания специалистов с высшим образованием.

Таким образом, будем понимать под педагогикой высшей школы – отрасль (раздел) общей (профессиональной) педагогики, изучающую основные составляющие (закономерности, принципы, формы, методы, технологии, содержание) образовательного процесса в вузе, а также особенности и условия (требования к процессу взаимодействия преподавателя и студента, требования к личности преподавателя и студента и др.) эффективного осуществления профессиональной подготовки будущего специалиста.

Приведем задачи профессиональной педагогики, которые можно отнести к задачам педагогики высшей школы как общее к частному. В них входят:

1. Разработка теоретико-методологических основ профессионального образования и методик проведения исследований в профессиональной педагогике.
2. Обоснование сущности, аспектов и функций профессионального образования.
3. Изучение истории развития профессионального образования и педагогической мысли.
4. Анализ современного состояния и прогнозирование развития профессионального образования в нашей стране и за рубежом.
5. Выявление закономерностей профессионального обучения, воспитания и развития личности.
6. Обоснование образовательных стандартов и содержания профессионального образования.
7. Разработка новых принципов, методов, систем и технологий профессионального образования.
8. Определение принципов, методов и способов управления профессионально-педагогическими системами, мониторинга профессионально-образовательного процесса и профессионального развития обучающихся.

Кроме этого можно выделить задачи педагогики высшей школы в практической области:

1. Формирование у преподавателей высшей школы умений и навыков методически обоснованного проведения всех видов учебной, научной и воспитательной работы.
2. Установление связи обучения, профессиональной подготовленности и формирование у студентов устойчивых навыков проведения исследовательской работы на основе этой связи.
3. Преобразование учебного процесса в процесс развития самостоятельного, творческого мышления.

4. Формирование, развитие, проявление педагогического мастерства с целью мобилизации студентов на разнообразные творческие действия.

5. Анализ социально-педагогического фактора, законов и особенностей формирования у студентов педагогических знаний, умений, навыков, педагогического сознания.

6. Вооружение педагогов психологическими знаниями.

Использование содержания педагогики высшей школы в качестве программы действий по организации и проведению многообразных видов педагогической деятельности.

К категориальному аппарату педагогики высшей школы, помимо общепедагогических, можно отнести профессионально-педагогические категории, такие как:

Профессиональное образование – процесс и результат профессионального развития личности посредством научно-организованного профессионального обучения и воспитания.

Профессиональное обучение – процесс и результат овладения обучающимися профессиональными знаниями, умениями и навыками.

Профессиональное воспитание – процесс и результат формирования профессионально важных качеств (различают общие и специальные ПВК).

Профессиональное развитие – развитие личности как субъекта профессиональной деятельности.

Профессиональное становление – результат профессионального развития: разряд, категория, класс, должность, степень, звание и др.

5.2. Принципы и методы педагогического исследования.

Научные исследования в области педагогики высшей школы представляют собой специфический вид познавательной деятельности, в ходе которой с помощью разнообразных методов выявляются новые, прежде не известные стороны, отношения, грани изучаемого объекта.

Любое научное исследование осуществляется в соответствии с теми или иными методологическими установками. Методология характеризует подход исследователя к анализу действительности. Она проявляется в его замысле, методике и результатах.

По своему характеру и содержанию исследования в области педагогики высшей школы разделяются на фундаментальные, прикладные и разработки.

Фундаментальные исследования призваны разрешать задачи стратегического характера. Их основными отличительными признаками являются: теоретическая актуальность, выражающаяся в выявлении закономерностей, принципов или фактов, имеющих принципиально важное значение, концептуальность, историзм, критический анализ научных положений и установок, новизну и научную достоверность полученных результатов. Однако главным критерием является решение перспективной задачи подготовки специалистов, а также те теоретические выводы, которые вносят серьезные изменения в логику развития самой науки.

Основными признаками прикладных исследований являются: приближенность их к актуальным запросам практики; сравнительная ограниченность выборки исследования; оперативность в проведении и внедрении результатов и др. Прикладные исследования опираются на исследования фундаментальные, которые вооружают их общей ориентацией, теоретическими и логическими знаниями, помогают определить наиболее рациональную методику исследования. В свою очередь, прикладные исследования дают ценный материал для фундаментальных исследований.

К разработкам в педагогике высшей школы относят методические рекомендации по тем или иным вопросам обучения и воспитания студентов, инструкции, методические средства и пособия. Они опираются на прикладные исследования и передовой

педагогический опыт. Отличительными чертами являются: целеустремленность, конкретность, определенность и сравнительно небольшой размер.

Специфический вид научно-педагогического исследования - изучение, обобщение и внедрение в практику передового опыта обучения и воспитания. Особенность исследований состоит в том, что они, как правило, вплетены в конкретную практику вузов и доступны каждому преподавателю.

Педагогический опыт - явление многоплановое, динамичное, противоречивое. С одной стороны - это живой, реальный педагогический процесс, повседневная педагогическая практика, а с другой - отражение этого процесса в сознании педагога - система его знаний, навыков и умений, привычек и личностных качеств, приобретенных в процессе обучения и воспитания людей.

В совершенствовании педагогического процесса вуза важная роль принадлежит передовому современному опыту и опыту педагогов-новаторов. Его отличают: высокая действенность, стабильность результатов, репрезентативность и перспективность. Он создается, как правило, на основе новой педагогической идеи и вместе с тем порождает новую идею. Главной целью изучения передового опыта учебно-воспитательной работы в учебных заведениях является повышение эффективности педагогического процесса, достижение более высокого уровня подготовки специалистов.

Применительно к исследованиям в области педагогики высшей школы системный подход предполагает изучение педагогического процесса вуза или его элементов как целостной системы. Исследуемая система, в свою очередь, состоит из множества элементов, каждый из которых сам является сложной системой. Между элементами системы существуют сложные связи и зависимости, совокупность которых составляет структуру данной системы. Система в целом и ее элементы выполняют различные функции, определенным образом связанные с ее структурой и внешним миром. Системный подход дает возможность моделировать изучаемые явления и исследовать их в состоянии развития и в разных условиях.

Системный подход предполагает многоуровневое и многоплановое изучение объекта, в ходе которого строится не одна, а ряд его моделей, отражающих объект на разных уровнях и срезам. При этом возможен синтез этих моделей в новой, целостной, обобщающей модели. Как показывает практика, диапазон системного подхода в педагогических исследованиях широк. Он позволяет разрабатывать как общие проблемы педагогики, так и многие вопросы методики воспитания и дидактики.

Исторически в педагогике сложились и активно используются следующие методы: теоретический анализ, наблюдение, эксперимент, беседа, письменный опрос (анкетирование), изучение результатов деятельности обучаемых и воспитуемых, педагогической документации, сравнительно-исторический метод. В педагогических исследованиях начали применяться математические и кибернетические методы. Имеются новые методы такие как: аналогия, формализация, моделирование.

Методология требует, чтобы методы исследования были адекватны цели и содержанию предмета науки. Особо следует сказать о математических методах и формализации. Их применение в педагогических исследованиях в соответствии с требованиями методологии и учетом своеобразия педагогических явлений оправданно и дает положительные результаты.

Педагогические исследования складываются из нескольких этапов. Основными из них являются:

- выбор и обоснование темы, предварительная разработка замысла и рабочего плана исследования;
- изучение литературных источников, соответствующей документации, предварительное ознакомление с опытом, уточнение условий обстановки, построение гипотезы, формулировка задач, разработка методики исследования;

– изучение педагогической практики, сбор фактического материала в целях проверки гипотезы;

- теоретический анализ добытого фактического материала;
- проверка выводов и рекомендаций;
- оформление результатов исследования;
- внедрение результатов исследования в повседневную практику.

Методика исследования педагогического процесса в целом и его отдельных сторон всегда индивидуальна. Она определяется замыслом, характером исследования, зависит от условий и имеющихся средств для намеченной работы. В ходе исследования она непрерывно совершенствуется.

Для педагогики высшей школы в настоящее время особенно важны экспериментальные исследования, позволяющие активно вмешиваться в изучаемое явление, вносить новые элементы в педагогический процесс в целях его совершенствования. Педагогическому эксперименту присущи следующие черты:

- преднамеренное, строго продуманное внесение в изучаемое явление чего-то принципиально нового в соответствии с задачами исследования и в целях проверки гипотезы;

- организация учебно-воспитательной деятельности, позволяющей видеть связи между явлениями и их взаимовлияние;

- проверка и контроль эксперимента, его сравнение с другими экспериментами, решающими аналогичную задачу с иных позиций;

- систематическая проверка количественных и качественных изменений, проведение контрольных срезов и, если требуется, внесение корректив; использование при необходимости вариационной статистики;

- объективный количественный и качественный анализ полученных результатов, теоретические обобщения, научные выводы и рекомендации.

Эффективность педагогического эксперимента состоит в том, что он позволяет создавать новый опыт в точно учитываемых условиях. Его успех во многом зависит от обстоятельной разработки и оригинальности гипотезы, от умелого выбора экспериментальных и контрольных групп (учебных групп, курсов, факультетов, вузов), четкости планирования, неуклонного учета всех основных условий и факторов, проверяемых в исследовании, их влияния на ход и результаты обучения, воспитания, развития человека. С особой тщательностью следует провести сравнение и сопоставление учебной деятельности и результатов экспериментальных и контрольных групп.

Результаты исследования оформляются в виде отчета о научной работе, научного доклада, статьи, брошюры, диссертации или монографии.

Оформлением результатов исследования не заканчивается работа над выдвинутой проблемой. Это должно быть внедрение обоснованных рекомендаций в повседневную практику вузов.

Таким образом, научные исследования в области педагогики высшей школы осуществляются в соответствии с методологическими принципами многими методами. Важнейшие требования методологии - постоянное стремление к внедрению результатов исследований в повседневную деятельность вузов и этим самым дальнейшее повышение качества подготовки будущих специалистов.

Вопросы для повторения:

1. Что составляет методологические основы педагогики высшей школы?
2. Раскройте сущность и содержание основных методов, используемых в педагогике высшей школы.
3. Раскройте сущность и содержание понятий «методология» и охарактеризуйте ее основные уровни.
4. Назовите предмет изучения педагогики высшей школы.

5. Каковы основные направления исследований в области педагогики высшей школы?

Тема 6. Приоритетные стратегии и тенденции развития высшего образования.

6.1. Современные стратегии модернизации высшего образования в России и за рубежом

Образование является одним из основных системообразующих институтов общества, реализующих широкий спектр общественно-значимых функций и находящихся под влиянием происходящих общественных трансформаций. Особенностью образования в современном мире является то, что оно одновременно выступает одним из самых консервативных институтов, сохраняющих и воспроизводящих традиционные формы и отношения, а с другой, – оно все более становится центром воспроизводства наиболее значимых инноваций и передовых практик, определяющих перспективы развития общества. Ряд глобальных трендов развития современного образования определяется общими мировыми тенденциями и находится под влиянием мировых общественных проблем.

Глобализация образования. Образование встроено в процесс всемирной экономической, политической, культурной интеграции и унификации, развертывающийся в последние десятилетия во всем мире. Проявлением этого является всеобщая унификация знания, в результате чего происходит выход национальных образовательных систем за пределы государственных границ, интернационализация образования и формирование единого мирового образовательного пространства и рынка образовательных услуг. Глобализация образования проявляется в гармонизации страновых систем образования между собой, унификации уровней образования и квалификационных рамок, открытости и трансграничности образования, возможности получать его из любой точки мира. В последнее время новый мощный импульс расширению глобализации образования дают информационные технологии и цифровизация образования, разрушающие национальные границы образования в принципе, и позволяющие говорить о формировании единого мирового цифрового образовательного пространства, определяющего новые конкурентные условия для всех игроков образовательного рынка.

Массовизация образования. Массовизация образования стала глобальным трендом образования в последние пятьдесят лет в связи расширением социальных функций государства, обеспечившего доступ к нему широких слоев населения, что привело к превращению образования из элитного в массовое. Влияние может оказать обратная сторона массовизации образования, выражающаяся в снижении его качества, определенной дискредитации образования, особенно более высоких ступеней, что уже приводит к снижению спроса на высшее образование среди населения разных стран. Это может стать угрозой финансовой стабильности университетов. Поэтому можно предположить, что новым драйвером станет идея «нового элитного» образования, которое будет ориентироваться на ограниченный круг людей и вернет принцип элитарности в школы и вузы, а реализовываться он будет на уникальном экспертном уровне очного (оффлайн) образования.

Демократизация образования. Демократизация образования проявляется в реализации и расширении прав каждого человека на образование, возможностей для самоорганизации и права выбора обучающихся и обучающихся в образовательном процессе, поливариативности способов образовательной деятельности, многообразии образовательных систем и форм получения образования. Важным проявлением демократизации образования во всем мире является сокращение государственных функций в регулировании образования, развитие общественного управления, самоуправления и автономии образовательных организаций. Одной из современных форм демократизации образования явилось появление феномена массовых открытых образовательных курсов, которые выложили ведущие университеты мира на открытых цифровых платформах для

широкого пользователя без всяких ограничений. Это позволило университетам преодолеть все институциональные границы, существовавшие в образовании, продвинуть себя в мировом образовательном пространстве, а образовательный контент сделать максимально доступным из любой точки мира любому пользователю.

Технологизация образования. Технологизация образования, вылившаяся в настоящее время в «цифровую революцию», стала ведущим трендом развития образования. Сначала информационные, а теперь цифровые технологии кардинальным образом изменили образовательный ландшафт и конфигурацию, способствовали появлению новых сущностей в образовании. Из межличностного коммуникативного процесса оно, по сути, превратилось в технологический процесс, зависимый от использования развивающихся стремительными темпами информационных технологий. За последние несколько лет возникли принципиально новые образовательные онлайн-проекты. Также появился ряд проектов и платформенных решений в таких областях как управление учебным процессом, оценка и сертификация результатов обучения, социальные сети для преподавателей и студентов, исследователей и работодателей и т.д.

Образование в современной экономике рассматривается не как затратная сфера наряду с социальной помощью, пенсионной системой, госаппаратом, обороной и безопасностью, а как инвестиционная сфера, определяющая темпы и качество экономического роста. Значение человеческого капитала еще более возросло в XXI веке в условиях увеличения роли знаний и инноваций в экономике и усиления неопределенности. Поэтому в последние десятилетия ключевым элементом человеческого капитала становится интеллектуальный капитал, который представляет собой способность генерировать и осваивать инновации. Он приобретает характер решающего фактора для модернизации экономики, перехода к новым технологическим укладам и для ответа на вызовы глобальной конкуренции. Растет спрос населения на высшее образование, что также способствует росту инвестиций в данную сферу. Определенным отражением этого является рост стоимости высшего образования.

Непрерывность и пожизненность образования. Среди общемировых тенденций образования особо следует выделить быстрое развитие непрерывного образования (образования на протяжении всей жизни). Необходимость этого обусловлена не только ускорившимися процессами технико-технологического и информационного прогресса, но и особенностями социально-экономического и демографического развития. Институционализация непрерывного образования осуществляется в разных формах и на разных уровнях. Оно формируется как в вертикальной (образование по уровням в течение всей жизни), так и в горизонтальной (параллельное обучение на программах разного уровня, самообразование) плоскостях. Реагируя на изменяющиеся потребности рынка труда, сфера образования все более приобретает многоуровневый и многоформатный характер. Особо бурное развитие получает неформальное образование, которое активно начинает конкурировать с формальным образованием, а дополнительное с основным. Для университетов появляется угроза оказаться в арьергарде этих процессов в том случае, если не удастся диверсифицировать спектр реализуемых образовательных услуг и выстроить у себя систему непрерывного образования для различных сегментов рынка труда и образовательных потребностей граждан разных возрастов на основе современных цифровых технологий. Поэтому вузы вынуждены не просто своевременно реагировать на эти процессы, а обеспечивать опережающие реакции, создавая конкурентоспособные условия для обучения на протяжении всей жизни. Растущая интернационализация одинаково характерна как для школьного, так и для университетского образования во всем мире. Студенты из-за рубежа – это наиболее мобильные молодые люди, финансово обеспеченные и обладающие большими способностями и талантом, что позволяет им поступать в ведущие вузы мира. Именно за таких людей сегодня развернулась серьезная мировая конкуренция, которую зачастую ведут уже не отдельные университеты, а образовательные консорциумы или даже страны. На этом фоне усиливается роль

международных образовательных стандартов и рейтингов, как в школьном образовании (PISA, PIRLS, TIMSS — наиболее известные международные системы оценки навыков школьников), так и в высшем образовании (TOEFL, рейтинги мировых университетов, Болонский процесс и система унификации результатов образования). Указанные глобальные тренды развития образования определяют общий контекст, в котором разворачиваются основные тенденции развития высшего образования в мире и в России.

Основные направления стратегического развития ведущих университетов мира.

В настоящее время под влиянием глобальных трендов развития образования активно идут процессы трансформации высшей школы. В условиях нарастающей конкуренции на мировом образовательном пространстве перед университетами мира встают более серьезные задачи, чем раньше. Они вынуждены конкурировать не только в учебной и научной работе, но и в сфере создания инноваций, влияния на экономический рост, в решении основных мировых проблем. В соответствии с этим выстраиваются основные направления стратегического развития ведущих университетов мира, которые находят отражение в их стратегических документах, программах развития, модельных решениях. В качестве наиболее значимых: Эпоха гринфилда в образовании. Исследование SEDeC. Центр образовательных разработок Московской школы управления Сколково.

Глобализация университетского образования и науки на основе цифровизации. Формирование глобального образовательного электронного пространства, создание нового типа открытых информационных ресурсов «без границ», наиболее известным из которых является Coursera (Free Online Courses From Top Universities) – это самый значимый инновационный тренд трансформации высшего образования в мире, меняющий представление о возможностях и формате деятельности университета в целом. По существу, это – одна из первых информационно-образовательных моделей реализации идеи мета-университета (Meta University) – глобальной сети консорциума университетов и корпораций. Целью такого университета должно стать решение глобальных проблем, выходящих за рамки региональных и национальных приоритетов – здравоохранения, экологии, международного взаимодействия и ряда других. ИКТ-форматы все в большей степени становятся реальностью образовательного процесса в западной традиции.

Коммерциализация научных идей. Современные ведущие университеты располагают собственными центрами трансфера технологий, технопарками, в частности, предоставляющими студентам возможности организации стартапов, малых фирм с целью разработки технологической продукции, ее последующего патентования и маркетингового продвижения. Фактически студенты создают рабочее место сами в кампусах университетов, совмещая исследовательскую деятельность с образовательным процессом и бизнесом. Так происходит постепенная и преимущественная приватизация университетской системы бизнесом, все больше превращающим научные технологии в коммерческий продукт. Что касается юникорнов (unicorn – англ. единорог) - стартап, оценка рыночной стоимости которого превышает 1 млрд дол.), то по оценкам экспертов Стэнфордский университет - самое популярное место, где учились основатели юникорнов - 63 лучших предпринимателя мира.

Личностно-ориентированное обучение студентов. Данная стратегия означает диверсификацию образования путем обеспечения индивидуального подхода к обучающемуся с целью раскрытия потенциала каждого студента. Образовательное учреждение все в большей степени соответствует электронному каталогу товаров и услуг, где можно одним кликом мыши выбрать необходимый образовательный контент. Таким образом, у студентов есть возможность индивидуального выбора образовательной программы.

Служение обществу. Направлено на построение партнерских отношений между подразделениями университета и за пределами кампуса, создание атмосферы инклюзивности, благополучия и вежливости. В рамках направления могут реализовываться следующие инициативы: - стратегическое корпоративное и общественное взаимодействие:

разработка эффективных моделей взаимодействия с бизнесом и сообществом, гражданами; использование отношений с корпоративными партнерами для максимального создания рабочих мест; - здоровье и оздоровление: повышение участия в инициативах, которые направлены на поддержание физического и психического здоровья научно-педагогического сообщества, персонала, студентов университета; построение взаимовыгодных деловых партнерств для борьбы с неравенством в отношении оздоровления местного населения; - осуществление положительного влияния на мир: постоянное совершенствование процесса оказания услуг, мониторинг расширения спектра услуг студентам, преподавателям, населению; рост вовлеченности выпускников в деятельность университета (Университета Говард, США).

Повышение производительности труда и эффективности затрат. Направление ставит своим результатом повышение эффективности и результативности деятельности университета за счет вложения инвестиций в обновленные технологии и системы в целях содействия автоматизации процессов. Основной задачей является эффективная работа на всех уровнях организации. В рамках данного направления могут реализовываться следующие инициативы: - операционное превосходство и соответствие: оптимизация ключевых процессов, процедур, ресурсов университета, сокращение времени завершения основных процессов на 30%; регулярное проведение мониторинга эффективности и результативности административных, академических процессов для обеспечения соответствия со стандартами качества, установленными нормами; - инфраструктура университета и её устойчивость: снижение потребления электроэнергии на 20%; определение приоритетов в сфере строительства и реконструкции зданий университета с учетом его научно-технического развития; - обслуживание клиентов и взаимодействие: проведение обучения по обслуживанию клиентов в рамках всего университета в целях улучшения качества обслуживания студентов, сотрудников и внешних клиентов; вовлечение всех сотрудников университета в общественную жизнь, включая обсуждение новых идей, процессов развития университета, повышения производительности труда, эффективного управления активами университета (Университет Говард, США).

Достижение финансовой устойчивости. Направлено на достижение финансовой устойчивости университета за счет диверсификации доходов, оптимизации процессов, обеспечения прозрачной отчетности и надежного управления персоналом. Целью является повышение качества подготовки финансовой отчетности, улучшение финансового состояния университета. В рамках данного направления развития могут реализовываться следующие инициативы: - стратегическое планирование долгосрочных инвестиций; - увеличение эндаумент-фонда в целях роста поддержки студенческих стипендиальных программ и инвестиций в инфраструктуру; - сбор средств (фандрайзинг): создание инфраструктуры для устойчивой корпоративной и частной благотворительности в целях проведения активной и успешной кампании по накоплению капитала; сбор средств согласно основным приоритетам и целям университета; повышение уровня участия выпускников университета в сборе средств; увеличение числа частных доноров на 10% ежегодно (Университет Говард, США).

Финансирование университета в биткоинах. Университет Никосии на Кипре стал первым в мире учебным заведением, принимающим оплату за обучение в криптовалюте (2013 г.). По мнению финансового директора университета Христоса Влахоса, биткоины позволят облегчить распространение финансовых услуг в тех регионах мира, где не хватает развитой банковской сети. Согласно информации с официального сайта Университета Никосии, оплата в биткоинах осуществляется и в настоящее время. Некоторые американские вузы 12–18 месяцев назад начали приобретать криптовалюту на различных биржах, включая Coinbase. Зачастую средства поступают в университетские фонды в форме благотворительных пожертвований. На них финансируется образовательная и исследовательская деятельность. Также цифровые активы используются в качестве инвестиций. Самый крупный фонд оказался у Гарвардского университета.

Рост рынка управления онлайн-программами университетов. Аутсорсинг в сфере управления онлайн-образованием. OPM (Online Program Management) модель. Развитие онлайн-программ и курсов университетами, растущее внедрение технологий, увеличение количества мобильных устройств являются одними из основных факторов, которые, как ожидается, будут стимулировать рост рынка управления онлайн-программами. Все больше университетов используют бизнес-модель OPM. Сейчас можно отметить резкий рост числа университетов, которые начали передавать управление своими онлайн-программами сторонним поставщикам. Фактически онлайн управление программами (OPM) — это набор услуг для образовательного учреждения от зачисления учащегося до маркетинга, от повышения осведомленности потенциальных студентов до их зачисления на первый семестр в университете и удержания их на протяжении всего периода обучения. Что касается высшего образования, учреждения, которые сосредоточены на разработке или расширении своих онлайн программ, должны принять твердое решение о том, использовать ли стороннюю организацию OPM (Online Program Management) или пытаться создать и реализовать собственными силами. Наряду с бизнес-моделью, в которой операторы управления программами занимаются всеми аспектами разработки и реализации программ, появилась альтернатива «а ля карт», позволяющая учреждениям выбирать компоненты управления программами для передачи на аутсорсинг. Стратегические задачи и передовые практики ведущих университетов мира позволяют определять основные ориентиры развития высшего образования на глобальном и национальных уровнях.

3. Тенденции развития российского образования в сравнении с мировыми образовательными системами. Общий уровень образования населения России В современном мире экономическое благосостояние страны и ее граждан тесно связано с образованием. Это подтверждается тем, что между величиной валового внутреннего продукта (ВВП) на одного жителя и долей населения со средним профессиональным и высшим образованием, что соответствует третичному образованию по Международной стандартной классификации образования (МСКО 11), имеется статистически значимая положительная связь. Население Российской Федерации — одно из наиболее образованных в мире, уступая Ирландии и Канаде.

Третичное образование. Интересная особенность была отмечена в нашей стране: в среднем, молодежь в нашей стране получают образование раньше, чем сверстники за рубежом. При этом, «...если охват образованием населения младшей возрастной группы в России значительно превышает показатели развитых стран, то начиная с 23-24 лет мы уже отстаем от средних значений..., а в возрасте 27-28 лет доля студентов в общей численности российского населения опускается на уровень экономически менее развитых, чем Россия, стран». Отсюда, возникают проблемы при трудоустройстве выпускников. Еще одно существенное отличие участия российских молодых людей в третичном образовании от сложившихся в мире моделей заключается в том, что значительная - бóльшая, чем в других странах, - их часть учится по программам СПО, готовящим специалистов среднего звена (эквивалент «коротких программ» третичного образования МСКО). Кроме того, в нашей стране гораздо выше доля студентов, обучающихся по неочной форме.

Структура выпуска по программам. В структуре выпуска по программам третичного образования обращает на себя внимание значительное преобладание в России выпускников по инженерным специальностям. Можно предположить, что это реакция на существенное снижение выпуска по промышленным специальностям. Еще одно направление, по которому доля выпускников в России превышает средние значения по развитым странам, — услуги. По всем остальным областям знаний доля выпускников в России ниже. Особенно велико это отставание в таких областях, как естественные науки и математика, здравоохранение и социальная защита, искусство и гуманитарные науки.

Привлекательность российского образования. На долю России приходится 6,8% мирового рынка третичного образования (процент иностранных студентов, обучающихся в стране в общей численности иностранных студентов в мире). По этому показателю наша

страна уступает только США, Соединенному Королевству и Австралии. Но если рассматривать уровни третичного образования по отдельности, то обнаруживается, что высокие показатели Российской Федерации обеспечены главным образом привлечением иностранных студентов на программы бакалавриата. По доле студентов на магистерских программах мы уступаем, помимо перечисленных, еще Франции и Германии, а по доли аспирантов опускаемся на 14-е место. Иными словами, к нам едут учиться в основном по наиболее простым программам. Если в вузах стран ОЭСР студенты из других стран, входящих в эту организацию, составляют 27%, то доля иностранных студентов из стран ОЭСР в России составляет только 1%. Среди иностранных студентов России 4% — из Китая, 2% — из Индии, а 43% — из соседних с Россией стран. Следует отметить также, что Россия — нетто-экспортер образования, то есть число приехавших к нам иностранных студентов больше, чем число молодых россиян, обучающихся за рубежом. Соотношение иностранных студентов в России и россиян, обучающихся за рубежом, составляет 5:1. Половина российских студентов за рубежом приходится на пять стран: Германию (17%), Чехию (10%), США (9%), Соединенное Королевство (7 %) и Францию (6%).

Расходы и финансирование в образовании. Уровень расходов в начальном и среднем образовании широко варьируется по странам. Средняя величина расходов на 1 школьника по странам ОЭСР составляет 10 тыс. долл. в год на 1 ученика. В России этот показатель ниже более чем в два раза (4,2 тыс. долл.). В высшем образовании размах вариации еще шире: от 7,1 тыс. долл. на студента в год в Латвии до 52 тыс. долл. в Люксембурге при среднем значении по странам ОЭСР 16,5 тыс. долл. Россия и здесь существенно отстает от других стран: в нашей стране расходы на 1 студента вуза составляют 9,5 тыс. долл. в год. Участие государства в расходах на дошкольное образование варьируется по странам очень широко: от 98% в Люксембурге до менее 50% в Великобритании и Японии. При этом наблюдается, хотя и слабо выраженная, тенденция: чем выше уровень экономического развития страны, тем большую долю расходов на дошкольное образование берет на себя государство. В России доля государства в расходах на дошкольное образование составляет 88%. Финансирование третичного образования в Российской Федерации на 64% происходит за счет государственных источников, 23% средства домохозяйств, около 12% - прочие негосударственные источники, 1% - международное финансирование. В России доля расходов на оплату труда в среднем образовании равна 83% от текущих расходов, в третичном - 67%, что близко к средним показателям развитых стран, которые составляют 80% и 66% соответственно. Но важно отметить, что по отдельным странам этот показатель варьируется значительно: от 62% (Чехия) до 95% (Колумбия) в среднем образовании и от 57% (Италия) до 100% (Колумбия) в третичном образовании.

Вопросы для повторения:

1. Как вы понимаете, что такое глобализация образования?
2. В чем минусы массовизации образования?
3. Что такое технологизация образования?
4. Как вы понимаете непрерывность и пожизненность образования?
5. Перечислите тенденции развития российского образования в сравнении с мировыми образовательными системами.
6. Перечислите основные направления стратегического развития ведущих университетов мира.

Тема 7, 8. Формы организации обучения в вузе: традиции и инновации.

7.1. Трехмерная модель систематики форм организации обучения.

7.2. Вузовская лекция. Игры. Семинары и конференции. Самостоятельная работа студентов. Проектно-творческая деятельность. Дистанционное обучение. Авторские технологии обучения.

8.1. Научно-исследовательская работа студентов.

8.2. УИР как часть профессиональной подготовки студентов. Формы организации НИР в вузе.

8.3. Защита интеллектуальной собственности.

7.1. Трехмерная модель систематики форм организации обучения.

В высшей школе сегодня успешно применяют такие формы организации обучения, как: лекции, семинары, лабораторные работы, практические занятия, специальные и функциональные (деловые) игры, теоретические (научно-практические) конференции, тестирование, консультации, подготовка рефератов, индивидуальные контрольные собеседования, самостоятельная работа обучающихся, производственная практика, курсовые работы (проекты, задачи), презентации, выпускные квалификационные работы и многие другие, обусловленные творческим подходом педагогических коллективов вузов к организации и проведению образовательного процесса.

В настоящее время с появлением в вузах современных информационных средств, мультимедийного и интерактивного оборудования стремительно развиваются такие организационные формы, как дистанционное и открытое обучение, обучение в компьютерных классах и лабораториях, обучение с помощью «Кейс технологий» и т. п.

Лекция и самостоятельная работа студентов остаются основными формами обучения. Форма организации обучения выбирается с учетом целей, особенностей содержания учебного материала, адекватных им методов и средств обучения, места и времени проведения занятий.

Цель лекции, как правило, в основном ориентирована на формирование знаний, а не практических умений.

Практические занятия имеют характерную особенность в целях, поскольку направлены на развитие практических умений и навыков студентов, а также особенности в выборе методов преподавания и учения.

Форма организации обучения - это целостная системная характеристика процесса обучения с точки зрения особенностей взаимодействия преподавателя и студента, соотношения управления и самоуправления, особенностей места и времени обучения, количества студентов, целей, средств, содержания, методов и результатов обучения.

Многомерный анализ существующих форм организации обучения позволяет представить их в виде трехмерной модели систематики форм организации обучения (рис. 3).

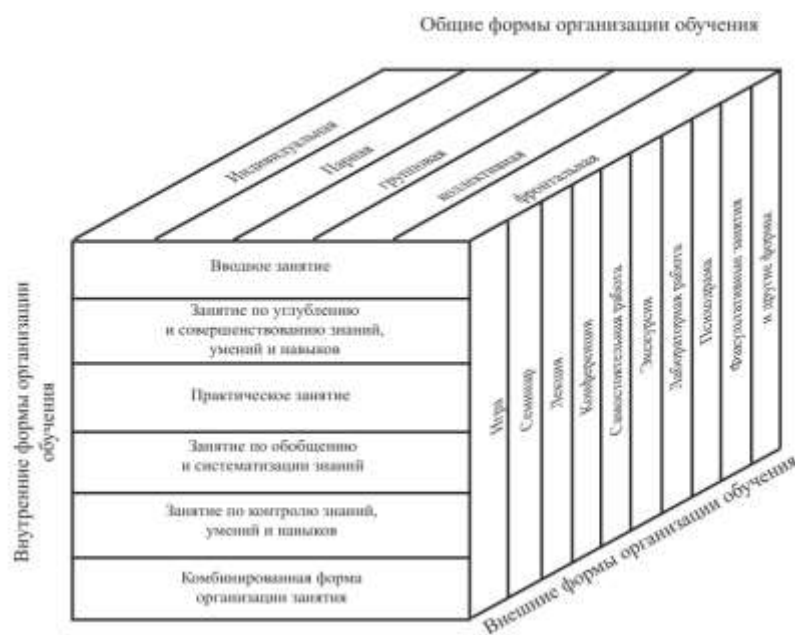


Рисунок 3 - Трехмерная модель систематики форм организации обучения

Внешняя форма организации обучения - это лекция, семинар, конференция, самостоятельная работа, игра, экскурсия, психодрама и др. Если взять цели, содержание, методы, средства, соотношение педагогического управления и самоуправления студентов, то достаточно видоизменить хотя бы один элемент, как видоизменяется внешняя форма организации обучения. Важное значение для понимания особенностей и возможностей эффективного функционирования той или иной формы организации обучения имеет ее структура, которая характеризует специфику ее внутренней организации, отражающей способы взаимодействия ее элементов. Основанием, для выделения внутренней формы организации обучения, является структурное взаимодействие элементов с точки зрения доминирующей цели обучения.

В основу общих форм организации обучения положены характеристики особенностей коммуникативного взаимодействия как между преподавателем и студентами, так и между самими студентами. При парном обучении преподаватель общается с двумя студентами, которые в свою очередь активно взаимодействуют. При групповом обучении общение преподавателя осуществляется с группой студентов из трех и более человек, которые в свою очередь имеют свои общие цели учебной деятельности и осуществляют активное взаимодействие как между собой, так и непосредственно с преподавателем. При коллективной форме обучения студенты рассматриваются как целостный коллектив, имеющий своих лидеров — руководителей из среды студентов. Общие цели, задачи и активное взаимодействие между всеми студентами обеспечивают достаточно высокий уровень их сплоченности и взаимопонимания в процессе коллективной учебной деятельности. При индивидуальной форме организации обучения преподаватель адаптирует степень сложности, трудности заданий, оказывает помощь с учетом знаний, умений и личностных качеств студентов. При фронтальной форме обучения преподаватель работает со студентами всей группы, следит, чтобы в едином темпе студенты продвигались к единой цели.

7.2. Вузовская лекция. Игры. Семинары и конференции. Самостоятельная работа студентов. Проектно-творческая деятельность. Дистанционное обучение. Авторские технологии обучения.

Выделяют следующие типы лекций: информационная, проблемная, лекция-визуализация, лекция вдвоем, лекция с заранее запланированными ошибками, лекция-пресс-конференция. Их цель - переход от простой передачи информации до активного освоения содержания обучения с одновременным запуском механизмов теоретического мышления и всей структуры психических функций. В данном процессе происходит усиление социального контекста в формировании профессионально важных качеств специалиста.

Информационная лекция – способ передачи готовых знаний обучающимся посредством монологической формы общения. Незаменим при передаче большого объема информации. Структура: вступление (формулировка темы, цели, изложение плана, характеристика рекомендуемой литературы, связь предыдущих лекций с новым материалом), основная часть (изложение содержания в соответствии с планом), заключение (подведение итога, возможности использования информации в практической деятельности, ответы на вопросы слушателей).

Проблемная лекция – тип лекции, на котором процесс познания студентов приближается к поисковой, исследовательской деятельности. При этом обеспечивается достижение трех основных целей: усвоение студентами теоретических знаний; развитие теоретического мышления; формирование познавательного интереса к содержанию учебного предмета и профессиональной мотивации. Задача преподавателя заключается не столько в передаче информации, сколько в развитии научного знания и способов их разрешения. Новое знание вводится как неизвестное для студентов. Студент не просто перерабатывает информацию, а переживает ее усвоение как субъективное открытие еще

неизвестного для себя знания. Учебный материал представляется в форме учебной проблемы, которая фиксирует некоторое противоречие (научные проблемы). Неизвестным является ответ на вопрос, который разрешает противоречие, переживаемое студентом как интеллектуальное затруднение. Для проблемного изложения отбираются узловые разделы курса, которые являются важными для будущей профессиональной деятельности. С помощью особых методических приемов (постановка проблемных и информационных вопросов, выдвижение гипотез, обращение к студентам «за помощью» и др.) преподаватель побуждает студентов к совместному размышлению, дискуссии. Лекции проблемного характера дополняются семинарскими занятиями, которые организуются как дискуссии.

Лекция-визуализация является результатом принципа наглядности, содержание которого меняется под влиянием данных психолого-педагогической науки, форм и методов активного обучения. Подготовка лекции-визуализации преподавателем состоит в перекодировании, переконструировании учебной информации по теме лекции в визуальную форму для предъявления студентам через технические средства обучения (схемы, рисунки, чертежи, презентации и т.п.). Чтение лекции-визуализации сводится к развернутому комментированию преподавателем подготовленных визуальных материалов, полностью раскрывающих тему данной лекции. Здесь важна определенная визуальная логика и ритм подачи материала. Важно учитывать цвет, графический дизайн, сочетание словесной и наглядной информации, дозировка подачи материала, мастерство и стиль общения преподавателя с аудиторией. Данный тип лекции лучше всего использовать на этапе введения студентов в новый раздел, тему, дисциплину. Основная трудность данного типа лекции состоит в выборе и подготовке средств наглядности, дидактически обоснованной режиссуре процесса ее чтения с учетом психофизиологических возможностей студентов, уровня образования и профессиональной принадлежности.

Телелекция – это один из видов лекций, опирающийся на современные аудио-, видеосредства и коммуникационные технологии обучения. Она может быть проведена без обратной связи, т.е. сначала записывается и затем тиражируется. Можно использовать и обратную связь, если лекция проводится в телестудии и применяется телефонная связь со слушателями, которые могут задать лектору, интересующий их вопрос.

Лекция вдвоем – процесс моделирования реальных профессиональных ситуаций, обсуждение теоретических вопросов с разных позиций двумя специалистами (теоретиком и практиком, сторонником и противником того или иного решения). В процессе «лекции вдвоем» создается проблемная ситуация, разворачивается система доказательств, обосновывается конечный вариант совместного решения. «Лекция вдвоем» особенно эффективна в случаях, когда целями обучения выступают формирование теоретического мышления, воспитание убеждений, необходимость развития у студентов умений оперативно анализировать профессиональные ситуации, выступать в роли экспертов, отбирать неверную или неточную информацию.

Лекции с заранее запланированными ошибками. Преподаватель закладывает в содержание лекции определенное количество ошибок содержательного, методического или поведенческого характера. Список таких ошибок преподаватель приносит на лекцию и предъявляет их студентам в конце. Лектор строит изложение таким образом, чтобы ошибки были тщательно «замаскированы». Задача студентов заключается в том, чтобы по ходу лекции отмечать в конспекте замеченные ошибки и назвать их в конце лекции. На разбор ошибок отводится 10–15 минут. В ходе этого разбора даются правильные ответы на вопросы – преподавателем студентам или совместно. Лекцию с запланированными ошибками целесообразно применять в качестве контроля знаний, диагностики трудностей усвоения материала.

Лекция-пресс-конференция близка к соответствующей форме профессиональной деятельности с некоторыми изменениями. Назвав тему лекции, преподаватель просит слушателей письменно задать ему вопрос по данной теме. Каждый слушатель должен в течение 2-3 минут сформулировать наиболее интересующий его вопрос, написать на

бумажке и передать преподавателю. Затем лектор в течение 3-5 минут сортирует вопросы по их смысловому содержанию и начинает читать лекцию. Изложение материала строится не как ответ на каждый заданный вопрос, а в виде связного раскрытия темы, в процессе которого формулируются соответствующие ответы. В завершение лекции преподаватель проводит оценку вопросов как отражение знаний и интересов слушателей. Данный тип лекции лучше всего проводить в начале изучения темы или раздела (выявить круг интересов и потребностей обучаемых, степень их готовности к работе, отношение к работе), в середине (на привлечение внимания студентов к узловым моментам содержания учебного предмета, уточнение представлений преподавателя о степени усвоения материала, систематизацию знаний студентов, коррекцию выбранной системы лекционной и семинарской работы по курсу) и в конце (подведение итогов лекционной работы, определение перспектив развития усвоенного содержания в последующих разделах). Лекцию данного рода можно провести и по окончании всего курса с целью обсуждения перспектив применения теоретических знаний на практике как средства решения задач освоения материала последующих учебных дисциплин, средства регуляции будущей профессиональной деятельности.

Педагогическая (дидактическая) игра — это такая форма организации обучения, воспитания и развития личности, которая осуществляется педагогом на основе целенаправленно организованной деятельности студентов, которая изначально мотивирована на успех, осуществляется по специально разработанному сценарию и правилам, максимально опирается на самоорганизацию обучающихся; воссоздает или моделирует опыт человеческой деятельности и общения. На рисунке 4 показана классификация игр.

По целевой ориентации среди педагогических игр могут быть выделены: дидактические (они позволяют организовать различные виды учебной деятельности; сформировать познавательные и практические умения, углубить знания); воспитывающие (ориентированные на воспитание нравственных, эстетических, коммуникативных, волевых и других качеств личности); контролирующие (они одновременно или специально могут выполнять и функции контрольно-оценочной деятельности).

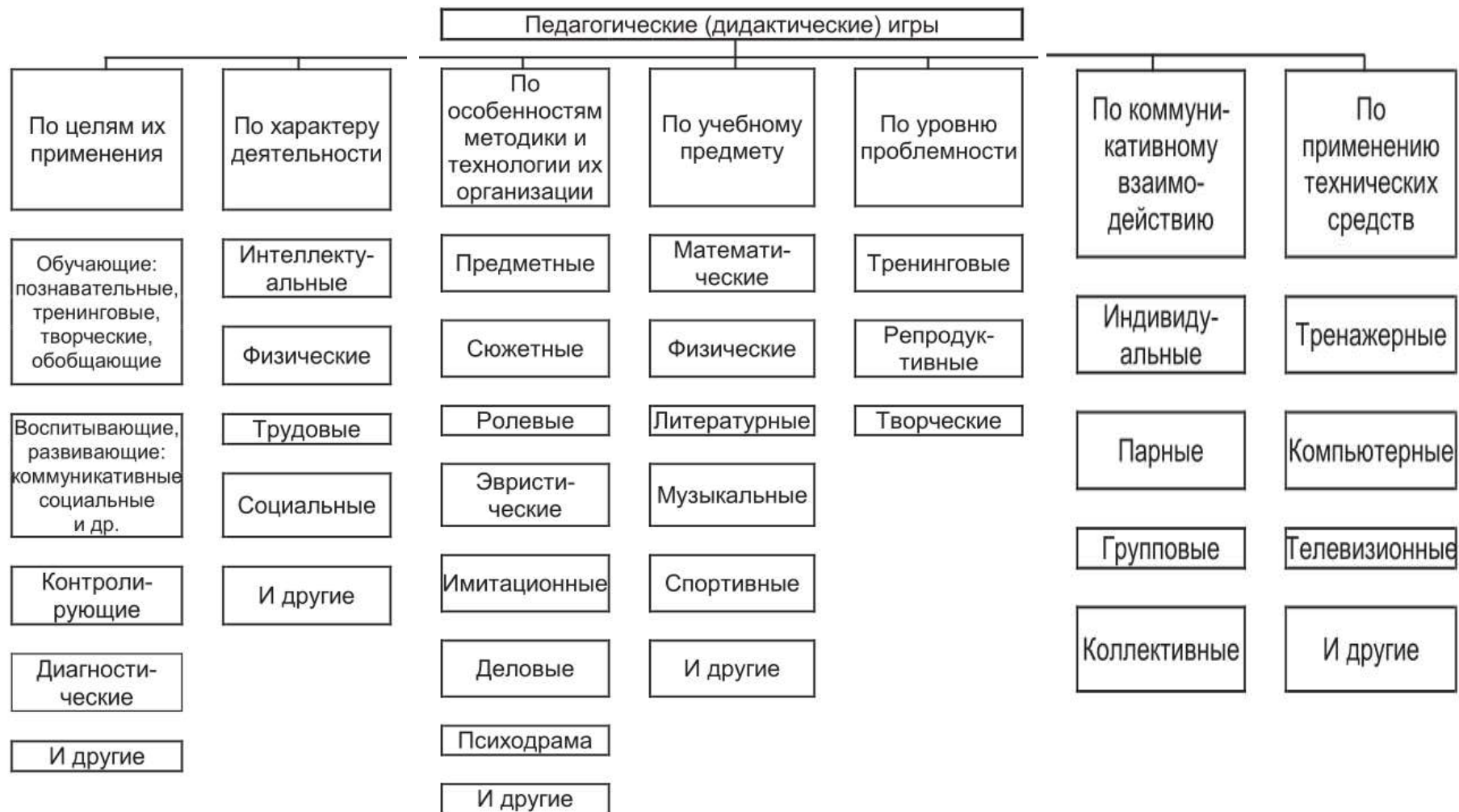


Рисунок 4 – Педагогические (дидактические) игры

Семинарская форма обучения. Главная цель – обеспечить студентам возможность практического использования теоретических знаний в условиях, моделирующих форм деятельности научных работников, предметный и социальный контексты этой деятельности. На семинарском занятии студенты должны научиться выступать в роли докладчиков и оппонентов, владеть умениями и навыками постановки и решения интеллектуальных проблем и задач, доказательства и опровержения, отстаивания своей точки зрения, демонстрации достигнутого уровня теоретической подготовки.

По большинству учебных дисциплин семинарские занятия целесообразно проводить в форме дискуссий, руководимых преподавателем. На семинаре отрабатываются важнейшие темы и разделы учебной программы. Широко распространено также обсуждение рефератов или докладов, подготовленных студентами.

Наибольшее распространение в последнее время получают спец-семинары – семинары исследовательского типа с независимой от лекционного курса тематикой, целью которых является углубленное изучение отдельных научно-практических проблем, с которыми столкнется будущий специалист. Спец-семинар, руководимый обычно крупным специалистом, приобретает характер научной школы. На семинарские занятия выносятся узловые темы курса, усвоение которых определяет качество профессиональной подготовки; вопросы, наиболее трудные для понимания и усвоения. Проработка этих тем осуществляется не в условиях индивидуальной (выступление студентов «по очереди»), а в условиях коллективной работы, обеспечивающей активное участие в ней каждого студента. Содержание семинарских занятий должно отражать принцип проблемности, быть методической основой для развертывания дискуссии, творческого применения имеющихся знаний.

Семинар-дискуссия организуется как процесс диалогического общения участников, в ходе которого происходит формирование практического опыта совместного участия в обсуждении и разрешении теоретических проблем, теоретико-практического мышления будущего специалиста. Особенностью семинарского занятия как формы коллективной творческой работы является возможность равноправного и активного участия каждого студента в обсуждении теоретических позиций, предлагаемых решений, в оценке их правильности и обоснованности. Преподаватель заранее должен ознакомить студентов с правилами ведения дискуссии, возможными ролями. Это целесообразно сделать на предшествующих семинарам проблемных лекциях с использованием метода микродискуссии. По окончании семинара-дискуссии преподаватель может сделать общие выводы, подвести итоги, оценить вклад каждого и группы в целом в решение проблемы семинара.

Семинар с использованием «сократовского» метода обучения - формой совместной творческой деятельности преподавателя и студентов. Их суть – в самостоятельном определении обучающимися основных понятий, в раскрытии сущности и закономерностей изучаемого явления и процессов путем последовательной постановки преподавателем вопросов и поиска ответов студентами. Данный метод обучения требует кропотливой самостоятельной подготовки студентов и преподавателя к занятию, в ходе которой у обучающихся формируются исследовательские умения и навыки (умение ставить проблемные вопросы, анализировать проблемные ситуации, выдвигать гипотезы – предполагаемые ответы и т.д.). Использование «сократовской» беседы в ходе семинара позволяет выявить пробелы в знаниях студентов, повышает интерес к изучаемой дисциплине, способствует активному усвоению знаний, формирует и развивает навыки самостоятельной работы и ведения беседы.

Семинар с использованием метода конкретных ситуаций. На семинарском занятии преподавателем создаются конкретные ситуации, взятые из профессиональной деятельности специалистов. От студентов требуются глубокий анализ ситуации и решение поставленной задачи. Ситуационная задача может иметь несколько вариантов решения, которые окажутся приемлемыми в данной ситуации, что требует от специалиста умения

выбрать из них наиболее оптимальные. В практике применения метода анализа конкретных ситуаций обычно используются следующие виды ситуаций: ситуация-иллюстрация (демонстрация конкретного примера из практики, в котором проявляются способы действия должностных лиц, типовые алгоритмы решения задач, эффективность использования методов и приемов руководства и т.д.); ситуация-упражнение не может быть разрешена без обращения студентов к специальным источникам информации, литературе и справочникам. Обучающий эффект обеспечивается деятельностью всех участников семинара по анализу и решению ситуационных задач; ситуация-проблема включает в себе проблемную задачу, которая стоит перед профессиональной практикой. Она может предъявляться студентам в виде текста, видеофрагмента, доклада, набора документов, отражающих состояние какого-либо объекта, процесса, события, или в форме выступления приглашенных специалистов перед студентами.

Практическое занятие – это, как правило, решение прикладных задач, образцы которых были даны на лекциях. Практические занятия преследуют следующие цели: помочь обучающимся систематизировать, закрепить и углубить теоретические знания; научить студентов приемам решения практических задач, способствовать овладению умениями и навыками в выполнении расчетов, графических и других видов заданий; научить студентов работать с книгой, документацией и схемами, пользоваться справочной литературой и прикладными программами; выработать у студентов умения учиться самостоятельно, т.е. овладеть способами и приемами самообразования и самоконтроля. Структура практического занятия может быть различной: вначале сам преподаватель демонстрирует способы решения определенного класса задач, а затем организует упражнения по решению подобных задач; студенты сразу приступают к самостоятельному решению задач, но при необходимости преподаватель дает пояснения, консультации.

Лабораторные занятия – одна из форм практической работы студентов, в которой путем проведения экспериментов осуществляются углубление и закрепление теоретических знаний, формирование умений и навыков в интересах профессиональной подготовки. Основными структурными элементами лабораторной работы являются: постановка темы и целей занятия; проверка уровня теоретических знаний, необходимых для работы; ознакомление студентов с содержанием лабораторной работы; групповое выполнение лабораторной работы; консультация преподавателя в процессе работы; обсуждение полученных результатов членами рабочей группы; письменный или устный отчет о выполненной работе; контроль и оценка результатов лабораторной работы.

В вузах применяют следующие виды лабораторных занятий: фронтальный (одновременное выполнение работы всеми студентами), по циклам (работы делятся на несколько циклов, соответствующих определенным разделам лекционного курса), индивидуальный (студенты могут одновременно работать над различными темами) и смешанный (комбинированный) тип.

Подготовка студентов к лабораторной работе осуществляется в часы самостоятельной работы с использованием учебников, конспектов лекций и методических материалов. Проведению лабораторного занятия предшествует сдача студентами коллоквиума (от лат. *colloquium* – беседа) – собеседования преподавателя со студентами по поводу предстоящей лабораторной работы, проверка глубины усвоения теоретического материала. Лабораторные занятия заканчиваются защитой результатов работы и полученных результатов.

Самостоятельная работа - планируемая, организационно и методически направляемая познавательная деятельность студентов, осуществляемая без прямой помощи преподавателя для достижения образовательных целей. Ядром самостоятельной работы является познавательная (учебная, научная, производственная) задача, предлагаемая студентам. Самостоятельная работа реализуется: непосредственно в процессе аудиторных занятий – на практических и семинарских занятиях, при выполнении лабораторных работ; в контакте с преподавателем вне рамок расписания – на консультациях по учебным

вопросам, в ходе творческого сотрудничества в рамках научно-исследовательских работ, при ликвидации задолженностей, выполнении индивидуальных заданий и т.д.; в библиотеке, дома, в общежитии, на кафедре при выполнении студентом творческих и учебных задач.

Рекомендуются следующие виды заданий: текущая работа с лекционным материалом, предусматривающая проработку конспекта лекций и учебной литературы; поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; изучение материала, вынесенного на самостоятельную проработку; решение задач; подготовка к лабораторным занятиям; подготовка к практическим и семинарским занятиям; практикум по учебной дисциплине с использованием программного обеспечения; написание реферата по заданной проблеме; выполнение расчетно-графической работы; подготовка к контрольной работе или коллоквиуму; подготовка к зачету, экзамену; выполнение курсовой работы или проекта; участие в научных студенческих конференциях и семинарах; аналитический разбор научной публикации по заранее определенной преподавателем теме; анализ статистических и фактических материалов по заданной теме, проведение расчетов, составление схем и моделей на основе статистических материалов. В качестве форм контроля самостоятельной работы могут быть: тестирование; проверка контрольных работ; доклад по самостоятельно изученной теме; веерный экспресс-опрос; отчет по результатам выполненного проекта.

Проектно-творческая деятельность. В процессе проектно-творческой деятельности студентов применяются самые разнообразные методы (эксперимент, моделирование, мозговой штурм и др.). Проектно-творческая деятельность студентов - это одна из форм самостоятельной работы студентов, направленная на решение учебных и (или) научных проблем, творческих (исследовательских) задач и заданий, выполнение (решение) которых осуществляется студентом преимущественно самостоятельно на основе педагогических методов и средств проблемного и эвристического обучения.

Проектно-творческая деятельность студентов имеет большую вариативность и по некоторым критериям ее удается классифицировать:

1. По доминирующему методу выполнения проекта: исследовательские; творческие; практико-ориентированные; теоретические; информационные и др.

2. По количеству участников: индивидуальные (личностные); парные; групповые; коллективные.

3. По содержанию деятельности: учебные; научные; практические.

4. По продолжительности выполнения: краткосрочные; среднесрочные; долгосрочные.

5. По степени вовлеченности организаций: внутривузовские; межвузовские; международные.

В проектно-творческой деятельности студентов можно выделить несколько этапов.

I этап. Самоопределение, самоактуализация, мотивация. На нем идет определение с выбором темы, проблемы, над которыми студент хотел бы работать.

II этап. Организационное и информационное обеспечение. Организация, создание реальных условий для доступа студента к необходимой справочной, учебной, научной литературе, включая и использование Интернета, позволяющие ему «войти» в проблему, познакомиться с базовой информацией, расширить и углубить свои знания по предложенной теме, проблеме.

III этап. Выдвижение предположений, формулирование гипотез, идей, разработка проекта. Выполнение творческого, исследовательского задания на этом этапе происходит индивидуально или с небольшой поддержкой преподавателя.

IV этап. Планирование. Происходит более детальное планирование выполнения проекта, конкретизация целей и задач в разработке проекта.

V этап. Сбор дополнительной информации и выполнение проекта. Собирается, систематизируется, анализируется дополнительная информация, проверяется ранее

выдвинутая гипотеза, систематизируются и анализируются данные, которые ее подтверждают или опровергают.

VI этап. Оформление результатов выполненного проекта. На этом этапе происходит не просто оформление результатов проектной деятельности студентов, но и их более целостное и глубокое осмысление.

VII этап. Защита проекта. Осуществляется публичная защита проекта, дается общая оценка результативности проектно-творческой деятельности студента.

На всех этапах проектно-творческой деятельности студента возможны консультации и помощь преподавателя, которые варьируются как по содержанию, так и по форме в зависимости от мотивации и творческого потенциала студента.

Очень важно, чтобы «защита проектов», их оценка происходили на основе четко выделенных критериев: новизны, оригинальности, обоснованности, системности и глубины проработки проблемы теоретической (или) практической значимости. Защиту проекта лучше всего проводить в форме презентации проекта с использованием схем, таблиц и других средств наглядности.

В условиях вузовской практики наибольшее внимание уделяется курсовым и дипломным проектам.

Дистанционное обучение. Это форма индивидуального обучения, в процессе которого осуществляется их погружение в интерактивную технотронную обучающую среду, обеспечивающую полный контроль студента, индивидуальное планирование учебного процесса и тестирования на расстоянии из единого центра дистанционного обучения. Дистанционное обучение - это одна из форм заочного обучения, плюс общение с преподавателем через Интернет, в ходе которого студент получает учебные материалы и задания на свой компьютер, выполняет тесты и контрольные работы и отправляет их преподавателю. В процессе дистанционного обучения изучается теоретическая часть, выполняются практические задания и решаются контрольные работы, которые затем отсылаются преподавателю по электронной почте. В процессе обучения возможно общаться и задавать вопросы преподавателю по e-mail. При дистанционном обучении Вы имеете возможность сами выбирать последовательность изучения предметов и темп работы. Можете решать, сколько времени потратить на изучение того или иного курса. «Дистанционный» студент получает комплект материалов сразу при зачислении на занятия. В такой комплект входят не только учебники, но и тексты лекций, практикумы, задания для самостоятельной работы на разных носителях — традиционных бумажных, CD, аудио- и видеоносителях.

Вместе с новыми методами и технологиями обучения дистанционное обучение привносит в теоретическую педагогику и образовательную практику новые понятия и термины, в первую очередь к ним относятся:

- виртуальный класс (группа);
- поддержка обучения (поддержка студентов);
- учебные телекоммуникационные проекты;
- обратная связь;
- диалоговая технология;
- компьютерная связь;
- телеконференция;
- координатор, модератор, фасилитатор телекоммуникационного проекта (телеконференции).

Под виртуальным классом (группой) понимается общность студентов, взаимодействие между которыми при совместном выполнении ими учебных заданий происходит по компьютерным сетям.

Под поддержкой обучения (или поддержкой обучаемого) понимают любые материалы, информацию, поступающую от преподавателя к студенту, находящемуся в другой географической точке.

Учебный телекоммуникационный проект - совместная (коллективная) деятельность студентов, направленная на достижение некоторой модельной цели, которая носит не учебный характер и моделирует цель какой-либо научной или производственной деятельности. Важными отличительными чертами учебного телекоммуникационного проекта являются:

- его временная определенность и ограниченность (от двух недель до трех месяцев);
- использование компьютерных телекоммуникационных сетей и программных средств для обмена информацией между всеми участниками проекта, которые часто образуют виртуальную или квазивиртуальную группу;
- необходимость четкой организации деятельности студентов, которая устанавливается координатором проекта.

Обратная связь в дистанционном обучении — поток информации от педагога к дистанционному студенту на стадии оценивания педагогом деятельности учащегося, его продвижения и успехов и несущая реакцию педагога на успехи студентов, оценку его деятельности (одобрение или неодобрение).

Диалоговая технология — конфигурация программного обеспечения, оборудования, а также межличностного взаимодействия и деятельности, обеспечивающая свободное общение.

Телеконференция — способ обмена текстовыми сообщениями с некоторыми сообществами заинтересованных в этом людей.

Компьютерная связь — совокупность способов использования компьютеров и телекоммуникационных сетей в качестве инструментов для организации связи. Компьютерная связь включает в себя: электронную почту, которая позволяет направлять сообщения в почтовые ящики пользователей сети; телеконференции, которые позволяют направлять сообщения всем участникам одновременно; доступ к удаленным информационным источникам, например, библиотечным ресурсам, базам данных, серверам.

Авторские технологии обучения. При разработке авторских технологий следует учитывать, что главное в арсенале преподавателя - это он сам. Его голос, жесты, доброе, заботливое отношение к студентам, стремление соприкоснуться с духовным миром каждого студента и открыть свой, всякий раз неожиданно загадочный, новый мир.

Большинство преподавателей при разработке собственных, более гибких авторских технологий опираются не на какую-то одну, а несколько дидактических (педагогических) концепций, выстраивая их с учетом специфики и приоритетности решаемых задач обучения, воспитания и развития личности. Приведем девять этапов или «слагаемые любой педагогической технологии», выделенные в результате исследования В.П. Беспалько:

- 1 — анализ будущей деятельности студента;
- 2 — определение содержания обучения на каждой ступени обучения;
- 3 — проверка степени нагрузки студента и расчет необходимого времени при заданном способе построения учебного процесса;
- 4 — выбор организационных форм обучения и воспитания, наиболее благоприятных для реализации намеченного дидактического процесса;
- 5 — подготовка материалов (текстов, ситуаций) для осуществления мотивационного компонента дидактического процесса;
- 6 — разработка системы учебных упражнений, нацеленных на усвоение предметов с заданными показателями качества;
- 7 — разработка материалов (тестов) для объективного контроля за качеством усвоения студентами знаний и действий соответственно целям обучения и критериям оценки степени усвоения;
- 8 — разработка структуры и содержания учебных занятий, нацеленных на эффективное решение образовательных и воспитательных задач;

9 — апробация проекта на практике и проверка завершенности учебно-воспитательного процесса (достижения цепей с показателями усвоения $K > 0,7$), коррекция проекта».

8.1. Научно-исследовательская работа студентов

Формы участия студентов в научно-исследовательской работе могут быть сведены к двум направлениям: учебно-исследовательской работе студентов, проводимой в учебном процессе и внесенной в учебные планы; учебно-исследовательской работе студентов, не связанной или косвенно связанной с учебным процессом. Эти направления и составляют систему научно-исследовательской работы студентов.

Научно-исследовательская работа студентов – одна из важнейших форм учебного процесса. Научные лаборатории и кружки, студенческие научные общества и конференции – все это позволяет студенту начать полноценную научную работу, найти единомышленников, с которыми можно посоветоваться и поделиться результатами своих исследований. Написание рефератов, курсовых, дипломных работ невозможно без проведения каких-то, пусть самых простых, исследований. Более глубокая научно-исследовательская работа, заниматься которой студента не обязывает учебный план, охватывает далеко не всех. Для тех же, кто проявляет интерес к творчеству и поиску, в вузах специально организуется дополнительная научно-исследовательская и творческая работа.

Внеучебная, вне сетки расписания, работа включает в себя большое многообразие видов деятельности студентов: участие в научных кружках; подготовку рефератов, докладов, сообщений; выступление с ними на факультетских и других научных конференциях; выполнение исследований по хозяйственным договорам или госбюджетной тематике в составе научных коллективов преподавателей; участие в студенческих конкурсах научно-исследовательских работ, олимпиадах, выставках студенческого творчества; подготовку статей в научные журналы, сборники научных работ, периодическую печать и др.

Часто старт студента в науку начинается именно с участия в работе кружка или иного студенческого научного объединения. Целями любого студенческого научного объединения являются развитие у молодежи творческого мышления через изучение методологических основ научной работы, освоение научной методики, способов и приемов изложения материала и обработки результатов научного исследования.

Демонстрация получаемых научных результатов обычно выходит за рамки собственно кружковой работы в виде выступлений с докладами и сообщениями по итогам научных исследований, участие в научных дискуссиях и т.д.

Существуют следующие виды студенческих научных объединений: студенческий предметный или тематический научный кружок, научная проблемная или творческая группа, научно-исследовательская лаборатория и др.

Научный кружок (тематический или предметный) - объединение студентов, аспирантов и других заинтересованных лиц, основанное на общности интересов, взглядов, идей с целью совместного научного творчества. Как правило, основными видами научных работ в таких объединениях являются: составление аннотаций по научной литературе и написание рефератов, овладение навыками проведения эксперимента и обработки результатов, проектирование и изготовление наглядных пособий, подготовка сообщений и выступлений на семинарах и конференциях и т.д. Предметный научный кружок чаще всего организуется при работе со студентами младших курсов по изучаемым ими учебным предметам и является первой ступенькой в «царство науки», и поэтому задачи перед его участниками ставятся несложные. Чаще всего это подготовка докладов и рефератов, которые потом заслушиваются на заседаниях кружка или на научной конференции.

Проблемные кружки. Все выше сказанное можно отнести и к проблемным, но следует учесть и некоторые отличия. Проблемный кружок может объединять собой студентов разных факультетов и курсов, а также, если при вузе имеются таковые, колледжей и лицеев.

Во главу угла может быть поставлена проблема, которой занимается научный руководитель кружка, или любая другая по его выбору. Проблемные кружки предполагают встречи с людьми, которые сталкиваются с проблемами, выбранными для рассмотрения, на работе и в быту, проведение различных викторин и КВН. Проблемный кружок может сочетать в себе элементы научного кружка, лаборатории и т.д.

Проблемные студенческие лаборатории. В них принимают участие студенты второго курса и старше. В рамках проблемных студенческих лабораторий осуществляются различные виды моделирования, изучение и анализ реальных документов, программ, деловых игр, а также практическая помощь предприятиям. Работа в такой лаборатории предполагает не столько изучение и анализ литературы, сколько постановку эксперимента, создание чего-то нового, способности студента к коллективной работе.

Научно-исследовательская лаборатория - студенческая группа, проводящая учебные исследования и научные эксперименты. В ней осуществляются различного вида пробы, опыты, моделирование, создание чего-то нового, изучение и анализ документов, проводятся деловые игры и т.д. Работа в лаборатории предполагает наличие определенного запаса знаний и навыков. Еще одной отличительной чертой лаборатории является преобладание коллективных форм работы над индивидуальными. Если в кружке студент отвечает, как правило, только за себя, то в лаборатории его тема исследования включена в общую тему и от правильности решения частных задач зависят общие результаты работы.

В лаборатории темы более конкретные, как правило, имеющие выход на практику. Поэтому студенты имеют больше возможностей быть приглашенными на работу в организации, выступающие заказчиками исследований. Студенческие лаборатории, работающие по научной хозяйственной теме кафедры, получают от нее не только моральную поддержку, но и материальное вознаграждение.

Участие в научно-практических конференциях. Включают в себя не столько теоретические научные доклады, сколько обсуждение путей решения практических задач. Проходят на территории завода, управляющего органа, с которым вуз поддерживает отношения. Они способствуют установлению тесных дружеских связей между вузом и предприятиями, помогают студентам учиться применять изученную теорию на практике.

Участие в научно-производственных структурах, временных творческих коллективах преподавателей кафедры, в бюджетных и внебюджетных научных исследованиях, в том числе включенных в планы НИР университета. Деятельность таких коллективов осуществляется под руководством ведущего по данному научному направлению преподавателя, чаще всего профессора, доктора наук. Каждый член такого научного объединения работает на постоянной, возможно даже частично платной, основе имеет научного консультанта и строго индивидуальное задание. Отвечая за себя, член такого коллектива понимает, что от его вклада зависит результат общего большого научного проекта.

Тьюторство. Оно предполагает, что студент в течение всего периода обучения в вузе последовательно разрабатывает определенную тему под руководством одного преподавателя. При такой организации обучения студенты наиболее полно осваивают методы и специфику научной деятельности, приобретают навыки работы в научных коллективах и организациях, а их научные руководители отбирают для себя потенциальных аспирантов.

Любая научно-исследовательская работа, независимо от того, в какой организационной форме она осуществляется, проходит через строго определенные этапы, последовательность и взаимосвязь которых отражает «технология научно-исследовательской работы»: постановка проблемы, определение объекта исследования; выбор и обоснование темы исследования; определение целей и задач исследования; выбор методов исследования; сбор и обработка информации об объекте исследования; построение модели функционирования объекта познания и его многоаспектное изучение с

применением различных методов исследования; оформление результатов исследования и их защита.

Выбор объекта исследования определяется объективными факторами, такими как его значимость, наличие нерешенной проблемы, ее актуальность, новизна и перспективность, и субъективными факторами – жизненным опытом, склонностями, интересами исследователя, научным руководителем и др.

От доказательства актуальности темы логично перейти к формулировке цели исследования, а также указать на конкретные задачи, которые предстоит решать в соответствии с этой целью (изучить, проанализировать, установить, выяснить, построить, обосновать, доказать и т.д.).

Первый этап научного исследования – выбор методов, которые служат инструментом в добывании фактического материала, что является необходимым условием достижения цели исследования. Второй этап – это проведение самого исследования с помощью выбранных методов и описание этого процесса, в котором освещаются методика, техника и результаты исследования. Третий этап научного исследования – обсуждение его результатов, которое ведется на заседаниях соответствующих кафедр, где дается предварительная оценка теоретической или практической значимости проведенной работы. Результаты исследования оформляются в виде научно-исследовательского отчета, а также докладываются студентом на научном семинаре или научно-практической конференции.

8.2. УИР как часть профессиональной подготовки студентов. Формы организации НИР в вузе

Источниками получения информации для студентов, кроме лекций преподавателя, являются учебники, учебные и методические пособия, научная литература (монографии, журналы), средства массовой информации и т.д.

Одним из основных видов самостоятельной учебно-воспитательной деятельности студентов является работа с книгой. Эффективность этой работы зависит от уровня сформированности таких умений, как: работа в библиотеке с каталогами и подбор литературы по определенной теме (проблеме); чтение и анализ текста; выделение узловых элементов информации; составление плана и конспекта по прочитанному тексту; цитирование; подготовка доклада к семинарскому занятию или к студенческой конференции; составление рецензии; оформление реферата.

Знакомство с книгой начинается с чтения аннотации (краткая характеристика печатного издания), включающая: сведения о целях, структуре и содержании работы, об авторе и достоинствах печатного издания, о тех, кому она предназначена. Помещается на обороте титульного листа книги.

Составление плана информационного текста. План текста – это перечень узловых вопросов, отражающих структуру его содержания (например, перечень вопросов, приведенных к каждой главе).

Конспектирование – это сжатое и последовательное письменное изложение содержания прочитанного. Нужно для того, чтобы: переработав любую информацию, передать ее в сокращенном виде; выделить в письменном тексте самое необходимое и нужное для решения учебной или исследовательской задачи; создать модель проблемы; упростить запоминание текста, облегчить овладение специальными терминами; накопить информацию для написания более сложной работы (доклада, реферата, курсовой, дипломной работы).

Цитирование. Цитата – это точная выдержка (часть текста) из какого-либо литературного источника. Она служит подтверждением выдвинутых автором положений и приводится в кавычках, точно по тексту оригинала (первоисточника). Пропуск слов, предложений, абзацев при цитировании обозначается многоточием; не допускается объединение в одной цитате нескольких отрывков, взятых из разных мест.

Библиографическая ссылка приводится с полной информацией о первоисточнике с указанием его номера из библиографического списка использованной литературы и страницы.

Рецензирование. Рецензия - статья, содержащая в себе критический обзор какого-либо научного или художественного произведения, либо отзыв на научную работу (диссертацию, монографию, учебник и т.д.). Она раскрывает содержание рецензируемой работы, дает критическую оценку работе.

Написание реферата (реферирование). Реферат – это сжатое изложение основной информации первоисточников на основе ее смысловой переработки. Он позволяет представить содержание печатных изданий в обобщенном виде. Реферат как разновидность учебной исследовательской работы студента должен включать: обоснование актуальности темы, цель, задачи исследования, анализ литературы по проблеме со ссылкой на первоисточники, основное содержание, заключение, список литературы.

Курсовые работы призваны приобщать студентов к исследовательской работе над проблемами, которые разрабатываются преподавателями; учитывать разнообразие интересов студентов в области выбранной специальности, а также тематику исследовательской работы на факультете. После того, как тема курсовой работы выбрана и согласована с научным руководителем, составляется календарный план, в котором определяются сроки выполнения курсовой работы.

Дипломные работы. Дипломная работа – одна из основных форм выпускных квалификационных работ, предусмотренных в качестве аттестационных испытаний. Тематика дипломных работ определяется вузом. Студенту предоставляется право выбора темы, он может также предложить свою тему с обоснованием целесообразности ее разработки. Дипломная работа проверяется научным руководителем, а затем после исправления ошибок направляется на рецензирование (кроме бакалавриата). В рецензии отмечаются: актуальности темы; полнота и обстоятельность изложения поставленной проблемы; эффективность использования выбранных методов для решения проблемы; достижение поставленной цели; практическая значимость результатов.

8.3. Защита интеллектуальной собственности.

Многообразие видов интеллектуальной деятельности обуславливает многообразие форм её результатов – объектов интеллектуальной собственности (ОИС). Выделяют две сферы возникновения ОИС: 1) научно-техническую и производственную; 2) гуманитарную (рис. 5). Разделяют ОИС на 3 группы в зависимости от институтов права, регулирующих правоотношения в связи с их созданием и использованием: 1) объекты авторского права и смежных прав; 2) объекты промышленной собственности (объекты патентного права); 3) производственные секреты (ноу-хау). Отличительная особенность авторского права в том, что охрана прав распространяется в отношении формы произведения, а не его содержания. В произведениях живописи, литературы охрана предоставляется не сюжету, а форме, в которой он выражен. Защищается не сюжет книги, а словесная (литературная) форма его выражения. Соответственно, один и тот же сюжет может быть использован и писателем, и киносценаристом. В отношении программ для ЭВМ охрана распространяется на совокупность команд, но не на решаемые программой задачи и алгоритмические процедуры, которые она реализует. Наоборот, для объектов промышленной собственности (изобретений, полезных моделей) и производственных секретов (ноу-хау) приобретаемые права распространяются в отношении их содержания и, как правило, не зависят от конкретной формы реализации.

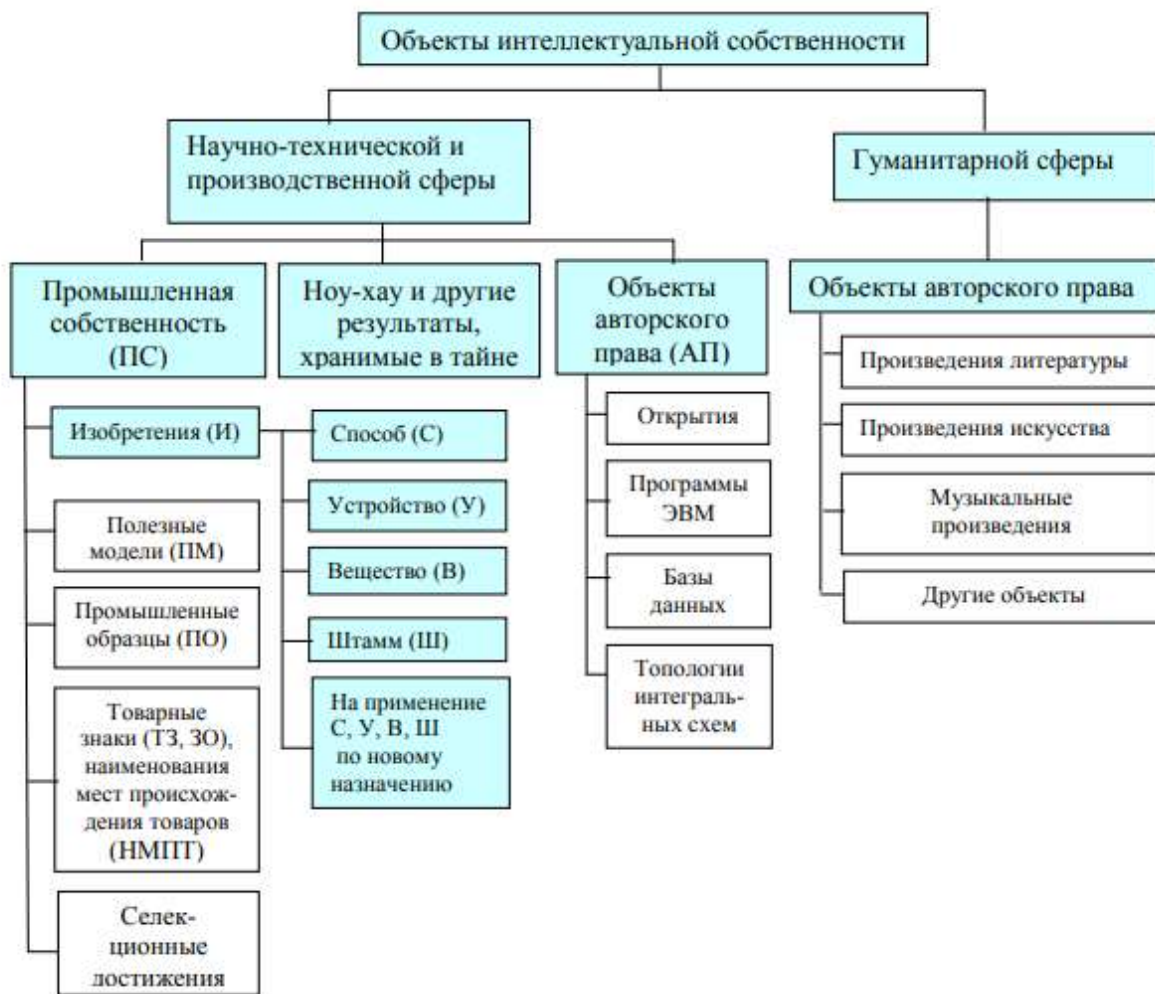


Рисунок 5 – Классификация объектов интеллектуальной собственности

Объекты промышленной собственности (ПС) составляют наиболее сильную подсистему ОИС (рис. 5). Охрана объектов ПС возникает только после признания их патентным ведомством патентоспособными и выдачи охранного документа – патента или свидетельства. Выдаче предшествует специальная экспертиза. Весьма специфично осуществляется охрана прав на ноу-хау: государство гарантирует обладателю ноу-хау защиту от незаконного использования этих сведений третьими лицами, но при условии, что: 1) эта информация имеет действительную или потенциальную коммерческую ценность в силу неизвестности её третьим лицам; 2) к этой информации нет свободного доступа на законном основании; 3) обладатель информации принимает надлежащие меры к охране ее конфиденциальности. Таким образом, пока выполняются эти условия, существует ноу-хау и существует охрана ИС в отношении этого ноу-хау. Следовательно, формой охраны ноу-хау является сохранение его в тайне.

Объекты авторского права (АП) - произведения (совокупность идей, мыслей и образов, получивших в результате творческой деятельности автора свое выражение в доступной для восприятия человеческими чувствами конкретной форме, допускающей возможность воспроизведения) науки, литературы и искусства независимо от назначения, а также от способа его выражения. Объектом АП следует считать не просто работу автора и не идеи, выраженные им, а произведение как комплекс идей и образов, получивших свое выражение в готовом труде, как индивидуальное и неповторимое творческое отражение объективной действительности. В ГК РФ определены следующие виды произведений: литературные произведения; драматические и музыкально-драматические произведения, сценарные произведения; хореографические произведения и пантомимы; музыкальные

произведения с текстом или без текста; аудиовизуальные произведения; произведения живописи, скульптуры, графики, дизайна, графические рассказы, комиксы и другие произведения изобразительного искусства; произведения декоративно-прикладного и сценографического искусства; произведения архитектуры, градостроительства и садово-паркового искусства, в т.ч. в виде проектов, чертежей, изображений и макетов; фотографические произведения и произведения, полученные способами, аналогичными фотографии; географические, геологические и другие карты, планы, эскизы и пластические произведения, относящиеся к географии, топографии и к другим наукам; другие произведения.

К объектам АП также относятся программы для ЭВМ, которые охраняются как литературные произведения. Каждый из объектов АП может быть классифицирован по многочисленным подвидам – по их внешним формам, жанрам и сферам применения. Так, литературные произведения могут быть художественного (1), научного (2), учебного (3) характера. Надо помнить, что АП не распространяется на идеи, концепции, принципы, методы, процессы, системы, способы, решения технических, организационных или иных задач, открытия, факты, языки программирования. Спорным в теории и практике является вопрос о включении в число объектов АП формул. К числу объектов, не охраняемых АП, относятся, прежде всего, те из них, которые не обладают хотя бы одним из признаков произведения науки, литературы и искусства. Если в ходе проделанной работы достигнут чисто технический результат, он также АП не охраняется. К ним относятся телефонные справочники, расписания движения, адресные книги и т.п. при условии, что составителем не применена оригинальная схема изложения справочных данных. Наряду с подобными объектами существуют произведения, обладающие всеми необходимыми для охраны признаками, но не охраняемые АП в силу прямого указания закона. К их числу относятся следующие четыре категории произведений: 1) произведения, срок охраны которых истек; 2) официальные документы, их официальные переводы, а также государственные символы и знаки; 3) произведения народного творчества; 4) сообщения о событиях и фактах, имеющие информационный характер. Значение имеет деление произведений на обнародованные и необнародованные, опубликованные и неопубликованные. АП охраняются и те, и другие. Однако если необнародованные произведения неприкосновенны и ни при каких условиях не могут быть использованы без согласия их авторов, то обнародованные произведения в исключительных, предусмотренных законом случаях, могут быть использованы заинтересованными лицами без согласия авторов и даже вопреки их возражениям. Аналогичные различия есть между опубликованными и неопубликованными произведениями. Под опубликованием в законе понимается выпуск в обращение экземпляров произведения, т.е. изготовление и выпуск в обращение копий произведения, изготовленных в любой материальной форме. Оно должно быть совершено с согласия автора. Произведения подразделяются на оригинальные, производные, составные. Практическое значение этой классификации – в том, что для создания и использования производных произведений надо получить разрешение обладателей авторских прав на те произведения, которые станут основой для производных. Оригинальным является такое произведение, все основные охраняемые элементы которого созданы самим автором. В производном (зависимом) произведении заимствованы охраняемые элементы чужого произведения.

Вторым условием возникновения авторских прав на такое произведение является соблюдение его создателем прав автора произведения, подвергнувшегося переводу, переработке, аранжировке или другой переработке. Помимо производных произведений к объектам АП также относятся сборники (энциклопедии, антологии, базы данных) и другие составные произведения, представляющие собой результат творческого труда по подбору или расположению материалов. Значение для определения авторских правомочий и режима использования произведения оказывает признание его служебным (произведения,

созданные в порядке выполнения служебных обязанностей или служебного задания работодателя).

Одним из объектов АП является программа для ЭВМ. Под ней понимается объективная форма представления совокупности данных и команд, предназначенных для функционирования ЭВМ и других компьютерных устройств с целью получения определенного результата. К числу программ для ЭВМ относят также подготовительные материалы, полученные в ходе ее разработки, и порождаемые ею аудиовизуальные отображения. Авторские права на все виды программ для ЭВМ (в т.ч. на операционные системы и программные комплексы), которые могут быть выражены на любом языке и в любой форме, включая исходный текст и объектный код, охраняются так же, как авторские права на произведения литературы. Охрана не распространяется на идеи и принципы, лежащие в основе программ, баз данных и топологий, в т.ч. на языки программирования.

Однотипны и личные права (право авторства, право на имя и право на неприкосновенность), имущественные права на программы, базы и топологии принадлежат как их создателям (авторам), так и их наследникам, а также другим физическим или юридическим лицам, получившим исключительные права в силу закона или договора. АП на программы для ЭВМ возникает с момента их создания и воплощения в объективной форме. Вместе с тем обладатель всех имущественных прав на программу вправе по своему желанию непосредственно либо через своего представителя зарегистрировать этот объект в Роспатенте. Использование программ для ЭВМ третьими лицами (пользователями) осуществляется, как правило, по договору с правообладателями. Допускается свободная перепродажа или передача иным способом права собственности либо иных прав на экземпляр программы или базы данных после первой продажи или другой передачи права на этот экземпляр. Лицу, правомерно владеющему экземпляром программы, разрешено свободно манипулировать ее данными, в т.ч. адаптировать их – вносить изменения, необходимые для функционирования программы на технических устройствах пользователя, а также осуществлять ее запись и хранение в памяти ЭВМ. Законный обладатель вправе изготавливать копию программы для архивных целей и для замены правомерно приобретенного и впоследствии утерянного, испорченного или ставшего непригодным к использованию оригинала. При определенных условиях обладатель экземпляра программы для ЭВМ может также ее декомпилировать – воспроизвести и преобразовать объектный код в исходный текст.

Интеллектуальная собственность научно-технической и производственной сфер – это изобретения (И), полезные модели (ПМ) и промышленные образцы (ПО), охрана которых осуществляется в рамках ГК РФ и подзаконных актов. Правовая охрана предоставляется на основании процедуры государственной регистрации, в ходе которой соответствующие РИД проверяются на охранную способность (патентоспособность). На РИД, признанный патентоспособным Роспатентом выдается официальный документ – патент, удостоверяющий исключительное право, авторство и приоритет И, ПМ либо ПО. Срок его действия исчисляется со дня подачи первоначальной заявки в Роспатент и составляет: для И – 20 лет, для ПМ – 10 лет, для ПО – 15 лет. Срок действия патента на ПМ может быть продлен по заявлению патентообладателя на срок, указанный в заявлении, но не более 3 лет, на ПО – на срок, указанный в заявлении, но не более 10 лет. По истечении срока действия исключительного права И, ПМ или ПО переходят в общественное достояние и могут использоваться любым лицом без чье-либо согласия или разрешения и без выплаты вознаграждения за использование. В соответствии со ст. 1349 ГК РФ объектами патентных прав не могут быть: а) способы клонирования человека; б) способы модификации генетической целостности клеток зародышевой линии человека; в) использование человеческих эмбрионов в промышленных и коммерческих целях; г) иные решения, противоречащие общественным интересам, принципам гуманности и морали.

Ключевым понятием патентного права является патентоспособность – совокупность свойств технического решения, без наличия которых оно не может быть признано

изобретением на базе действующего законодательства (в нашем случае – России). Статья 1350 ГК РФ, определяет три условия патентоспособности изобретения – новизну, изобретательский уровень и промышленную применимость. В ГК РФ есть указание на объекты, которые не могут выступать в качестве изобретения. Так, не являются изобретениями: 1) открытия; 2) научные теории и математические методы; 3) решения, касающиеся только внешнего вида изделий и направленные на удовлетворение эстетических потребностей; 4) правила и методы игр, интеллектуальной или хозяйственной деятельности; 5) программы для ЭВМ; 6) решения, заключающиеся только в представлении информации. Не предоставляется правовая охрана в качестве изобретения: 1) сортам растений, породам животных и биологическим способам их получения, за исключением микробиологических способов и продуктов, полученных такими способами; 2) топологиям интегральных микросхем. Срок действия исключительного права на изобретение (согласно ст. 1363 ГК РФ) составляет 20 лет.

Лицензионный договор – это соглашение, по которому обладатель исключительного права на объект интеллектуальной собственности (лицензиар) предоставляет или обязуется предоставить другой стороне (лицензиату) право использования этого объекта в предусмотренных договором пределах. Сторонами лицензионного договора, т.е. лицензиаром — обладателем исключительного права и лицензиатом — временным пользователем объекта интеллектуальной собственности, могут быть любые субъекты гражданских прав при соблюдении правил о право- и дееспособности.

Предметом лицензионного договора является право использования определенного объекта интеллектуальной собственности определенными в договоре способами, и лицензиат вправе использовать объект только в пределах тех прав и теми способами, которые предусмотрены договором. Договор должен быть заключен в письменной форме, если Кодексом не предусмотрено иное.

Вопросы для повторения:

1. Определите цели и содержание научно-исследовательских работ студентов.
2. Каковы функции участия студентов в исследовательских работах?
3. Раскройте содержание этапов исследовательских работ.
4. Каковы особенности различных форм их организации?
5. Какие задачи решаются советом вуза по исследовательским работам?
6. Назовите основные группы ОИС и раскройте состав этих групп.
7. В чем состоит отличие особенностей объектов авторского права от объектов промышленной собственности?
8. Что стоит за словом «патент» и что не может быть объектом патентных прав?

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

Информационная безопасность в профессиональной деятельности

Профиль подготовки

Искусственный интеллект и бизнес-аналитика

Квалификация выпускника

Магистр

Формы обучения

очная

г. Ульяновск, 2021

Тема 1.1. Информационная безопасность

1. Информация. Определение, особенности, виды информации.
2. Компрометация информации. Базовые критерии информационной безопасности. Конфиденциальность, целостность, доступность.
3. Информационная безопасность. Определение и структура ИБ. Подходы к обеспечению и управлению ИБ. Классификация способов защиты информации

Информация – это философская категория, в зависимости от контекста обозначающая:

- смысл/содержание формы;
- данные, содержащиеся в хранилищах;
- сигналы, передающиеся по линиям связи;
- сведения (сообщения, данные) независимо от формы их представления (ФЗ № 149-ФЗ, ст. 2);
- множество строк из определённого алфавита (Теория формальных грамматик).

Какую бы форму ни принимала информация, она обладает следующими особенностями:

- нематериальна, следовательно: неотделима от носителя, неисчерпаема, не локализована в пространстве;
- упорядочивает хаос (энтропия).

Различают следующие виды информации:

- вербальная («мягкая»): носитель – канал передачи, форма – язык;
- невербальная («твёрдая»): носитель – хранилище, форма – код;
- смешанная – сочетание различных видов;

- комплексная – наложение различных видов.

С точки зрения защиты информации ее можно классифицировать по типам:

12. открытая – незащищённая;

- конфиденциальная – предприняты меры по защите;
- публичная – предприняты меры по подготовке к публикации.

К определённому типу информации относит обладатель информации – лицо, самостоятельно создавшее информацию либо получившее на основании закона или договора право разрешать или ограничивать доступ к информации, определяемой по каким-либо признакам (ФЗ № 149-ФЗ, ст. 2).

В зависимости от обладателя информации и типа защищенности выделяют следующие категории информации (табл. 1).

Таблица 1.

Категории информации

Индивид	Организация	Государство
Конфиденциальная	Секретная	Тайная
Открытая	Рабочая	Свободная
Публичная	Декларированная	Официальная

Всякая информация обладает определённой ценностью. Ценность (полезность) информации можно оценить со стороны:

- полезности (результативности) – что можно сделать с её помощью;
- правильности (полноты, точности) – соответствует ли она действительности;
- своевременности (актуальности) – можно ли использовать её.

Компрометация информации – негативные последствия от угроз для информации. Выделяют основные виды компрометации информации, связанные со снижением её ценности с определённой стороны:

- утечка – информация неисчерпаема, но утрачивается её результативность (к);
- искажение – информация не уничтожима, но утрачивается полнота-точность (ц);
- потеря – информация не локализована, но утрачивается актуальность (д);

Информационная безопасность должна обеспечивать всестороннюю защиту ценности имеющейся информации.

Критерии информационной безопасности – основные (базовые) направления защиты ценности информации. Критерии могут быть использованы для определения приоритетов, выбора средств защиты, и оценки защищённости (табл. 2).

Таблица 2.

Модель CIA (Confidentiality-Integrity-Availability)

Критерий	Определение	Операция
Конфиденциальность Confidentiality	Легальное использование	Чтение, r (передача)
Целостность Integrity	Отсутствие вмешательства	Запись, w (модификация)
Доступность Availability	Беспрепятственный доступ	Активация x (преобразование)

Дополнительно выделяют (производные) критерии:

18. секретность (ca) – защита данных и канала передачи (защита от перехвата);

- неотказуемость (ci) – возможность установить авторство (подтверждение подлинности);
- сохранность (ia) – соответствие и готовность к использованию (сохранность улик);
- идентичность (cia) – соответствие целям (от происхождения, до применения).

На основании критериев защиты формируются конкретные требования к защите:

- требования обеспечения конфиденциальности:

1. неразглашение - требования к защите от утечек,
2. категорирование - требования к разделению информации по степени защиты (виды информации),
3. скрывание - требования к мерам по засекречиванию наличия информации,
4. ответственность - требования к соответствию законодательству (назначение ответственных, поиск виновных),

- требования обеспечения целостности:

1. сохранность - требования к хранению (что, место, время),
2. неизменность - требования к отсутствию вмешательства,
3. корректность - требования к проверкам (кто, когда, как),
4. неотказуемость - требования к контролю изменений,

- требования обеспечения доступности:

1. разграничение - разделение информации по способу обращения (роли-операции),
2. производительность - количество одновременно/за период обрабатываемых запросов,
3. надежность – степень сохранения возможности выполнения требуемых функций.

Надежность можно оценить количественно по формулам:

$$\frac{(Д(\text{время обещанной доступности}) - П(\text{время простоя}))}{И(\text{интервал: 24ч, 7д, 31д})} \times 100\%$$

$$\frac{MTTF(\text{средняя наработка до отказа})}{(MTTF + MTTR(\text{среднее время до восстановления}))} \times 100\%$$

Или качественно:

- Низкая.
- Средняя.
- Высокая доступность – наиболее распространённый уровень, ожидаемый пользователями, при котором система или приложение доступны в обозначенные требованиями дни и часы без незапланированных простоев, а о запланированных остановках в работе объявлено заранее.
- Непрерывный режим работы (continuous operations) -- система доступна 24 часа в сутки 7 дней в неделю без запланированных простоев.
- Постоянная доступность (continuous availability) -- сочетание высокой доступности с непрерывным режимом работы, система доступна 24 часа в сутки 7 дней в неделю без запланированных или незапланированных простоев.

Информационная безопасность (ИБ) – процесс защиты конфиденциальности, целостности и доступности информации.

Для защиты информации по каждому направлению существуют следующие методы:

- шифрование – конфиденциальность;
- хеширование – целостность;
- аутентификация, авторизация, аккаунтинг (протокол AAA) – доступность;
- стеганография – секретность;
- цифровая подпись – неотказуемость;
- цифровое архивирование – сохранность.

Применение этих способов уменьшает потерю ценности информации, но не предотвращает её, так как информационная система организации нуждается в постоянной защите от негативных внешних факторов (опасности).

Опасность – возможность пострадать от угрозы. Состоит из угрозы, наличия уязвимости и возможного вреда.

Безопасность – процесс нейтрализации угроз, уязвимостей, вреда. Составные элементы процесса обеспечения безопасности:

28. неприступность – отсутствие угроз;

- защищенность – невосприимчивость к угрозам;
- надёжность – отсутствие отрицательных последствий.

Обеспечение информационной безопасности – применение необходимых и достаточных мер по защите информации от угроз. Основные направления:

31. защита от атак;

- устранение уязвимостей;
- борьба с последствиями.

Области обеспечения информационной безопасности:

34. теоретическая – на основе закономерностей и опыта:

2. принципы безопасности;

3. модели безопасности;

- нормативно-правовая – на основе юридической ответственности и требованиях:

1. правовые акты;

2. нормативные документы;

- организационно-режимная – на основе правил и порядка:

1. регламенты;

2. режимные меры;

- техническая – на основе техники и автоматизации:

1. инженерно-технические средства;
2. программно-технические средства.

Используемые меры защиты информации необходимо контролировать, обслуживать, обновлять.

Управление информационной безопасностью – деятельность по созданию и поддержанию системы защиты информации, включающая: планирование архитектуры системы, выбор и применение средств защиты информации, выявление и реагирование на инциденты.

Комплексное управление информационной безопасностью называют системой управления информационной безопасностью.

Контрольные вопросы и задания

Чем информация отличается от данных, знаний?

Откуда берётся информация?

Оцените стоимость информации одной из сегодняшних новостей.

В чем разница между критериями, требованиями, свойствами информационной безопасности?

Что такое информационная безопасность?

Проанализируйте с точки зрения конфиденциальности, целостности и доступности информационную безопасность ваших конспектов.

Приведите примеры для каждого способа обеспечения информационной безопасности?

Чем управление информационной безопасностью отличается от обеспечения информационной безопасности?

1.2. Риски информационной безопасности

1. Понятие риска. Определение и структура риска. Термины риск-менеджмента.
2. Классификация угроз, уязвимостей, последствий. Особенности рисков ИБ.
3. Управление рисками. Процесс риск-менеджмента: анализ, оценка, обработка.

Современный системный подход к управлению информационной безопасностью основан на понятии риска и методах управления им.

Риск – это:

- следствие влияния неопределенности на достижение поставленных целей;
- вероятностно-стоимостная оценка потерь;
- сочетание вероятности и последствий наступления неблагоприятных событий;
- неопределённое событие или условие, которое в случае возникновения имеет позитивное или негативное воздействие на репутацию компании, приводит к приобретениям или потерям в денежном выражении;
- вероятностные последствия (отрицательные и положительные);
- угрозы + уязвимости + последствия.

Риск, его структура и способы управления им изучаются в дисциплине «риск-менеджмент».

Термины риск-менеджмента, касающиеся информационной безопасности (ГОСТ Р ИСО/МЭК 31000-2019 Менеджмент риска):

- объект (актив) – непосредственно связанное с целями (деятельностью) (информация, ПО, устройство);
- источник риска – объект или деятельность, которые самостоятельно или в комбинации с другими обладают возможностью вызывать повышение риска (хакеры, вредоносы, сбои);

- событие (инцидент) – возникновение или изменение специфического набора условий (нарушение конфиденциальности, целостности и доступности);
- правдоподобность (появления события) – характеристика возможности и частоты появления события (уязвимость информационной безопасности (недекларированные возможности ПО, ошибки, скрытые каналы);
- последствие (consequence) – результат воздействия события на объект (последствия нарушения конфиденциальности, целостности и доступности (утечка, искажение, отказ);
- угроза – событие без последствий.

Существует множество подходов к классификации риска. Рассмотрим некоторые из них.

Классификация источников угроз:

1. По вероятности:
 1. случайные;
 2. неумышленные;
 3. преднамеренные.
- По источнику:
 1. стихийные – природа;
 2. техногенные – аппаратные/программные средства;
 3. антропогенные – человек-нарушитель:
 - i. Сотрудники:
 1. халатный;
 2. некомпетентный;
 3. излишне компетентный (активный);
 4. вредитель;
 5. инсайдер.

2. Хакеры:

1. любители;
2. исполнители;
3. мстители.

3. Криминал:

1. вымогатели;
2. вандалы;
3. воры;
4. мошенники.

- По направленности:

1. нарушение конфиденциальности;
2. нарушение целостности;
3. нарушение доступа.

Причины уязвимостей:

1. обстоятельства;
- невнимательность;
- 3) попустительство;
- 4) совпадение;
- 5) злой умысел.

Классификация последствий:

- б) по ущербу:
 - а) без последствий;

- б) незначительные;
 - в) некритичные;
 - г) критичные, серьёзные;
 - д) опасные;
 - е) катастрофические.
- 7) по реакции:
- а) игнорируемые;
 - б) учитываемые;
 - в) обрабатываемые;
 - г) приоритетные.

Для выявления совокупности условий и факторов (угрозы, уязвимости), которые приводят или могут привести к риску, а так же для их классификации и анализа составляется модель угроз. Общим для различных моделей угроз является: составление перечня угроз; определение границ и условий реализации угроз, оценка уровня опасности.

Что бы уменьшить количество анализируемых угроз рассматривают только актуальные для конкретной организации угрозы. Определить актуальность угрозы помогает ландшафт угроз – совокупность наиболее распространённых и опасных угроз для определенных активов, определенного вида систем, типовой организации.

Основными требованиями к источникам данных по рискам являются: регулярность появления данных, их объективность и значимость. Все источники данных по ИБ можно разделить на 3 группы: отчеты IT-компаний, содержимое международных и национальных баз, публикации информационно-аналитических центров (табл. 3.).

Таблица 3.

Источники данных по рискам

Источник	Открытость	Многосторонность	Масштаб
Статистика киберугроз от Касперского	+	-	+
ESET Threat Intelligence	-	-	+
Microsoft Security Intelligence Report	+	+	-***
Отчеты CISCO по ИБ	+*	+	+***

Отчеты Positive Technologies	+	+	-
Отчеты Kaspersky ICSysystems CERT	+	_**	+
Ежегодный обзор уязвимостей Flexera	+*	-	+
Отчеты Internet Storm Center	+	-	+
База уязвимостей CVE	+	_**	+
Отчеты CWE	+	_**	+
Аналитические отчеты US-CERT	+	+**	_***
Топ-10 сообщества OWASP	+	-	+
БДУ ФСТЭК	+	_**	-
Данные по киберпреступлениям Statista	-	+	+
Сообщество BISA	+*	+	-
Аналитическая центр компании InfoWatch	+*	+	-
Подписки по безопасности и приватности IDC	-	+	+***
Исследования Ponemon Institute	+	+	+***
Данные Федеральной торговой комиссии США	+	+	-
Портал databreaches	+	_**	+***
«Anti-Malware.ru»	+	+**	+

Деятельность по управлению риском – это непрерывный процесс, в котором можно выделить ключевые этапы. Процесс риск-менеджмента представлен на рис. 1 [1].

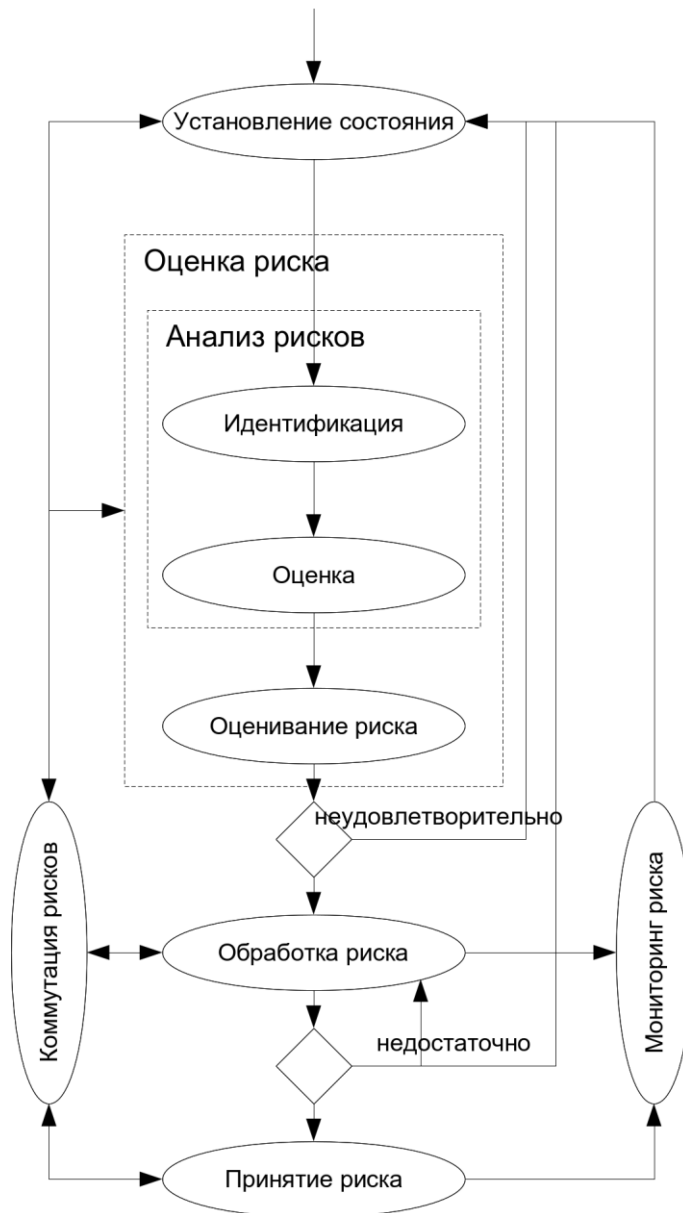


Рис. 1. Процесс риск-менеджмента

Установление состояния включает:

- определение текущей ситуации (контекста);
- установление приоритетов и стратегии;
- формулирование целей и задач;
- согласование критериев оценки.

Виды критериев:

- 8) количественные – величины;
- 9) качественные – метки;
- 10) полуколичественные – шкалы.

В процессе оценки рисков обычно придерживаются следующего порядка:

5. идентификация активов – составление списка активов, которые необходимо защищать;
 - оценка стоимости активов – определение стоимости актива с учётом его роли в деятельности организации;
 - идентификация уязвимостей – составление списка уязвимостей активов, условий их возникновения;
 - идентификация угроз для активов – составление списка источников угроз с учетом их актуальности для защищаемых активов;
 - оценка уязвимостей – определение вероятности возникновения инцидента с учетом имеющихся уязвимостей;
 - вероятностная оценка последствий – определение последствий стоимости инцидента с учетом стоимости активов;
 - оценка стоимости контрмер – определение стоимости средств защиты, мер по устранению уязвимостей, профилактики инцидентов.

При проведении оценки риска важно идентифицировать все риски. Рекомендуется следовать одному из проверенных методов оценки риска, которые можно классифицировать по принципу оценки [2]:

- 11) экспертные – на основе мнения людей-экспертов (структурированный анализ «что-если» (SWIFT));
- 12) аналитические (логико-вероятностные) – на основе расчётов (анализ причин и последствий);

- 13) статистические – на основе имеющихся данных (деревья решений);
- 14) модельно-расчетные – сочетание предыдущих на основе симуляций (метод Монте-Карло).

После оценки риски разбиваются на группы по способу воздействия (обработки):

- 15) нейтрализация – устранение источника риска;
- 16) избежание (предотвращение, уклонение) – отказ от деятельности, обуславливающей риск;
- 17) уменьшение – изменение вероятности или возможности;
- 18) компенсация (снижение) – реагирование на последствия;
- 19) передача (перенос) – делегирование обработки риска другой стороне;
- 20) принятие (сохранение, но не игнорирование) – готовность терпеть последствия;

Также к способам обработки риска относят:

- 21) коммуникацию – обмен информацией о риске;
- 22) мониторинг – наблюдение за инцидентами.

Управление рисками опирается на инфраструктуру риск-менеджмента – систему организации и поддержки процессов. Цель создания инфраструктуры – обеспечить необходимый уровень оценки и обработки риска.

Контрольные вопросы и задания

Что такое риск?

Опишите структуру риска.

Сформулируйте определение риска ИБ.

Приведите примеры современных угроз, уязвимостей, последствий ИБ.

Опишите процесс риск-менеджмента.

Как оценить риск несдачи экзамена?

Приведите примеры способов обработки рисков.

Тема 1.3. Шифрование

1. Криптология. Цели и задачи криптографии и криптологии. Шифрование и расшифровывание.
2. Шифры. Принципы и способы шифрования. Типы шифров.
3. Атаки на шифры. Классификация способов атак на шифры.
4. Цифровая подпись. Виды, принцип создания. Удостоверяющий центр.

Криптология – наука, изучающая методы шифрования и дешифрования. Подразделяется на криптографию (науку обеспечения конфиденциальности при хранении и передаче данных) и криптоанализ (науку об уязвимостях криптографических методов).

Шифрование – это:

- криптопреобразование данных;
- обратимое преобразование информации для сокрытия от недопущенных лиц (в общем смысле).

Шифрование обеспечивает:

- конфиденциальность – доступ только по ключу;
- перманентность – обнаружение вмешательства;
- идентифицируемость – только владельцы ключа.

Российский стандарт шифрования ГОСТ Р 34.12-2015 (Кузнечик) определяет следующие термины.

Открытый текст (plaintext) – незашифрованная информация.

Шифр (cipher) – криптографический метод, используемый для обеспечения конфиденциальности данных, включающий алгоритм зашифрования и алгоритм расшифрования.

Зашифрование (encryption) – обратимое преобразование данных с помощью шифра, которое формирует шифртекст из открытого текста.

Ключ (key) – изменяемый параметр в виде последовательности символов, определяющий криптографическое преобразование.

Итерационный ключ (roundkey) – последовательность символов, вычисляемая в процессе развёртывания ключа шифра и определяющая преобразование на одной итерации блочного шифра.

Шифртекст (ciphertext) – данные, полученные в результате зашифрования открытого текста с целью скрытия его содержания.

Расшифрование (decryption) – операция, обратная к зашифрованию.

Принципы шифрования:

- *принцип независимости* – защита информации, а не ее носителя;
- 2) *принцип обратимости* – должен существовать эффективный способ восстановить исходное сообщение;
- 3) *принципы Керкгоффса* – чем меньше секретов, тем выше безопасность:
 - а) система должна быть физически, если не математически, невскрываемой;
 - б) необходимо, чтобы не требовалось сохранение системы в тайне; попадание системы в руки врага не должно причинять неудобств;
 - в) хранение и передача ключа должны быть осуществимы без помощи бумажных записей; корреспонденты должны располагать возможностью менять ключ по своему усмотрению;
 - г) система должна быть пригодной для сообщения через телеграф;
 - д) система должна быть легко переносимой, работа с ней не должна требовать участия нескольких лиц одновременно;
 - е) от системы требуется, учитывая возможные обстоятельства её применения, чтобы она была проста в использовании, не требовала значительного умственного напряжения или соблюдения большого количества правил;
- 4) *принцип нулевого разглашения* – предоставлять только необходимую информацию.

5) *принцип трудозатрат* – на преодоление защиты необходимы большие затраты.

Защита конфиденциальности с помощью шифрования основана на криптопреобразовании сообщения (рис. 2).

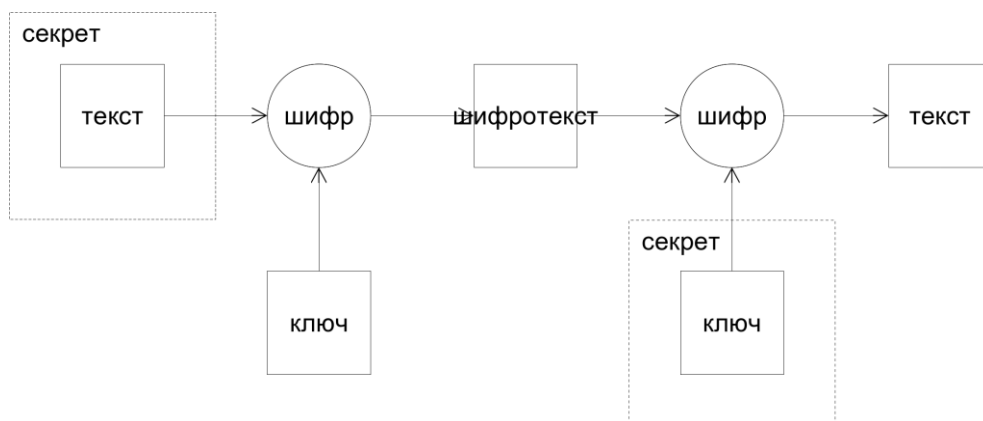


Рис. 2. Процессы шифрования и расшифровывания

Способы шифрования:

- б) перестановки – изменение порядка символов сообщения;
- 7) подстановки – моноалфавитная или полиалфавитная замена символов сообщения;
- 8) динамическое – изменение ключа;
- 9) квантовое – изменение сообщения при попытке перехватить.

Типы шифров по работе с данными:

- 10) Блочные – шифрование по блокам в порядке определяемом режимом. Режимы блочных шифров:
 - а) *electronic code book* (ECB) – режим электронной кодовой книги;
 - б) *cipher block chaining* (CBC) – режим сцепления блоков шифротекста;
 - в) *propagating cipher block chaining* (PCBC) – режим распространяющегося сцепления блоков шифра;

- 4) *cipher feed back* (CFB) – режим обратной связи по шифротексту;
- д) *output feed back* (OFB) – режим обратной связи по выходу;
- е) *counter mode* (CTR) – режим сцепки за шифрованным значением счётчика.

Особенности:

- а) низкая скорость работы – несколько раундов, перекрёстные связи;
 - б) размножение ошибки – невозможность исправления.
- 11) Поточные – шифрование последовательно по символам, с учётом позиции: Особенности:
- а) шаблонность – одинаковое преобразование;
 - б) коррелированность выходного потока с потоком ключа – предсказуемый и неравномерный ключевой поток.

Типы шифров по работе с ключами:

- 12) Симметричные – $K_{ш} = K_p$ (по степени секретности). Проблема: передача ключа.
- 13) Асимметричные – $K_{ш} < K_p$ (по степени секретности). Проблема: низкая скорость работы.
- 14) Комбинированные – $K_{сим}(K_{ас})$. Асимметричный ключ шифруется симметричным шифром.

Шифрование не является 100%-м способом защиты информации. Клод Шеннон доказал существование абсолютно стойких шифров, при соблюдении требований: ключ используется один раз, ключ статистически надежен, длина равна или больше длины сообщения, исходное сообщение избыточно (оценка правильности расшифровки). На практике используются достаточно стойкие шифры – не вскрываемые в течении достаточного времени. Несоблюдение принципов, «слабости» реализации криптосистем могут стать целями атак на зашифрованный текст.

Проблемы шифрования:

- 15) соблюдение принципов;
- 16) сохранение секрета;

17) управление ключами.

Целью атаки может быть восстановление исходного сообщения или вычисление используемого ключа.

Классификация атак и методов «взлома» шифров:

1. *Пассивные* – основаны на возможности перехвата сообщения:
 - 1) если есть только шифротекст – слишком мало информации;
 - 2) если известно исходное сообщение:
 - а) *линейный криптоанализ*.
2. *Активные* – основаны на использовании той же криптосистемы:
 - 1) если есть возможность повторно шифровать с искомым ключом:
 - а) *повтор сообщения*;
 - б) *дифференциальный криптоанализ*;
 - 2) если есть возможность дешифровывать заданные сообщения:
 - а) *перебор (брутфорс)*;
 - б) *алгебраический криптоанализ*;
 - в) *статистический криптоанализ*;
3. *Адаптируемые* – изменение метода в зависимости от результатов.
4. *Побочных каналов* – использование уязвимостей реализации

Современные шифры:

- 18) ГОСТ Р 34.12-2015 (Кузнечик);
- 19) 3DES;
- 20) AES;

- 21) Blowfish;
- 22) IDEA.

Контрольные вопросы и задания

Чем криптография отличается от шифрования?

Опишите принципы шифрования на примере замка и ключа.

Приведите примеры шифров.

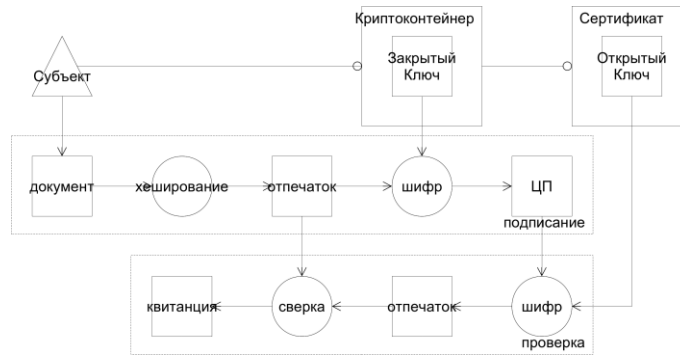
Напишите, какой способ, тип, режим используются в конкретном шифре.

Какие проблемы может решить шифрование, а какие – не решает или создаёт?

Приведите примеры успешных атак на шифр.

Цифровая подпись

-
- Цифровая подпись – способ проверки неотказуемости авторства.
- В отличие от хеширования цифровая подпись формирует специальные зашифрованные данные, содержащие информацию о документе и авторе.
- Принцип работы цифровой подписи (рис. 10):
подписание – асимметричное шифрование хеша документа закрытым ключом;
- проверка – расшифрование и сверка хеша документа с помощью открытого ключа.
-



• Рис. 10. Схема работы цифровой подписи

Ключевым элементом цифровой подписи является удостоверяющий центр (УД, центр сертификации, СА) -- тот, кому доверяют обе стороны, а открытый ключ широко известен. Задача центра сертификации — подтверждать подлинность ключей шифрования с помощью сертификатов электронной подписи.

Контрольные вопросы и задания

Чем цифровая подпись отличается от хеширования с аутентификацией.

Придумайте и проанализируйте свою систему цифровой подписи документов.

Тема 2.1. Иерархия нормативно-правовых документов по информационной безопасности.

1. Иерархия нормативно-правовых документов РФ
3. Виды тайн
2. Государственная система обеспечения информационной безопасности
4. Ответственность за нарушения в сфере ИБ

Иерархия нормативно-правовых документов:

- нормативно-правовые:
 1. конституция РФ,
 2. международные договоры и соглашения,
 3. законы РФ(кодексы, федеральные законы),
 4. указы и распоряжения Президента РФ,
 5. постановления и распоряжения Правительства РФ;
- нормативно-технические:
 1. технические регламенты,
 2. нормативно правовые акты федеральных органов исполнительной власти (приказы),
 3. стандарты государственные (национальные) и организации.

Основополагающие документы по ИБ:

- Конституция РФ
- Статья 23.
- 1. Каждый имеет право на неприкосновенность частной жизни, личную и семейную тайну, защиту своей чести и доброго имени.

- 2. Каждый имеет право на тайну переписки, телефонных переговоров, почтовых, телеграфных и иных сообщений. Ограничение этого права допускается только на основании судебного решения.
- Статья 29.4. Каждый имеет право свободно искать, получать, передавать, производить и распространять информацию любым законным способом. Перечень сведений, составляющих государственную тайну, определяется федеральным законом.
- Статья 29.5. Гарантируется свобода массовой информации. Цензура запрещается.
- Статья 24.1. Сбор, хранение, использование и распространение информации о частной жизни лица без его согласия не допускаются.
- Статья 42. Каждый имеет право на благоприятную окружающую среду, достоверную информацию о ее состоянии и на возмещение ущерба, причиненного его здоровью или имуществу экологическим правонарушением.
- Статья 44. Каждому гарантируется свобода литературного, художественного, научного, технического и других видов творчества, преподавания. Интеллектуальная собственность охраняется законом.
- Доктрина информационной безопасности РФ (Указ Президента от 05.12.2016 г. №646)
 - Национальные интересы в информационной сфере.
 - Основные информационные угрозы.
 - Стратегические цели и основные направления обеспечения.
 - Принципы обеспечения ИБ.
- ФЗ "Об информации, информационных технологиях и о защите информации" №149-ФЗ, 27.07.2006
 - ст.1.1. Настоящий Федеральный закон регулирует отношения, возникающие при:
 - 1) осуществлении права на поиск, получение, передачу, производство и распространение информации;
 - 2) применении информационных технологий;
 - 3) обеспечении защиты информации.

Классификация областей ИБ.

Указ Президента РФ от 06.03.1997 N188 "Об утверждении Перечня сведений конфиденциального характера".

Объекты защиты: информационные системы (ИС), автоматизированные системы управления (АСУ), информационно-телекоммуникационные сети (ИТКС).

- Публичная:
- общедоступная – к общедоступной информации относятся общеизвестные сведения и иная информация, доступ к которой не ограничен. [№149-ФЗ, ст.7.1.].
- Результаты интеллектуальной деятельности и средства индивидуализации -- творческим трудом которого создан такой результат (произведение или логотип) [ГК РФ, Ч.4, авторское право].
- Массовая информация -- предназначенные для неограниченного круга лиц печатные, аудио-, аудиовизуальные и иные сообщения и материалы; ["О средствах массовой информации" 27.12.1991 N2124-1, ст.2]. "О рекламе".
- Государственная:
- Информация, содержащаяся в информационных системах общего пользования [№149-ФЗ] – информация, содержащаяся в государственных информационных системах (реестр Минкомсвязи или вводится приказом гос.органа), а также иные имеющиеся в распоряжении государственных органов сведения и документы являются государственными информационными ресурсами (ГИС), ... в целях реализации полномочий государственных органов и обеспечения обмена информацией между этими органами, а также в иных установленных федеральными законами целях [№149 ст.9,14].
- Государственная тайна -- защищаемые государством сведения в области его военной, внешнеполитической, экономической, разведывательной, контрразведывательной и оперативно-розыскной деятельности, распространение которых может нанести ущерб безопасности Российской Федерации ["О государственной тайне" от 21.07.1993 N 5485-1, ст.2].
- Тайна следствия и судопроизводства -- сведения, составляющие тайну следствия и судопроизводства, сведения о лицах, в отношении которых в соответствии с федеральными законами от 20 апреля 1995 г. N 45-ФЗ "О государственной защите судей, должностных лиц правоохранительных и контролирующих органов" и от 20 августа 2004 г. N 119-ФЗ "О государственной защите потерпевших, свидетелей и иных участников уголовного судопроизводства", другими

нормативными правовыми актами Российской Федерации принято решение о применении мер государственной защиты, а также сведения о мерах государственной защиты указанных лиц, если законодательством Российской Федерации такие сведения не отнесены к сведениям, составляющим государственную тайну [N188 Ук-Пр, п.2]. Сведения, содержащиеся в личных делах осужденных, а также сведения о принудительном исполнении судебных актов, актов других органов и должностных лиц, кроме сведений, которые являются общедоступными в соответствии с Федеральным законом от 2 октября 2007 г. N 229-ФЗ "Об исполнительном производстве" [N188 Ук-Пр, п.7].

- Служебная тайна -- служебные сведения, доступ к которым ограничен органами государственной власти в соответствии с Гражданским кодексом Российской Федерации и федеральными законами [N188 Ук-Пр, п.3].
- «Инфраструктура, которая при выведении из строя или разрушении приведет к катастрофическому и далеко идущему ущербу»(КИИ), научных и кредитно-финансовых организациях, а также предприятиях, работающих в стратегически важных для государства областях: оборонной, топливной и атомной промышленности, в сферах транспорта, энергетики, здравоохранения, связи, в ракетно-космической, горнодобывающей, металлургической и химической промышленности [№187-ФЗ от 26 июля 2017 г. «О безопасности критической информационной инфраструктуры Российской Федерации»].
- Коммерческая
- Коммерческая тайна (N188 Ук-Пр, п.5) -- сведения, связанные с коммерческой деятельностью, доступ к которым ограничен в соответствии с Гражданским кодексом Российской Федерации и федеральными законами). режим конфиденциальности информации, позволяющий ее обладателю при существующих или возможных обстоятельствах увеличить доходы, избежать неоправданных расходов, сохранить положение на рынке товаров, работ, услуг или получить иную коммерческую выгоду ["О коммерческой тайне" от 29.07.2004 N 98-ФЗ].
- Тайна изобретения, ноу-хау -- сведения о сущности изобретения, полезной модели или промышленного образца до официальной публикации информации о них [N188 Ук-Пр, п.6].
- Секрет производства, ноу-хау -- сведения любого характера (производственные, технические, экономические, организационные и другие) о результатах интеллектуальной деятельности в научно-технической сфере и о способах осуществления профессиональной деятельности, имеющие действительную или потенциальную коммерческую ценность вследствие неизвестности их третьим лицам,

если к таким сведениям у третьих лиц нет свободного доступа на законном основании и обладатель таких сведений принимает разумные меры для соблюдения их конфиденциальности, в том числе путем введения режима коммерческой тайны [ГК РФ Ч.4, секрет производства].

- Профессиональная тайна -- сведения, связанные с профессиональной деятельностью, доступ к которым ограничен в соответствии с Конституцией Российской Федерации и федеральными законами (врачебная, нотариальная, адвокатская тайна, тайна переписки, телефонных переговоров, почтовых отправлений, телеграфных или иных сообщений и так далее) [N188 Ук-Пр, п.4] -- «информация, которая становится доступной некоторому кругу лиц при осуществлении своих проф обязанностей».
- Банковская тайна -- тайна об операциях, о счетах и вкладах своих клиентов и корреспондентов. ["О банках и банковской деятельности" от 02.12.1990 N 395-1, ст.26].
- Частная
- Личная, семейная тайна.
- Персональная информация -- сведения о фактах, событиях и обстоятельствах частной жизни гражданина, позволяющие идентифицировать его личность (персональные данные), за исключением сведений, подлежащих распространению в средствах массовой информации в установленных федеральными законами случаях.[N188 Ук-Пр, п.1, "О персональных данных" от 27.07.2006 N 152-ФЗ].
- Тайна связи -- тайна переписки, телефонных переговоров, почтовых отправлений, телеграфных и иных сообщений, передаваемых по сетям электросвязи и сетям почтовой связи [N188 Ук-Пр, п.2, "О связи" от 07.07.2003 N126-ФЗ, ст.63].

ГСЗИ

Государственная система защиты информации представляет собой совокупность органов и исполнителей, используемой ими техники защиты информации, а также объектов защиты, организованная и функционирующая по правилам, установленным соответствующими правовыми, организационно-распорядительными и нормативными документами в области защиты информации. Так же является составной частью системы обеспе-

чения национальной безопасности Российской Федерации и призвана защищать безопасность государства от внешних и внутренних угроз в информационной сфере.

Функционирование государственной системы защиты информации осуществляется на основании законности:

- Конституция Российской Федерации
- ФЗ «О безопасности»
- ФЗ «О государственной тайне»
- ФЗ «Об информации, информатизации и защите информации»
- ФЗ «Об участии в международном информационном обмене»
- Доктрина информационной безопасности Российской Федерации
- Положение о государственной системе защиты информации в Российской Федерации от иностранных технических разведок и от утечки по техническим (утверждено Постановлением Совета Министров – Правительства Российской Федерации от 15 сентября 1993 г. №912–51)
- Указы президента Российской Федерации (№1085 от 16.8.2004 г.)
- Постановления правительства Российской Федерации
- Другие правовые акты федеральных органов власти в области защиты информации

Главными направлениями работ по защите информации являются:

- обеспечение эффективного управления системой защиты информации;
- определение сведений, охраняемых от технических средств разведки, и демаскирующих признаков, раскрывающих эти сведения;
- анализ и оценка реальной опасности перехвата информации техническими средствами разведки, несанкционированного доступа, разрушения (уничтожения) или искажения информации путем преднамеренных программно-технических воздействий в процессе ее обработки, передачи и хранения в технических средствах, выявление возможных технических каналов утечки сведений, подлежащих защите;

- разработка организационно-технических мероприятий по защите информации и их реализация;
- организация и проведение контроля состояния защиты информации.

Деятельность организуют следующие организации (рис.4):

Федеральная служба технического и экспортного контроля (ФСТЭК России) и ее территориальные органы (региональные управления в субъектах Российской Федерации)

Федеральные органы исполнительной власти, другие органы и организации Российской Федерации, руководящие работники которых входят в состав коллегии ФСТЭК России по должности (Минюст, Минобороны, МЧС, МВД, МИД, Минпромэнерго, Минэкономразвития, Минприроды, ФСО, ФСБ, СВР, ГУСП, РАН, ЦБР)

Структурные подразделения по защите информации федеральных органов исполнительной власти, других органов государственной власти и организаций Российской Федерации

Предприятия, проводящие работы с использованием сведений, отнесенных к информации ограниченного доступа, и их подразделения по защите информации

Научно-исследовательские организации по проблемам защиты информации

Организации-разработчики средств защиты информации, защищенных технических средств и средств контроля эффективности защиты информации

Предприятия, оказывающие услуги в области защиты информации

Организации Федерального агентства по техническому регулированию и метрологии (бывшего Госстандарта России), выполняющие работы по стандартизации в области защиты информации

Органы системы лицензирования деятельности в области защиты информации

Органы системы сертификации средств защиты информации

Органы системы аттестации объектов защиты по требованиям безопасности информации

Ответственность за нарушение ИБ

Виды ответственности [Конституция РФ 118.2.]: Судебная власть осуществляется посредством конституционного, гражданского, административного и уголовного судопроизводства.

Уголовный кодекс РФ:

- Статья 2. Задачи Уголовного кодекса Российской Федерации. 1. Задачами настоящего Кодекса являются: охрана прав и свобод человека и гражданина, собственности, общественного порядка и общественной безопасности, окружающей среды, конституционного строя Российской Федерации от преступных посягательств, обеспечение мира и безопасности человечества, а также предупреждение преступлений.
- Статья 44. Виды наказаний. Видами наказаний являются:
 - а) штраф;
 - б) лишение права занимать определенные должности или заниматься определенной деятельностью;
 - в) лишение специального, воинского или почетного звания, классного чина и государственных наград;
 - г) обязательные работы;
 - д) исправительные работы;
 - е) ограничение по военной службе;
 - ж) конфискация имущества;(утратил силу);
 - з) ограничение свободы;
 - з.1) принудительные работы;
 - и) арест;
 - к) содержание в дисциплинарной воинской части;
 - л) лишение свободы на определенный срок;
 - м) пожизненное лишение свободы;
 - н) смертная казнь.

- Глава 28. Преступления в сфере компьютерной информации
- Статья 272. Неправомерный доступ к компьютерной информации. 1. Неправомерный доступ к охраняемой законом компьютерной информации, если это деяние повлекло уничтожение, блокирование, модификацию либо копирование компьютерной информации,
- Статья 273. Создание, использование и распространение вредоносных компьютерных программ. 1. Создание, распространение или использование компьютерных программ либо иной компьютерной информации, заведомо предназначенных для несанкционированного уничтожения, блокирования, модификации, копирования компьютерной информации или нейтрализации средств защиты компьютерной информации,
- Статья 274. Нарушение правил эксплуатации средств хранения, обработки или передачи компьютерной информации и информационно-телекоммуникационных сетей. 1. Нарушение правил эксплуатации средств хранения, обработки или передачи охраняемой компьютерной информации либо информационно-телекоммуникационных сетей и окончного оборудования, а также правил доступа к информационно-телекоммуникационным сетям, повлекшее уничтожение, блокирование, модификацию либо копирование компьютерной информации, причинившее крупный ущерб,
- Статья 274.1. Неправомерное воздействие на критическую информационную инфраструктуру Российской Федерации – для КИИ.
- Статья 146. Нарушение авторских и смежных прав.
- Статья 159.6. Мошенничество в сфере компьютерной информации.
- Статья 137. Нарушение неприкосновенности частной жизни.
- Статья 138. Нарушение тайны переписки, телефонных переговоров, почтовых, телеграфных или иных сообщений.
- Статья 283. Разглашение государственной тайны. Статья 283.1. Незаконное получение сведений, составляющих государственную тайну
-
- Кодекс РФ об административных правонарушениях:
- Статья 1.2. Задачи законодательства об административных правонарушениях. Задачами законодательства об административных правонарушениях являются защита

личности, охрана прав и свобод человека и гражданина, охрана здоровья граждан, санитарно-эпидемиологического благополучия населения, защита общественной нравственности, охрана окружающей среды, установленного порядка осуществления государственной власти, общественного порядка и общественной безопасности, собственности, защита законных экономических интересов физических и юридических лиц, общества и государства от административных правонарушений, а также предупреждение административных правонарушений.

- Статья 3.2. Виды административных наказаний. 1. За совершение административных правонарушений могут устанавливаться и применяться следующие административные наказания:
 - 1) предупреждение;
 - 2) административный штраф;
 - 3) возмездное изъятие орудия совершения или предмета административного правонарушения;(утратил силу. - Федеральный закон от 28.12.2010 N 398-ФЗ);
 - 4) конфискация орудия совершения или предмета административного правонарушения;
 - 5) лишение специального права, предоставленного физическому лицу;
 - 6) административный арест;
 - 7) административное выдворение за пределы Российской Федерации иностранного гражданина или лица без гражданства;
 - 8) дисквалификация;
 - 9) административное приостановление деятельности;(п. 9 введен Федеральным законом от 09.05.2005 N 45-ФЗ)
 - 10) обязательные работы;(п. 10 введен Федеральным законом от 08.06.2012 N 65-ФЗ)
 - 11) административный запрет на посещение мест проведения официальных спортивных соревнований в дни их проведения.(п. 11 введен Федеральным законом от 23.07.2013 N 192-ФЗ)
- 2. В отношении юридического лица могут применяться административные наказания, перечисленные в пунктах 1 - 4, 9 части 1 настоящей статьи.

- 3. Административные наказания, перечисленные в пунктах 3 - 11 части 1 настоящей статьи, устанавливаются только настоящим Кодексом.
-
- Гражданский кодекс РФ:
- Статья 2. Отношения, регулируемые гражданским законодательством. 1. Гражданское законодательство определяет правовое положение участников гражданского оборота, основания возникновения и порядок осуществления права собственности и других вещных прав, прав на результаты интеллектуальной деятельности и приравненные к ним средства индивидуализации (интеллектуальных прав), регулирует отношения, связанные с участием в корпоративных организациях или с управлением ими (корпоративные отношения), договорные и иные обязательства, а также другие имущественные и личные неимущественные отношения, основанные на равенстве, автономии воли и имущественной самостоятельности участников.
- Статья 12. Способы защиты гражданских прав. Защита гражданских прав осуществляется путем:
 - признания права;
 - восстановления положения, существовавшего до нарушения права, и пресечения действий, нарушающих право или создающих угрозу его нарушения;
 - признания оспоримой сделки недействительной и применения последствий ее недействительности, применения последствий недействительности ничтожной сделки;
 - признания недействительным решения собрания;(абзац введен Федеральным законом от 30.12.2012 N 302-ФЗ)
 - признания недействительным акта государственного органа или органа местного самоуправления;
 - самозащиты права;
 - присуждения к исполнению обязанности в натуре;
 - возмещения убытков;
 - взыскания неустойки;
 - компенсации морального вреда;
 - прекращения или изменения правоотношения;

- неприменения судом акта государственного органа или органа местного самоуправления, противоречащего закону;
- иными способами, предусмотренными законом.

Ответственность так же предусмотрена за нарушение технических регламентов и требований к лицензионным видам деятельности.

Контрольные вопросы и задания

Зачем нужна иерархия нормативно-правовых документов?

Что такое конфиденциальная информация?

Какой государственный орган координирует деятельность по защите информации в Российской Федерации?

Чем отличается уголовная ответственность от административной?

Тема 2.2. Система обеспечения информационной безопасности организации

1. Архитектура системы обеспечения информационной безопасности
2. Политика информационной безопасности
3. Проектирование системы менеджмента информационной безопасности
4. Регламенты и правила информационной безопасности

Систему управления защитными мерами организации (рис. 5) называют:

1. СОИБ – система обеспечения информационной безопасности.
2. СУИБ – система управления информационной безопасности.
3. СМИБ – система менеджмента информационной безопасности.
4. ISMS – information security management system.

Принципы СУИБ:

- 1) многоуровневая защита – защита от различных видов атак;
- 2) многослойная (многоэшелонная) защита – подстраховка, дополнение, компенсация;
- 3) модульная архитектура – упрощение создания. Применение модулей обеспечивает:
 - а) стандартизованность,
 - б) тестируемость,
 - в) обновляемость;
- 4) компонентный подход – оптимальное распределение функций;
- 5) минимальное вмешательство человека – меньше зависимость от «человеческого фактора»;
- б) прозрачность – обнаружение недостатков и «узких мест»;

7) секретная безопасность – скрытие недостатков, уязвимостей.

Типичная СУИБ состоит из:

- политика, архитектура,
- регламенты:
- тех./бизнес- процесс,
- контроль доступа к информации, инфраструктуре,
- управление программно-аппаратной инфраструктурой;
- регламент использования средств защиты;
- управление персоналом,
- служба безопасности,
- контроли: средства защиты, правила,
- мероприятия:
- жизненный цикл,
- режимы,
- реагирование.

По функциям СУИБ согласно стандартам Certified Information Systems Security Professional (CISSP) состоит из:

- 1) из подсистемы защиты периметра сети, включающую:
 - а) подсистему обеспечения безопасности межсетевых взаимодействий;
 - б) подсистему фильтрации контента и предотвращения утечки конфиденциальной информации;
- 2) подсистемы обнаружения и предотвращения атак, крупной частью которой является:
 - а) подсистема защиты от вредоносного ПО;

- 3) криптографической подсистемы;
- 4) подсистемы администрирования безопасности, включающую:
 - а) подсистему резервного копирования и восстановления данных;
 - б) подсистему установки обновлений ПО;
- 5) подсистемы мониторинга и аудита безопасности, содержащую:
 - а) подсистему контроля целостности данных.

В структуре СУИБ выделяют :

- 1. Руководящий комитет по надзору, в который входят:
 - 1) исполнительное руководство,
 - 2) владельцы данных,
 - 3) владелец системы,
 - 4) и сотрудник ответственный за безопасность.
- 2. Политика безопасности организации, в которую входят принципы, критерии.
- 3. Проект управления безопасностью, который регулирует правила применения защитных мер.
- 4. Проект управления персоналом, который регулирует правила работы сотрудников.
- 5. Проект управления ресурсами, который регулирует правила использования инфраструктуры.
- 6. Проект мониторинга (надзор), в который входят задачи наблюдение и оценка состояния.
- 7. Проект непрерывности бизнеса, который определяет действия при возникновении инцидентов.
- 8. Проект управления зрелостью, который регулирует совершенствование системы.
- 9. План создания и обслуживания СУИБ, который определяет жизненный цикл системы.

Политика безопасности организации – это:

- совокупность руководящих принципов, правил, процедур и практических приёмов в области безопасности, которые регулируют управление, защиту и распределение ценной информации;
- документ определяющий и регулирующий отношения между защитными мерами.

Виды политик безопасности:

1. По составу:

- 1) организационная – акцент на общую безопасность;
- 2) проблемная – акцент на существенные вопросы безопасности;
- 3) конкретная – акцент на конкретные системы.

2. По требованиям:

- 1) регулирующая – требования;
- 2) рекомендательная – рекомендации;
- 3) информирующая – общие положения безопасности.

Типовое содержание политики безопасности:

1) цели и задачи безопасности:

- а) стратегические;
- б) тактические;
- в) оперативные;

2) требования и базисы:

- а) законы и стандарты;

- б) распределение ответственности;
- 3) модели безопасности и модели защиты;
- 4) архитектура системы защиты;
- 5) общая стратегия построения и использования СУИБ.

Проектирование СУИБ – важный этап защиты организации. Специалистами по ИБ сформулированы рекомендации, которых необходимо придерживаться при проектировании.

Подходы к проектированию:

- 1) сверху-вниз – построение СУИБ начинается с целей руководства;
- 2) снизу-вверх – построение СУИБ исходит от требуемых защитных мер;
- 3) горизонтальный – системный подход, отвечающий на основные вопросы:
 - а) зачем, т. е. определить цели;
 - б) почему, т. е. определить причины создания СУИБ (угрозы);
 - в) что, т. е. определить объект защиты;
 - г) кто, т. е. определить субъекты, ответственные за работу СУИБ;
 - д) где, т. е. определить место, на которое распространяется защита;
 - е) когда, т. е. определить время и режимы;
 - ж) чем и как, т. е. определить способы защиты;
- з) сколько, т. е. определить стоимость СУИБ.

Разработка проекта состоит:

- 1) в назначении ответственных;
- 2) в выборе целей и приоритетов;
- 3) в определении необходимых действий;

4) в спецификации результатов (артефакты), которые необходимо получить.

Проект СУИБ разрабатывается в определённом порядке (плане). План проектирования СУИБ:

1. Определение целей.
2. Формирование политики безопасности и структуры проекта. Этот этап включает:
 - 1) категорирование ресурсов;
 - 2) оценка рисков;
 - 3) права и обязанности;
 - 4) нормативные акты;
 - 5) составление плана и сметы создания СУИБ.
3. Выбор мер и средств защиты.
4. Регламентирование мер безопасности информационных процессов организации, составление правил и режимов работы организации.
5. Обеспечение непрерывности бизнес-процессов. На данном этапе составляются планы:
 - 1) реагирования на инциденты;
 - 2) восстановления от инцидентов;
 - 3) расследования инцидентов.
6. Проведение аудита и тестирования.

В целом СУИБ может следовать каскадной или итеративной модели жизненного цикла.

Каскадная модель состоит из следующих этапов:

- 1) проектирование;

- 2) планирование;
- 3) внедрение;
- 4) аттестация;
- 5) эксплуатация;
- 6) ликвидация.

В итеративной модели неограниченное количество циклов, и каждый цикл проходит фазы:

- 1) PLAN – планирование действий;
- 2) DO – выполнение мероприятия;
- 3) CHECK – проверка полученных результатов;
- 4) ACT – анализ и принятие решений.

В курсах Certified Information Systems Security Professional (CISSP) жизненный цикл СУИБ включает следующее:

1. Планирование и Организация:

- 1) получение одобрения руководства;
- 2) создание руководящего комитета по надзору (oversight steering committee);
- 3) оценка бизнес-драйверов (люди, информация или задачи, которые обеспечивают реализацию бизнес-целей компании);
- 4) создание профиля угроз компании;
- 5) проведение оценки рисков;
- 6) разработка архитектуры безопасности на организационном, прикладном, сетевом и компонентном уровнях;
- 7) определение решений на каждом уровне архитектуры;

8) получение согласия руководства на дальнейшие действия.

2. *Реализация (внедрение):*

1) распределение ролей и обязанностей;

2) разработка и внедрение политики безопасности, процедур, стандартов, базисов и различных руководств;

3) выявление критичных данных на этапах хранения и передачи;

4) реализация следующих проектов:

а) идентификация и управление активами;

б) управление рисками;

в) управление уязвимостями;

г) соответствие требованиям;

д) управление идентификацией и доступом;

е) управление изменениями;

ж) жизненный цикл разработки программного обеспечения;

з) планирование непрерывности бизнеса;

и) обучение и повышение осведомлённости;

к) физическая безопасность;

л) реакция на инциденты;

5) внедрение решений (административных, технических, физических) по каждому проекту;

6) разработка решений по аудиту и мониторингу для каждого проекта;

7) установка целей, соглашений об уровне обслуживания (SLA) и метрик по каждому проекту.

3. *Функционирование и Поддержка:*

- 1) соблюдение установленных процедур для обеспечения базисных уровней в каждом реализованном проекте;
 - 2) проведение внутренних и внешних аудитов;
 - 3) выполнение задач, намеченных в каждом проекте;
 - 4) управление соглашениями об уровне обслуживания по каждому проекту.
4. *Мониторинг и Оценка:*
- 1) анализ лог-файлов, результатов аудита, собранных значений метрик и SLA по каждому проекту;
 - 2) оценка достижения целей по каждому проекту;
 - 3) проведение ежеквартальных встреч с руководящими комитетами;
 - 4) совершенствование действий каждого этапа и их интеграция в фазу *Планирования и Организации*.

Регламенты информационной безопасности

Основные регламенты:

- регламент изменения аппаратно-программной конфигурации системы;
 - процедура выбора и компонентов;
 - процедура установки и модификации компонентов;
 - процедура обслуживания и тестирования;
 - права внесения изменений в конфигурацию аппаратно-программных средств;
 - процедура утилизации;
- регламент использования средств защиты:
 - инструкция по организации антивирусной защиты,

- инструкция по организации защиты сети,
- инструкция по применению криптографической защиты;
- инструкция по защите отдельных узлов и каналов;
- регламент доступа к системе;
- регламент управления персоналом;
- регламент службы безопасности;
- регламент обеспечения непрерывности.

- Вопросы управления доступом:
 - К чему каждый пользователь должен иметь доступ?
 - Кто дает разрешение на доступ и сам доступ?
 - Как принимаются решения о доступе в соответствии с политиками?
 - Остается ли доступ у уволенных сотрудников?
 - Как поддерживать в порядке нашу динамичную и постоянно меняющуюся среду?
 - Каков процесс отзыва прав доступа?
 - Каким образом осуществляется централизованное управление правами доступа и их мониторинг?
 - Почему сотрудники должны помнить по восемь паролей?
 - Как нам централизовать доступ?
 - Как мы управляем доступом наших сотрудников, клиентов, партнеров?
 - Как мы можем убедиться, что мы соответствуем необходимому набору требований?

Правила управления доступом [15]:

- Запретить доступ к системам пользователям, не прошедшим аутентификацию, и анонимным учетным записям.
- Ограничить и контролировать использование административных и иных привилегированных учетных записей.
- Блокировать учетную запись или вносить задержку после нескольких неудачных попыток регистрации.
- Удалять учетные записи уволенных сотрудников сразу же после их ухода из компании.
- Блокировать учетные записи, которые не использовались 30-60 дней.
- Внедрить строгие критерии доступа.
- Применять принцип «должен знать» и принцип минимальных привилегий.
- Отключить ненужные функции системы, службы и порты.
- Заменить пароли «по умолчанию» для встроенных учетных записей.
- Ограничить и контролировать правила глобального доступа.
- Убедиться, что названия учетных записей не раскрывают должностных обязанностей пользователей, которым они принадлежат.
- Удалить излишние правила использования ресурсов учетными записями и группами.
- Удалить из списков доступа к ресурсам излишние идентификаторы пользователей, учетные записи и роли.
- Организовать периодическую смену паролей.
- Установить требования к паролям (по длине, содержанию, сроку действия, распространению, хранению и передаче).
- Организовать журналирование системных событий и действий пользователей, а также периодический просмотр журналов.
- Обеспечить защиту журналов регистрации событий.
- Управление ключами:
 1. хранить только в закрытом виде,
 2. делать резервные копии,

3. передавать только безопасным способом,
4. ограничить срок действия ключа.

Управление персоналом

Основные роли:

- владельцы, руководство;
- персонал: служащие, сотрудники, специалисты, управляющие;
- охранники;
- клиенты, посетители, поставщики.

Принципы работы с посетителями, поставщиками, клиентами:

- разделение контролируемой территории на зоны по степени секретности;
- регламент посещения: часы, области, сопровождение;
- принцип «чистого» рабочего стола;
- множественные критерии выбора поставщиков;
- ответственность поставщика за поставляемое оборудование и работу;
- контроль выполняемых работ;
- мониторинг, анализ и проверка цепочек поставок;
- своевременное информирование клиентов, поставщиков об инцидентах.

Возможные последствия для организации от действий сотрудников:

- разглашение (Р) информации;
- утечка (У) информации;

- несанкционированный доступ (НСД): предоставление нелегального доступа, неправильное использование средств вычислительной техники.
- Основные причины, приводящие к нарушениям со стороны сотрудников:
 - недостаточный уровень знания положений нормативных актов и внутренних организационно-распорядительных документов предприятия, регламентирующих деятельность по защите информации;
 - слабый контроль со стороны руководителей всех уровней за состоянием защиты информации и эффективностью принимаемых мер по недопущению утечки этой информации;
 - недостаточное внимание к вопросам организации работы с персоналом предприятия, изучению морально-деловых качеств сотрудников предприятия, допущенных к конфиденциальной информации;
 - несвоевременное принятие эффективных и действенных мер по предотвращению разглашения персоналом предприятия конфиденциальной информации, а также мер по фактам нарушения норм и правил защиты информации сотрудниками предприятия.

Этапы работы с персоналом:

- найм,
- обучение,
- контроль,
- увольнение.

Принципы принятия на работу:

- уровень фильтрации зависит от выбираемой позиции;
- сотрудники это инвестиции;
- не нанять проще чем уволить;
- учет конфликтов интересов;

- раннее заключение договоров: договор о не разглашении конфиденциальной информации.

При принятии на работу проверяются:

- кандидат:

1. характер,
2. навыки,
3. вредные привычки;

- резюме:

1. образование,
2. стаж,
3. сертификаты;

- биография:

1. долги,
2. судимости,
3. наркотики.

Формы обучения:

- допуск к работе:

1. ознакомление с правилами допуска к работе,
2. инструктажи на рабочем месте,

- повышение осведомленности:

1. ознакомление с действующей политикой и регламентами (регламент действий пользователя), стандартами,
2. семинары, вебинары, конференции и т.п.;

- повышение квалификации;

- переподготовка;
- стажировка.

Обучение включает в себя так же:

1. контроль сотрудника при выездных формах обучения;
2. оценка результатов;
3. получение подписей о прохождении.

Принципы контроля персонала:

- отчет снизу-вверх,
- контроль сверху-вниз,
- выявление необычных действий,
- информируемость.

Направления деятельности по предотвращению нарушений:

- изучение морально-деловых качеств сотрудников предприятия;
- повышение ответственности сотрудников всех категорий за сохранение в тайне доверенных по службе сведений конфиденциального характера, например: ротация обязанностей, разделение обязанностей, разделение знаний, двойное управление;
- проведение профилактической работы по предупреждению (исключению) утечки конфиденциальной информации путем ее разглашения;
- повышение уровня теоретических знаний и практических навыков сотрудников в вопросах защиты конфиденциальной информации – обучение;
- создание и поддержание устойчивого морально-психологического климата в коллективе предприятия;
- создание и применение системы стимулирования труда сотрудников, допущенных к конфиденциальной информации.

Методы мотивации:

- непосредственная:
 1. убеждение,
 2. внушение,
 3. агитация;
- властная:
 1. указание,
 2. приказ,
 3. распоряжение и др.;
- стимулирующая:
 1. моральная,
 2. материальная,
 3. трудовая.

Действия при увольнении:

- немедленное лишение прав доступа к системе;
- запрет свободного перемещения по территории организации;
- возврат имущества организации;
- наблюдение за увольняемым;
- договор о лояльности.

Контрольные вопросы и задания

Опишите систему управления защитой выбранной организации.

Чем план отличается от проекта?

Приведите примеры элементов структуры проекта.

Опишите процесс перехода от проекта СМИБ к плану его построения.

Что входит в минимальный набор (ядро) составляющий политику безопасности?

Как «заставить» сотрудника соблюдать правила безопасности?

Как регламентировать телефонные звонки и переписку сотрудников?

Допустимо ли требовать от уволенного сотрудника соблюдение конфиденциальной информации?

Тема 3.1. Средства защиты информации

1. Защита служб. Антивирусы.
2. Система восстановления. Резервные копии, Транзакции. RAID.
3. Контроль периметра. Сетевые экраны. Демилитаризованная зона (DMZ)
4. Средства мониторинга. Система обнаружения атак (IDS, IPS). Системы защиты от утечек (DLP).

Цель применения *программно-аппаратных средств защиты информации (СЗИ)* – защита информационных процессов.

Для повышения надёжности и эффективности СЗИ часть функций реализуется в виде аппаратных устройств.

По назначению СЗИ подразделяется:

- 1) на *средства администрирования* – внедрение СЗИ и управление их работой;
- 2) *средства контроля периметра* – создание и поддержание границы периметра;
- 3) *средства защиты служб* – обеспечение правильной работы информационных процессов;
- 4) *средства восстановления* – обеспечение непрерывной работы информационных процессов;
- 5) *средства мониторинга* – выявление и реагирование на инциденты;
- 6) *вспомогательные средства защиты* – реализация отдельных методов безопасности.

Рассмотрим наиболее значимые программно-аппаратные средства защиты информации.

Вредоносная программа (зловред, malware) – код, осуществляющий негативное воздействие на состояние и работу системы.

Классификация зловредов:

1) вредоносные:

- а) virus – вредоносный, саморазмножающийся код;
- б) logic bomb – вредоносный код, запускающийся при определенных условиях;
- в) worm – самостоятельно распространяющийся вредонос;
- г) trojan – вредонос, распространяющийся обманом;
- д) EICAR-Test-File (European Institute for Computer Antivirus Research) – 16-битный COM-файл, выводящий сообщение;

2) загрязняющие (вспомогательные):

- а) rootkit – утилита, действующая в ядре операционной системы;
- б) backdoor – средство скрытого несанкционированного доступа;
- в) exploit – способ использования уязвимости системы, позволяющий внедрять вредоносный код;
- г) shellcode – код предоставляющий доступ к командной консоли системы;
- д) spyware – утилиты для незаконного сбора данных;

3) потенциально-вредоносные (нежелательные):

- а) adware – «надоедливые», мешающие программы;
- б) shovelware – бесполезные, самоустанавливающиеся программы;
- в) вредоносные утилиты – специальные инструменты для проведения анализа и атак;
- г) подозрительные упаковщики – утилиты для сокрытия зловредов.

Антивирусные технологии:

- 1) сигнатурный анализ – обнаружение по заданному шаблону;

- 2) эвристический анализ – обнаружение по критериям;
- 3) анализ поведения – обнаружение подозрительного поведения с помощью: эмуляции кода и песочницы;
- 4) анализ целостности – обнаружение вмешательства;
- 5) облачная проверка – проверка внешними экспертами и экспертными системами.

Режим работы антивирусных средств:

- 1) по требованию;
- 2) фоновый – во время выполнения операций;
- 3) по расписанию;
- 4) автоматический – при появлении внешних событий.

Параметры антивирусов:

- 1) расписание проверок: время, период, события;
- 2) области проверки: оперативная память, загрузочная область, диски, внешние хранилища;
- 3) объекты проверки: типы вредоносных, типы файлов, архивы, выполняемые действия;
- 4) способ лечения: удаление, очистка, карантин, уведомление;
- 5) расписание обновлений: время, период, события;
- 6) источники обновления: официальные, локальные;
- 7) самозащита: изменение настроек, остановка задач.

Современные антивирусные программы позволяют также контролировать активность программ и пользователя, содержат функции настройки и оптимизации системы.

Система резервного копирования

Система резервного копирования в соответствии с политикой, автоматически создает резервные копии данных и системы. В случае сбоев, после расследования инцидента, возможно восстановить испорченные данные и систему. Восстановление обычно происходит вручную, по требованию пользователя.

В регламенте резервного копирования определяется:

- Частота резервного копирования:
 - периодическая;
 - плановая;
 - по событию.
- Принцип копирования:
 1. Физическое (клонирование) – копирование по блокам. Достоинства: возможен пропуск свободных блоков. Недостатки: сохраняются ненужные элементы, дефектные блоки.
 2. Логическое (копирование) – копирование рекурсивно по структуре данных. Достоинства: сохраняется структура, возможно пропустить ненужный элемент, возможно восстановление отдельного элемента. Недостатки: восстановление возможно только в такой же системе, требуется больше места на сохранение структуры и связей, занимает больше времени.
- Процесс копирования: на горячую – система продолжает работать, на холодную – система отключается во время копирования.
- Вид резервной копии:
 1. полная – содержит все данные, нужные для восстановления;
 2. разностная – содержит только новые и изменённые данные по сравнению с предыдущей полной копией;
 3. инкрементная – содержит новые и изменённые данные от другой инкрементной копии;
 4. декрементная – содержит старые и изменённые данные, по сравнению с полной текущей копией;

- Место хранения резервных копий:

1. локально,
2. на отдельном носителе,
3. удаленно, по сети
4. в облаке.

- Ротация резервных копий – схема хранения и обновления:

1. одноразовая – хранится только одна копия;
2. простая – хранится последовательность копий (износ хранилища);
3. дед-отец-сын – периодически (раз месяц) создается полная копия (дед), по событию или раз в неделю создается дифференциальная копия (отец), регулярно, например один раз в день делается инкрементная копия (сын);
4. хайнойская башня – одна полная, много инкрементных;
5. 10 наборов - циклическая смена носителей;

- Дополнительные параметры: сжатие, шифрование.

Для повышения надежности хранения данных используется RAID (Redundant Array of Independent Disks) – избыточный массив независимых (самостоятельных, раньше было – дешовых) дисков). Несколько физических дисковых устройств объединяются в логический модуль для повышения отказоустойчивости и (или) производительности.

Петтерсон с коллегами из Беркли представили спецификации пяти уровней RAID, которые стали стандартом де факто:

- RAID 1 — зеркальный дисковый массив;
- RAID 2 — зарезервирован для массивов, которые применяют код Хемминга;
- RAID 3 — дисковый массив с выделенным диском чётности;
- RAID 4 — дисковый массив с чередованием и выделенным диском чётности;
- RAID 5 — дисковый массив с чередованием, в том числе данных чётности (нет диска, выделенного для хранения чётности — блоки чётности чередуются с блоками данных на каждом диске).

Межсетевой экран

Сетевой экран (firewall, brandmauer) это фильтрующий сетевой шлюз.

Цель применения – блокировка нежелательного сетевого трафика.

По уровню вмешательства, сетевые экраны классифицируют следующим образом:

- 1) канальный – мост с фильтрацией;
- 2) сетевой – пакетный фильтр, отдельно выделяют пакетный сетевой экран с отслеживанием состояний (statefull);
- 3) сеансовый – исключение прямых соединений;
- 4) приложений – проксирование соединений;
- 5) экспертный уровень – на нескольких уровнях, специализация.

По расположению сетевые экраны бывают:

- 1) периметровые – на границе с внешней сетью;
- 2) межсетевые – на сетевых устройствах;
- 3) персональные – на каждом компьютере.

Используя сетевые экраны, организуются специальные сетевые зоны, с определёнными правилами.

Демилитаризованная зона (DMZ) – изолированный сегмент сети. Виды зон:

- 1) «сэндвич» – между сетевыми экранами;
- 2) «тупик» – выделенный сегмент;
- 3) «остров» – на периметре сети.

Кроме ограничения трафика сетевые экраны, благодаря их расположению, могут осуществлять функции транзита и преобразования трафика.

Система выявления и предотвращения вторжений

Система выявления вторжений (IDS, intrusion detection system) – система детектирования и уведомления о подозрительных действиях.

Система предотвращения вторжений (IPS, intrusion prevention system) – это IDS с возможностью принимать действия по защите от атак.

Структура IDS:

1. Сенсоры – собирают трафик и данные о действиях.
2. Анализаторы – ищут подозрительные действия.
3. Административные интерфейсы – принимают сообщения от анализатора для последующей обработки.

Виды IDS:

- 1) host-based (HIPS/HIDS) – следят за состоянием компьютеров;
- 2) network-based (NIPS/NIDS) – следят за сетевым трафиком.

Методы обнаружения вторжений:

1. Сигнатурный (signature based) – ищет сигнатуры (признаки) атак, выделяют:
 - 1) отслеживающий шаблоны (pattern matching) – анализирует данные (файлы, трафик), свидетельствующие об атаке;
 - 2) отслеживающий состояние (stateful matching) – анализирует последовательность действий, приводящих в запрещённые состояния.
2. Основанный на аномалиях (anomaly based) – рассчитывает и сравнивают рейтинг «аномальности» данных с порогом, выделяют:

- 1) основанный на статистических аномалиях (statistical anomaly-based) – строит профиль «нормальной» деятельности;
- 2) основанный на аномалиях протоколов (protocol anomaly-based) – выделяет действия, эксплуатирующие уязвимости;
- 3) основанный на аномалиях трафика (traffic anomaly-based) – выделяет необычное (новое) использование системы.
3. Основанный на правилах (rule-based) или эвристический (heuristic-based) – экспертные системы, основанные на базе знаний (knowledge base), механизме логических выводов (inference engine) и программировании на основе правил (rule-based programming).

Система предотвращения утечек

Data Leak Prevention (DLP) – технологии предотвращения утечек конфиденциальной информации из информационной системы вовне.

DLP-системы строятся на анализе потоков данных, пересекающих периметр защищаемой информационной системы. Точками анализа являются: носители, сеть, потоки ввода-вывода.

Распознавание конфиденциальной информации в DLP-системах производится двумя способами:

- 1) анализ формальных признаков (например грифа, документа, специально введённых меток, сравнением хеш-функции);
2. анализ контента (содержимого).

Ошибки при работе DLP:

3. ложное срабатывание (более вероятны для 2-го способа проверки);
4. пропуск конфиденциальной информации (более вероятны для 1-го способа).

Проблемы использования DLP-систем:

5. шифрование, стеганография – для проверки необходимо дешифровать/найти данные;
- б) право на частную жизнь – проверка трафика не может противоречить праву на частную жизнь и тайну связи.

Другие СЗИ

DDOS-prevention, криптоутилиты, сканеры уязвимостей, системы аутентификации, службы каталогов, SIEM-системы, программно-аппаратные замки, система резервного копирования.

Контрольные вопросы и задания

Существуют ли только аппаратные средства защиты информации?

Какие СЗИ используются в вашей организации для обеспечения конфиденциальности, целостности, доступности?

***Является ли вредоносным объект с хеш-суммой
d4b7b4b1aac1c51d1eec18bc5ca26fbb3053af5b9fee2f6ae0966b2981edbbcb?***

Какие порты используются протоколом SSH?

Тема 3.2. Проверка информационной безопасности

1. Проверка информационной безопасности. Цели и задачи, способы оценки ИБ.
2. Аудит. Цели, принципы, виды аудита. Требования к аудитору.
3. Пентестинг. Методы и средства тестирования.

Лицензирование

Контроль деятельности организации в области ИБ осуществляется с помощью лицензирования.

ФЗ «О лицензировании отдельных видов деятельности» № 99-ФЗ, 04.05.2011:

- ст.2. Цели, задачи лицензирования отдельных видов деятельности и критерии определения лицензируемых видов деятельности. 3. К лицензируемым видам деятельности относятся виды деятельности, осуществление которых может повлечь за собой нанесение указанного в части 1 настоящей статьи ущерба и регулирование которых не может осуществляться иными методами, кроме как лицензированием.
- ст.3. Основные понятия, используемые в настоящем Федеральном законе: 2) лицензия - специальное разрешение на право осуществления юридическим лицом или индивидуальным предпринимателем конкретного вида деятельности (выполнения работ, оказания услуг, составляющих лицензируемый вид деятельности), которое подтверждается документом, выданным лицензирующим органом на бумажном носителе или в форме электронного документа, подписанного электронной подписью, в случае, если в заявлении о предоставлении лицензии указывалось на необходимость выдачи такого документа в форме электронного документа;
- ст.1. 2. Положения настоящего Федерального закона не применяются к отношениям, связанным с осуществлением лицензирования:
 1. 1) использования атомной энергии;
 2. 2) производства и оборота этилового спирта, алкогольной и спиртосодержащей продукции;
 3. 3) деятельности, связанной с защитой государственной тайны;
 4. 4) деятельности кредитных организаций;

5. 5) деятельность по проведению организованных торгов;
6. 6) видов профессиональной деятельности на рынке ценных бумаг;
7. 7) деятельности акционерных инвестиционных фондов, деятельности по управлению акционерными инвестиционными фондами, паевыми инвестиционными фондами, негосударственными пенсионными фондами;
8. 8) деятельности специализированных депозитариев инвестиционных фондов, паевых инвестиционных фондов и негосударственных пенсионных фондов;
9. 9) деятельности негосударственных пенсионных фондов по пенсионному обеспечению и пенсионному страхованию;
10. 10) клиринговой деятельности;
11. 11) страховой деятельности.

- ст.12. Перечень видов деятельности, на которые требуются лицензии (57 видов):

1. 1) разработка, производство, распространение шифровальных (криптографических) средств, информационных систем и телекоммуникационных систем, защищённых с использованием шифровальных (криптографических) средств, выполнение работ, оказание услуг в области шифрования информации, техническое обслуживание шифровальных (криптографических) средств, информационных систем и телекоммуникационных систем, защищённых с использованием шифровальных (криптографических) средств (за исключением случая, если техническое обслуживание шифровальных (криптографических) средств, информационных систем и телекоммуникационных систем, защищённых с использованием шифровальных (криптографических) средств, осуществляется для обеспечения собственных нужд юридического лица или индивидуального предпринимателя);
 1. ПП "Об утверждении Положения о лицензировании деятельности по разработке, производству, распространению шифровальных (криптографических) средств..." №313, 16.04.2012
 2. 2) разработка, производство, реализация и приобретение в целях продажи специальных технических средств, предназначенных для негласного получения информации;
1. ПП РФ "Об утверждении Положения о лицензировании деятельности по разработке, производству, реализации и приобретению в целях продажи специальных технических средств, предназначенных для негласного получения информации" №287, 12.04.2012

3. 3) деятельность по выявлению электронных устройств, предназначенных для негласного получения информации (за исключением случая, если указанная деятельность осуществляется для обеспечения собственных нужд юридического лица или индивидуального предпринимателя);
 1. Постановление Правительства РФ "Об утверждении Положения о лицензировании деятельности по выявлению электронных устройств, предназначенных для негласного получения информации (за исключением случая, если указанная деятельность осуществляется для обеспечения собственных нужд юридического лица или индивидуального предпринимателя)" №314, 16.04.2012
4. 4) разработка и производство средств защиты конфиденциальной информации;
 1. Постановление Правительства РФ "О лицензировании деятельности по разработке и производству средств защиты конфиденциальной информации" №171, 03.03.2012
5. 5) деятельность по технической защите конфиденциальной информации;
 1. Постановление Правительства РФ "О лицензировании деятельности по технической защите конфиденциальной информации" №79, 03.02.2012

Ответственность за нарушения требований к лицензионной деятельности: обычно административная – штраф, конфискация, при крупном ущербе или получении крупного дохода – уголовная (171 УК РФ).

Сертификация

Требования к продукции и услугам устанавливаются с помощью сертификации и техническим регулированием.

ФЗ "О техническом регулировании" №184-ФЗ, 27.12.2002:

1. ст.2. Основные понятия:
 - 1.2. стандарт – документ, в котором в целях добровольного многократного использования устанавливаются характеристики продукции, правила осуществления и характеристики процессов проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации и утилизации, выполнения работ или оказания услуг. Стандарт также может содержать правила и методы исследований (испытаний) и измерений, правила отбора

образцов, требования к терминологии, символике, упаковке, маркировке или этикеткам и правилам их нанесения;

- 1.3. технический регламент – документ, который принят международным договором Российской Федерации, подлежащим ратификации в порядке, установленном законодательством Российской Федерации, или в соответствии с международным договором Российской Федерации, ратифицированным в порядке, установленном законодательством Российской Федерации, или федеральным законом, или указом Президента Российской Федерации, или постановлением Правительства Российской Федерации, или нормативным правовым актом федерального органа исполнительной власти по техническому регулированию и устанавливает обязательные для применения и исполнения требования к объектам технического регулирования (продукции или к продукции и связанным с требованиями к продукции процессам проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации и утилизации);
- 1.4. сертификация – форма осуществляемого органом по сертификации подтверждения соответствия объектов требованиям технических регламентов, положениям стандартов, сводов правил или условиям договоров;
- 1.5. сертификат соответствия - документ, удостоверяющий соответствие объекта требованиям технических регламентов, положениям стандартов, сводов правил или условиям договоров;
- 1.6. декларация о соответствии - документ, удостоверяющий соответствие выпускаемой в обращение продукции требованиям технических регламентов;
2. ст.5. Особенности технического регулирования в отношении оборонной продукции (работ, услуг), поставляемой по государственному оборонному заказу, продукции (работ, услуг), используемой в целях защиты сведений, составляющих государственную тайну или относимых к охраняемой в соответствии с законодательством Российской Федерации иной информации ограниченного доступа, продукции (работ, услуг), сведения о которой составляют государственную тайну, продукции, для которой устанавливаются требования, связанные с обеспечением безопасности в области использования атомной энергии, процессов проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации, утилизации, захоронения указанной продукции
- 2.2.2. Особенности технического регулирования в части разработки и установления обязательных требований государственными заказчиками, федеральными органами исполнительной власти, уполномоченными в области обеспечения безопасности, обороны, внешней разведки, противодействия техническим разведкам и технической защиты информации, государственного управления использованием атомной

энергии, государственного регулирования безопасности при использовании атомной энергии, в отношении продукции (работ, услуг), указанной в пункте 1 настоящей статьи, а также соответственно процессов ее проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации, утилизации, захоронения устанавливаются Президентом Российской Федерации, Правительством Российской Федерации в соответствии с их полномочиями.

2.3.4. Особенности оценки соответствия продукции (работ, услуг), указанной в пункте 1 настоящей статьи, а также соответственно процессов ее проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации, утилизации, захоронения устанавливаются Правительством Российской Федерации или уполномоченными им федеральными органами исполнительной власти.

3. ст.6. Цели принятия технических регламентов

3.2.1. Технические регламенты принимаются в целях:

3.2.1. защиты жизни или здоровья граждан, имущества физических или юридических лиц, государственного или муниципального имущества;

3.2.2. охраны окружающей среды, жизни или здоровья животных и растений;

3.2.3. предупреждения действий, вводящих в заблуждение приобретателей, в том числе потребителей;

3.2.4. обеспечения энергетической эффективности и ресурсосбережения.

3.3.2. Принятие технических регламентов в иных целях не допускается.

4. ст.11. Цели стандартизации. Целями стандартизации являются:

4.2. повышение уровня безопасности жизни и здоровья граждан, имущества физических и юридических лиц, государственного и муниципального имущества, объектов с учетом риска возникновения чрезвычайных ситуаций природного и техногенного характера, повышение уровня экологической безопасности, безопасности жизни и здоровья животных и растений;

4.3. обеспечение конкурентоспособности и качества продукции (работ, услуг), единства измерений, рационального использования ресурсов, взаимозаменяемости технических средств (машин и оборудования, их составных частей, комплектующих изделий и материалов), технической и информационной совместимости, сопоставимости результатов исследований (испытаний) и измерений, технических и экономико-статистических данных, проведения анализа характеристик продукции (работ,

- услуг), планирования и осуществления закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд, добровольного подтверждения соответствия продукции (работ, услуг);
- 4.4.содействие соблюдению требований технических регламентов;
- 4.5.создание систем классификации и кодирования технико-экономической и социальной информации, систем каталогизации продукции (работ, услуг), систем обеспечения качества продукции (работ, услуг), систем поиска и передачи данных, содействие проведению работ по унификации.
5. Статья 18. Цели подтверждения соответствия. Подтверждение соответствия осуществляется в целях:
- 5.2.удостоверения соответствия продукции, процессов проектирования (включая изыскания), производства, строительства, монтажа, наладки, эксплуатации, хранения, перевозки, реализации и утилизации, работ, услуг или иных объектов техническим регламентам, стандартам, сводам правил, условиям договоров;
- 5.3.содействия приобретателям, в том числе потребителям, в компетентном выборе продукции, работ, услуг;
- 5.4.повышения конкурентоспособности продукции, работ, услуг на российском и международном рынках;
- 5.5.создания условий для обеспечения свободного перемещения товаров по территории Российской Федерации, а также для осуществления международного экономического, научно-технического сотрудничества и международной торговли.
6. Статья 19. Принципы подтверждения соответствия
- 6.2.1. Подтверждение соответствия осуществляется на основе принципов:
- 6.2.1. доступности информации о порядке осуществления подтверждения соответствия заинтересованным лицам;
- 6.2.2. недопустимости применения обязательного подтверждения соответствия к объектам, в отношении которых не установлены требования технических регламентов;
- 6.2.3. установления перечня форм и схем обязательного подтверждения соответствия в отношении определенных видов продукции в соответствующем техническом регламенте;

- 6.2.4. уменьшения сроков осуществления обязательного подтверждения соответствия и затрат заявителя;
- 6.2.5. недопустимости принуждения к осуществлению добровольного подтверждения соответствия, в том числе в определенной системе добровольной сертификации;
- 6.2.6. защиты имущественных интересов заявителей, соблюдения коммерческой тайны в отношении сведений, полученных при осуществлении подтверждения соответствия;
- 6.2.7. недопустимости подмены обязательного подтверждения соответствия добровольной сертификацией.

Для подтверждение соответствия необходимы процедуры проверки существующих мер защиты информации.

Задачи проверки информационной безопасности:

- 1) выбор и обоснование базисов (ориентиры) – проверка адекватности политики;
- 2) анализ активов, ресурсов и функций – требования стандартов:
 - а) классификация/категорирование,
 - б) определение приоритетов,
 - в) определение уровней критичности;
- 3) анализ рисков;
- 4) проверка состояния.

В проверках принимают участие:

- 1) владельцы активов;
- 2) специалисты;
- 3) служащие безопасности;
- 4) эксперты по рискам;

- 5) аналитики по безопасности;
- 6) аналитики данных;
- 7. аудиторы.

Способы проверки ИБ:

- 1) Испытание – проверка возможностей ИБ. Выделяют:
 - а) *сканирование* – проверка работоспособности защитных мер и наличие уязвимостей;
 - б) *проникновение* – использование проверочных атак.
- 2) Сравнение – сверка с нормами и требованиями. Выделяют:
 - а) *аудит* – проверка отчетности на соответствие требованиям;
 - б) *оценка рисков* – сравнение средств защиты с опасностями.
- 3) Анализ – изучение реальных случаев. Выделяют:
 - а) *статистический анализ* – анализ различных инцидентов;
 - б) *заманивание* – анализ действий злоумышленника.

В результате проверки дается оценка состояния защиты, необходимая для принятия соответствующих мер защиты.

Рассмотрим подробнее такие способы проверки, как аудит и тестирование на проникновение.

Аудит

Аудит – систематический, независимый и документированный процесс установления степени соответствия установленным критериям.

Цели проведения аудита:

- 1) оценка соответствия стандартам и нормам (выдача сертификата);
- 2) оценка уровня качества;
- 3) выработка рекомендации, локализация узких мест.

Принципы аудита:

- 1) *этичность* – информация об уязвимостях и проблемах не должна выйти за пределы организации;
2. *беспристрастность* – непредвзятость оценки;
3. *независимость* – отсутствие заинтересованности в результатах;
4. *компетентность* – должна проводиться специалистом, имеющим необходимые знания, умения;
5. *документальность* – результаты проверки фиксируются в отчётах.

Выделяют внешний и внутренний аудит.

Внешний – проводится сторонней организацией.

Достоинства внешнего аудита:

6. независимость,
7. квалификация специалистов,
8. опыт.

Недостатки внешнего аудита:

9. стоимость,
10. возможность утечки,
11. необходимость искать, приглашать специалистов.

Внутренний (самоаудит) – проводится сотрудниками организации.

Достоинства:

12. осведомлённость о внутреннем устройстве организации,
13. возможность регулярного проведения.

Недостатки:

14. субъективность,
15. загруженность проверяющих другими обязанностями.

В сфере информационной безопасности дополнительно выделяют аудит:

- 1) *активный* – исследование защищенности с точки зрения злоумышленника;
- 2) *экспертный* – сравнение с «идеальной» системой безопасности, с точки зрения опыта экспертов;
- 3) *нормативный* – на соответствие требованиям стандартов по информационной безопасности.

Результаты аудита зависят от аудитора и его компетенции. Компетентность аудитора:

- 1) *Образование*: высшее техническое образование в области, в которой проводится аудит.
- 2) *Знания*:
оцениваемые процессы и критерии оценки;
- 2) принципы аудита;
- 3) методология аудита;
- 4) инструментарий аудита;
- 5) методология анализа и управления рисками;
- е) законодательная и нормативная базу.

3) *Навыки и умения:*

- а) технический писатель – оформление отчетов;
- б) консультант – составление заключений и рекомендаций;
- в) переговорщик – опрос сотрудников;
- г) посредник – между организацией и проверяющими органами;
- д) исследователь – проведение испытаний.

4) *Личные качества (этика аудитора):*

- а) Институт SANS, основанный в 1989 году и готовящий профессионалов в компьютерной безопасности, имеет свой собственный «кодекс этики SANS»:
 - 1) я буду стремиться к познанию себя и к тому, чтобы быть честным в своих возможностях;
 - 2) я буду вести свой бизнес так, чтобы ИТ-профессия считалась честной и профессиональной;
 - 3) я уважаю частную жизнь и конфиденциальность.
- б) Кодекс этики аудитора и правила ассоциации аудита и контроля информационных систем ISACA.
- в) Кодекс международного консорциума по сертификации в области безопасности информационных систем (ISC)²[17]:
 - 1) Поступай честно, справедливо, ответственно, в рамках закона. Защищай всеобщее благополучие.
 - 2) Усердно трудись, предоставляй качественные услуги и развивай сферу безопасности.
 - 3) Поощряй увеличение количества исследований: обучай, направляй и отдавай должное сертификации.
 - 4) Избегай небезопасных действий, оберегай и усиливай целостность общественных инфраструктур.
 - 5) Придерживайся соглашений, гласных и негласных. Давай разумные советы.

- б) Избегай любого конфликта интересов, уважай веру других людей в себя, берись только за ту работу, выполнить которую тебе под силу.
- 7) Сохраняй и обновляй навыки, не участвуй в мероприятиях, которые могут навредить репутации других профессионалов.;
- г) Кодекс института внутренних аудиторов ИА4.
- д) Совет по аудиторской деятельности при Минфине РФ – Кодекс этики аудитора России (протокол № 16 от 28.08.2003).
- е) этические принципы сообщества профессионалов в области информационной безопасности RISSPA5 (Россия).
- 5) *Опыт* (сертификация аудитора):
 - а) certified information systems auditor (ISACA);
 - б) certified information security manager (ISACA2);
 - в) certified internal auditor (IIA);
 - г) certified public accountant (AICPA);
 - д) certified information systems security professional (ISCP);
 - е) certified systems security practitioner (ISC2);
 - ж) chartered accountant (CICA);
 - з) GIAC certified security engieener (SANS);
 - и) certified protection professional (ASIS).

Процесс проведения аудита:

- 1) Определение входных данных для проведения аудита:
 - а) цели;
 - б) сфера, особенности;
 - в) ограничения;

- Г) подходы: нисходящий (исходя из требований), восходящий (в зависимости от имеющихся проблем);
- Д) критерии.
- 2) Определение ролей и обязанностей:
 - а) заказчик – определяет цели;
 - б) руководитель – определяет критерии и ограничения;
 - в) организатор – планирует;
 - г) специалист – проводит аудит.
- 3) Выбор модели проведения аудита. В зависимости от целей выбирают:
 - а) модель оценки по показателям ИБ – оценка соответствия;
 - б) модель зрелости – оценка качества.
- 4) Проведение оценивания. Типичные шаги:
 - а) планирование;
 - 2) сбор данных;
 - 3) проверка достоверности.
- 5. Оформление результатов. Могут быть получены:
 - а) аналитический отчёт;
 - б) заключение аудита;
 - 3) свидетельство соответствия степени;
 - 4) сертификаты, аттестаты соответствия.

Тестирование на проникновение

Тестирование на проникновение (тесты на преодоление защиты, penetration testing, pentest, пентест) – метод оценки безопасности компьютерных систем или сетей средствами моделирования атаки злоумышленника.

Цель тестирования на проникновение – оценить возможность осуществления атаки злоумышленника и спрогнозировать экономические потери в результате ее успешного осуществления.

Виды:

6. Проверка закрытых систем – атакующий не имеет первоначальных сведений об устройстве атакуемой цели. Первоначальная задача такого вида проверки – сбор необходимой информации о расположении целевой системы, её инфраструктуры.
7. Проверка открытых систем – доступна полная информация о целевой системе.
8. Проверка полужакрытых систем – имеется лишь частичная информация.

Результат пентеста – отчёт, содержащий в себе все найденные уязвимости системы безопасности, а также рекомендации по их устранению.

Стандарт исполнения тестирования на проникновение состоит из семи основных этапов:

9. Предварительные взаимодействия. Заключение договора на пентест (статья 272 УК РФ).
10. Сбор разведывательных данных.
11. Моделирование угроз.
- 12) Анализ уязвимости.
13. Эксплуатация.
14. Послеэксплуатационный.
15. Составление отчётов.

Классификация инструментов пентестера:

4. инструменты для сбора информации – поиск уязвимостей;
- 5) инструменты для реверс-инжиниринга – восстановление;
- б) эксплойты – использование найденных уязвимостей;

- 7) инструменты для взлома – проведение атак;
- 8) инструменты для стрес-тестинга – проверка надежности.

Наиболее популярным инструментом для пентеста является бесплатный сборник эксплойтов и вспомогательных скриптов – Metasploit. Для обучения специалистов по пентесту используется виртуальная машина Metasploitable – намеренно уязвимая версия Ubuntu Linux, предназначенная для тестирования средств безопасности и демонстрации распространённых уязвимостей. В ней открыты все порты и присутствуют все известные уязвимости,

Контрольные вопросы и задания

Назовите наиболее достоверный и наиболее используемый способы оценки ИБ.

Что нужно для проведения аудита системы управления информационной безопасностью?

Опишите особенности аудита в сфере ИБ.

Какие курсы и сертификаты есть по пентесту?