

На правах рукописи



Згуральская Екатерина Николаевна

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ПОИСКА СКРЫТЫХ ЗАКОНОМЕРНОСТЕЙ В
БАЗАХ ДАННЫХ ПРИМЕНЕНИЕМ ИНТЕРВАЛЬНЫХ МЕТОДОВ НА ПРИМЕРАХ В
ПРОМЫШЛЕННОСТИ И ДРУГИХ ОБЛАСТЯХ**

Специальность 05.13.01 –Системный анализ, управление и обработка
информации (информационные технологии и промышленность)

Автореферат

диссертации на соискание ученой степени
кандидата технических наук

Ульяновск - 2021

Работа выполнена в федеральном государственном бюджетном образовательном учреждении высшего образования «Ульяновский государственный технический университет»

Научный руководитель: **Крашенинников Виктор Ростиславович**, доктор технических наук, профессор

Официальные оппоненты: **Граничин Олег Николаевич**, доктор физико-математических наук, профессор, профессор кафедры системного программирования математико-механического факультета Санкт-Петербургского государственного университета

Смагин Алексей Аркадьевич, доктор технических наук, профессор, заведующий кафедрой «Телекоммуникационных технологий и сетей» Ульяновского государственного университета

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева»

Защита состоится «29» декабря 2021 года в 15:00 часов на заседании диссертационного совета Д212.277.04 при ФГБОУ ВО «Ульяновский государственный технический университет» по адресу: 432027, г. Ульяновск, ул. Северный Венец, 32 (ауд. 211, Главный корпус).

С диссертацией можно ознакомиться в библиотеке и на сайте ФГБОУ ВО «Ульяновский государственный технический университет». Диссертация и автореферат размещены на сайте <http://www.ulstu.ru/>.

Автореферат разослан «___» _____ 2021 года

Ученый секретарь
диссертационного совета Д212.277.04
д.т.н., доцент



А.М. Наместников

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В диссертации исследуется использование интервальных методов для поиска скрытых закономерностей в базах данных в предметных областях.

Актуальность избранной темы. Использование цифровых технологий для управления в научной, производственной и социальной сферах является одним из главных факторов инновационного развития современного общества. Важную роль для совершенствования цифровых технологий играют информационные модели, основанные на знаниях. Как правило, неявные знания содержатся в базах и хранилищах данных в форме скрытых закономерностей. Отсюда возникает задача выявления этих закономерностей. Идет поиск путей повышения эффективности управления с учетом системных связей и новых знаний о функционировании объектов в предметных областях. Большой вклад в развитие системного анализа, управления и обработки информации внесли российские и зарубежные ученые: Вапник В.Н., Воронцов К. В., Граничин О.Н., Гудфеллоу Я., Дюк В.А., Журавлев Ю.И., Загоруйко Н.Г., Пятецкий-Шапиро Г. и др.

К числу основных проблем построения информационных моделей в слабо структурированных предметных областях является высокая комбинаторная сложность алгоритмов для поиска логических закономерностей.

Важное значение имеет выбор способов предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных, разработка новых методов оценки обобщающей способности алгоритмов распознавания и отбор информативных наборов признаков в описании допустимых объектов.

В рамках метода логической геометрии, разработанного Дюком В.А., предложено поиск логических закономерностей производить в окрестности указанного объекта через снижение размерности исходного пространства с использованием линейного отображения его на плоскость. Такое отображение существенно искажает структуру отношений объектов в исходном пространстве, что сильно ограничивает возможности для интерпретации результатов анализа на плоскости.

Метод опорных векторов SVM, предложенный Вапником В.Н., очень чувствителен к наличию шумовых объектов в обучающей выборке. Жадная стратегия обучения алгоритма метода рассматривает шумовые объекты как граничные, что оказывает существенное влияние на снижение обобщающей способности метода.

Селекция обучающих выборок в работах Загоруйко Н.Г. через поиск минимального покрытия эталонами реализуется путем ручной настройки параметров алгоритма FRIS STOLP. Кроме того, не решена проблема численного решения обнаружения начала переобучения.

Основным инструментарием в диссертации для поиска закономерностей в данных являются интервальные методы. С помощью этих методов упорядоченные значения признака (исходного или латентного) требуется разбить на интервалы так, чтобы каждый интервал содержал как можно больше значений признака объектов одного класса и как можно меньше значений признака объектов других классов. С этим требованием согласуются

два используемых в диссертации критерия качества, в которых оценивается степень однородности, устойчивости значений признака по интервалам. Оптимальность разбиения понимается в смысле экстремума значения критерия, выбор которого определяется спецификой задачи.

Следует отметить, что иногда данные могут иметь пропуски, то есть у некоторых объектов имеются значения не всех признаков, что усложняет задачу поиска закономерностей вообще, и построения границ интервалов, в частности. Кроме того, имеются трудности, связанные с выбором методов преобразования шкал измерений признаков с минимальной потерей информации и селекции обучающих выборок на данных с большой размерностью. Эти вопросы рассматриваются в диссертации.

Поиск оптимального разбиения (то есть границ интервалов) представляет собой сложную задачу, решить которую простым перебором практически невозможно. Поэтому тема данной диссертации является актуальной, так как ее основной задачей является нахождение способов снижения вычислительной сложности поиска оптимального разбиения. Для этого разработан рекурсивный алгоритм, который позволяет вычислять границы интервалов при отсутствии измеренных значений некоторых признаков в описании части объектов, а также способы предобработки данных (численный алгоритм вычисления экстремума критерия разбиения значений признака на непересекающиеся интервалы) и методы отбора информативных наборов признаков по выборке данных в целом и для формирования собственного пространства объекта.

Показано, что применение полученных в диссертации результатов позволяет находить закономерности интервальными методами даже в очень больших объемах данных (с возможными пропусками) при приемлемых вычислительных затратах.

Обнаруженные закономерности можно использовать для решения практических задач системного анализа, управления и обработки информации.

Эффективность разработанных алгоритмов проиллюстрирована примерами обработки данных при обнаружении неисправностей ультразвуковых расходомеров жидкости, классификации изображений, медицинской диагностике сердечно-сосудистых заболеваний и анализе причин, повлиявших на продолжительность срока выживаемости у больных хроническим лимфолейкозом.

Цель. Повышение эффективности поиска скрытых закономерностей по базам и хранилищам данных и многообразиям структур отношений объектов как нового знания из предметных областей за счет применения интервальных методов.

Объект исследования. Базы (хранилища) данных из предметных областей.

Методология и методы диссертационного исследования. В диссертационной работе использованы методы интеллектуального анализа данных, нечеткой логики, дискретной оптимизации.

Задачи:

1. Разработать численный алгоритм разделения значений признаков в описании допустимых объектов классов на непересекающиеся интервалы с использованием предобработки данных при числе интервалов, равном числу

классов. Оценить сложность алгоритмов при использовании и без использования предобработки данных.

2. Разработать способ оценивания устойчивости разбиения значений признаков в границах непересекающихся интервалов для выборки данных из двух классов при числе интервалов, больше либо равном двум. Значение устойчивости является обобщающим показателем доминирования представителей объектов классов по каждому интервалу.

3. Разработать способ отбора информативных наборов признаков по выборке данных в целом и для формирования собственного пространства объекта. Исследовать результаты отбора для принятия решений о наличии и виде неисправностей по данным калибровки ультразвуковых расходомеров жидкости.

4. Разработать рекомендации по выбору правил для распознавания объектов, формируемых с использованием интервальных методов. Исследовать эффективность такого выбора правил на примерах данных по сегментации изображений и данных по медицинской диагностике больных хроническим лимфолейкозом.

Научные новизна. В диссертационной работе впервые получены следующие **результаты:**

1. Разработан численный алгоритм вычисления экстремума критерия качества разбиения значений признака на непересекающиеся интервалы с использованием предобработки данных. Показано, что оценка сложности алгоритма с использованием предобработки значительно ниже, чем у алгоритма без предобработки. Описан способ выбора границ интервалов при условии, что число различных значений признака равно числу классов.

2. Предложен способ отбора информативных наборов разнотипных признаков для описания объектов класса, новизна которого заключается в применении рекурсивного алгоритма для упорядочивания признаков по отношению информативности с использованием предобработки данных путем формирования матрицы близости по парам признаков.

3. Разработаны способы использования интервальных методов в рамках информационных моделей, основанных на знаниях:

а) синтезированы латентные признаки, эффективность принятия решений по которым с точки зрения истинности гипотезы о компактности выше, чем по исходным признакам, используемым для их синтеза;

б) способ отбора информативного набора разнотипных признаков для собственного пространства объекта и значение оценки его по этому набору;

в) способ выбора границ между классами как логических закономерностей в форме полуплоскостей;

г) способ формирования *if...then* правил, отбираемых по значениям устойчивости разбиения признака на непересекающиеся интервалы, для классификации объектов;

д) способ вычисления обобщенных оценок объектов по нелинейным преобразованиям признаков с использованием значений функции принадлежности к классам.

Полученные **результаты соответствуют** следующим пунктам **паспорта специальности 05.13.01. – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)»**, а именно:

п.5 - разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации (Результаты 1 и 2 Заключения);

п.13 - методы получения, анализа и обработки экспертной информации (Результат 3 Заключения).

Теоретическая значимость диссертационной работы заключается в разработке новых методов предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных.

Практическая значимость. Обнаружение скрытых закономерностей с помощью интервальных методов позволяет повысить обобщающую способность алгоритмов распознавания и обосновывать процесс принятия решений в технических и других системах.

Степень достоверности полученных результатов обосновывается корректным применением математического аппарата, подтверждается вычислительными экспериментами и результатами практического использования.

Основные положения, выносимые на защиту.

– Разработанный численный алгоритм для разбиения значений признаков в описании объектов классов на непересекающиеся интервалы с применением предобработки данных требует значительно меньших вычислительных ресурсов, чем алгоритм без предобработки.

– Показано, что поиск оптимальной эвристики для отбора информативных наборов разнотипных признаков целесообразно проводить на основе результатов минимального покрытия обучающей выборки объектами-эталоны. Для оптимальной эвристики среднее число объектов выборки, описываемое информативным набором и притягиваемое одним эталоном минимального покрытия, имеет максимальное значение и лучшую обобщающую способность при распознавании по алгоритму «ближайший сосед».

– Разработанный рекурсивный алгоритм позволяет вычислять границы интервалов и их число при частичном отсутствии измеренных значений признаков в описании части объектов классов.

– Синтез латентных признаков по операциям умножения и деления значений исходных признаков позволяет увеличить внутриклассовое сходство и межклассовое различие в процессе принятия решений.

Апробация результатов. Основные положения диссертации докладывались на конференциях: V Международная конференция и молодежная школа «*Информационные технологии и нанотехнологии*» (г. Самара, 2019 г.), VI Международная конференция и молодежная школа «*Информационные технологии и нанотехнологии*» (г. Самара, 2020 г.), XI Всероссийская научно-практическая конференция «*Современные проблемы проектирования, производства и эксплуатации радиотехнических систем*» (г. Ульяновск, 2019 г.), I Всероссийская научно-техническая конференция

«Теоретические и практические аспекты развития отечественного авиастроения» (г. Ульяновск, 2012 г.), V Всероссийская научно-техническая конференция *«Теоретические и практические аспекты развития отечественного авиастроения»* (г. Ульяновск, 2018 г.), Международная конференция *«Инфокоммуникационные и вычислительные технологии в науке, технике и образовании»* (г. Ташкент, 2004 г.).

Публикации. По теме диссертационной работы опубликовано 16 печатных работ, из них 4 в изданиях из перечня ВАК, 2 Scopus, 1 патент на изобретение.

Внедрение результатов. Результаты диссертационной работы внедрены в гематологическом отделении Государственного учреждения здравоохранения «Ульяновская областная клиническая больница».

Сведения о личном вкладе автора. Постановка задач исследования осуществлялась совместно с научным руководителем. Все основные теоретические и практические исследования проведены автором диссертационной работы самостоятельно. Подготовка к публикации некоторых результатов проводилась совместно с соавторами, вклад соискателя был определяющим.

Структура диссертации. Диссертация состоит из введения, трех глав, заключения, списка литературы и приложения. Работа изложена на 124 страницах, включающих 12 рисунков и 31 таблицу. Список использованной литературы включает в себя 101 наименование.

Содержание работы

Во введении обоснована актуальность работы, сформулирована цель и задачи диссертационного исследования, изложена его научная новизна, раскрыты теоретическое значение и практическая ценность полученных результатов, кратко излагается содержание диссертационной работы.

В первой главе дается формулировка основной решаемой в диссертации задачи, обзор литературных источников в области выявления скрытых закономерностей и применения интервальных методов, описание и обоснование вычислительных алгоритмов для поиска оптимального разбиения значений признаков на непересекающиеся интервалы по двум критериям. В качестве примера приведен вычислительный эксперимент на данных по сегментации изображений.

Постановка задачи. Пусть задано множество объектов $E_0=\{S_1, \dots, S_m\}$, содержащее представителей l непересекающихся классов K_1, \dots, K_l . Описание объектов производится с помощью набора из n разнотипных признаков $X(n)=(x_1, \dots, x_n)$, δ из которых измеряются в номинальной, $n-\delta$ в интервальной шкалах ($\delta < n$). Допускается наличие пропусков и повторяющихся значений в данных. Считается, что задан критерий $F(*)$ для разбиения значений каждого количественного признака (как исходного, так и латентного) на непересекающиеся интервалы. Латентные признаки могут представлять собой комбинации из номинальных и количественных признаков. Требуется определить значения границ l интервалов, при которых $F(*)=extr$. Совокупность таких границ будем называть оптимальным решением задачи.

Оптимальное решение этой задачи также является ответом на следующие вопросы.

1. Насколько компактно расположены значения признаков в описании объектов классов на числовой оси?
2. Где проходят границы между значениями признаков объектов классов?

В данной постановке задачи следует отметить два важных момента. Во-первых, использование интервального метода, в основе которого лежит предположение, что у схожих объектов близки значения их признаков. Во-вторых, оптимальность решения задачи понимается в смысле назначенного или выбираемого критерия близости получающегося разбиения на классы к заданному разбиению, то есть результат существенно зависит от используемого критерия F^* . В диссертации используются два критерия, описание и применение которых дается ниже.

Критерий для разбиения значений признаков на число интервалов, равное числу непересекающихся классов.

Рассмотрим критерий

$$\left(\frac{\sum_{p=1}^l \sum_{i=1}^l (u_i^p - 1) u_i^p}{\sum_{i=1}^l |K_i| (|K_i| - 1)} \right) \left(\frac{\sum_{p=1}^l \sum_{i=1}^l u_i^p \left(m - |K_i| - \sum_{j=1}^l u_j^p + u_i^p \right)}{\sum_{i=1}^l |K_i| (m - |K_i|)} \right) \rightarrow \max_{\{A\}} \quad (1)$$

где l – множество индексов количественных признаков, u_i^p – множество измеренных значений признака x_j , $j \in I$ класса K_i в интервале $(c_{2p-1}; c_{2p}]$, $p=1, \dots, l$, $A=(a_0, \dots, a_l)$, $a_0=0$, $a_l=m$, a_p – порядковый номер элемента в упорядоченной по возрастанию последовательности r_1, \dots, r_m значений x_j из E_0 , определяющий правую границу интервала $c_{2p} = r_{a_p}$.

Выражение в первых скобках этого критерия является мерой внутриклассового сходства, а во вторых скобках – мерой межклассового различия. Таким образом, значение этого критерия для заданного разбиения объектов на классы является показателем того, насколько схожи объекты, находящиеся в одном классе, и насколько различны объекты, отнесенные к разным классам. Значение (1), равное 1, интерпретируется как идеальное разбиение, при котором в границах каждого интервала содержатся значения признака всех объектов одного класса.

Оптимальное значение критерия (1) будем считать мерой компактности признака на числовой оси в задачах с пересекающимися классами.

Критерий, в котором число интервалов изначально не задано и выборка объектов представлена только двумя классами. Значения признака объектов одного класса могут быть расположены на числовой оси двумя и более группами, между которыми находятся значения этого же признака другого класса. В этом случае целесообразно при интервальном методе каждому классу поставить в соответствие не один, а большее количество интервалов значений признака.

Произведем разбиение на непересекающиеся интервалы для каждого количественного признака, в границах которых доминируют значения объектов класса K_t или K_{3-t} , $t=1,2$. Отношение «доминирование» используется для вычисления устойчивости разбиения на интервалы по каждому признаку. Устойчивость – это численная характеристика разбиения значений признака на

непересекающиеся интервалы. Для поиска закономерностей значения c -го (латентного или исходного) признака упорядочим по возрастанию:

$$r_{c_1}, r_{c_2}, \dots, r_{c_m}. \quad (2)$$

Согласно определяемому ниже критерию, последовательность (2) разбивается на τ_c ($\tau_c \geq 2$) непересекающихся интервалов $[r_{c_u}; r_{c_v}]^i$, $1 \leq u, u \leq v, v \leq m$, $i=1, \dots, \tau_c$. Значения, лежащие в интервале $[r_{c_u}; r_{c_v}]^i$, далее могут рассматриваться как градация номинального признака.

Обозначим через $d_t^i(u, v)$, $d_{3-t}^i(u, v)$ количество представителей соответственно классов K_t , K_{3-t} в интервале $[r_{c_u}; r_{c_v}]^i$. Для рекурсивной процедуры поиска значений r_{c_u}, r_{c_v} используется критерий

$$\left| \frac{d_t^i(u, v)}{|E_0 \cup K_t|} - \frac{d_{3-t}^i(u, v)}{|E_0 \cup K_{3-t}|} \right| \rightarrow \max. \quad (3)$$

Устойчивость признака по множеству интервалов разбиения.

Определим значение функции принадлежности c -го признака к K_1 по интервалу $[r_{c_u}; r_{c_v}]^i$ как

$$f_1(i) = \frac{\eta_{1i}}{\eta_{1i} + \eta_{2i}}, \quad (4)$$

где $\eta_{1i}(t) = \frac{d_t^i(u, v)}{|E_0 \cup K_t|}$, $\eta_{2i}(t) = \frac{d_{3-t}^i(u, v)}{|E_0 \cup K_{3-t}|}$.

Так как $f_t(i) = 1 - f_{3-t}(i)$, $t=1, 2$, то устойчивость признака по множеству интервалов разбиения определим, как

$$U(c) = \frac{1}{m} \sum_{\{[r_u, r_v]^i\}} \begin{cases} f_t(i)(v - u + 1), & f_t(i) > 0.5, \\ (1 - f_t(i))(v - u + 1), & f_t(i) < 0.5. \end{cases} \quad (5)$$

Модифицированный критерий (3) для случая наличия пропусков в данных. Как уже отмечалось, иногда в практических приложениях у части объектов могут быть пропуски в их признаках. С учетом таких пропусков критерий (3) принимает вид

$$\left| \frac{d_t^i(u, v)}{T_p^c} - \frac{d_{3-t}^i(u, v)}{T_{3-p}^c} \right| \rightarrow \max, \quad (6)$$

где T_p^c, T_{3-p}^c – количество значений признака $x_c \in X(n)$ без пропусков у объектов E_0 соответственно из классов K_p и K_{3-p} . Условиями для реализации (6) являются:

- число различных значений признака больше или равно 2;
- значения $T_p^c > 0, T_{3-p}^c > 0$.

С учетом пропусков в данных значением устойчивости будет

$$U(c) = \frac{1}{\mu} \sum_{\{[r_u, r_v]^i\}} \begin{cases} f_t(i)(v - u + 1), & f_t(i) > 0.5, \\ (1 - f_t(i))(v - u + 1), & f_t(i) < 0.5, \end{cases} \quad (7)$$

где $\mu = T_p^c + T_{3-p}^c$.

Показатели устойчивости (5) и (7) имеют более приоритетное значение для обнаружения скрытых закономерностей, чем число непересекающихся интервалов, полученных по критериям (3) и (6). Это утверждение продемонстрировано на выборке объектов по сегментации изображений [http://archive.ics.uci.edu/ml/datasets/Statlog+%28Image+Segmentation%](http://archive.ics.uci.edu/ml/datasets/Statlog+%28Image+Segmentation%29).

Выборка состоит из 2300 объектов (изображений, сделанных на улице) описанных 19 признаками. Изображения этой выборки были вручную сегментированы (разбиты) экспертами на 7 классов (кирпич, небо, листва, цемент, окно, дорога, трава.). Случайным образом на этой выборке были сгенерированы пропуски. Значения устойчивости разбиения на данных без пропусков и содержащих пропуски до 30% различались не более чем на 5%, максимальная разница между числом непересекающихся интервалов равнялась 3, т.е. на выборке из генеральной совокупности значение устойчивости стремится к ее постоянной величине вне зависимости от числа разбиений по (3).

Разбиения на интервалы по (3) и (6) дают возможность наглядного представления знаний в виде дизъюнкций элементарных конъюнкций. Элементарные конъюнкции нужны для проверки принадлежности значения признака к одному из интервалов. Запись правила для отнесения объекта классу K_t , $t=1,2$, может иметь следующий вид: $a_1 \leq x_i \leq b_1$ or $a_2 \leq x_i \leq b_2$ or ... or $a_{\eta-1} \leq x_i \leq b_{\eta-1}$, где $a_j, b_j, j \in \{1, \eta\}$ – границы интервалов, η – число непересекающихся интервалов.

Во **второй главе** разрабатываются способы снижения сложности алгоритмов вычисления границ интервалов и реализация отбора информативных признаков на их основе за счет предобработки данных и иерархической агломеративной группировки признаков.

Сложность критерия разбиения значений признаков на число интервалов, равное числу непересекающихся классов.

Для уменьшения комбинаторной сложности реализации алгоритмов оптимизации по критерию (1) в диссертации предложено проводить предобработку данных. Оценка сложности представляет произведение количества всевозможных разбиений значений признака на непересекающиеся интервалы и операций, необходимых для определения значений u_p^t . Количество непересекающихся интервалов в (1) определяется как

$$\psi(l, m) = l C_{m-1}^{l-1}. \quad (8)$$

Оценка сложности без учета предобработки вычисляется как

$$F(l, m) = \left(2m \binom{1+l}{2} + m \right) \psi(l, m) = m(2+l)\psi(l, m), \quad (9)$$

в которой значение $m(2+l)$ представляет собой количество элементарных операций для подсчета частоты встречаемости представителей класса в интервале.

В процессе предобработки по упорядоченной последовательности значений признака r_1, \dots, r_m , сформируем целочисленную матрицу вида: $D=(d_{pi})$, в которой индекс столбца элемента d_{pi} , $p=1, \dots, l$, $i=1, \dots, m$, соответствует объекту $S \in E_0$ со значением признака r_i . Элементы матрицы вычислим как

$$d_{pi} = \begin{cases} 0, & i = 0, \\ d_{p,i-1} + g(p, i), & i > 0, \end{cases} \quad g(p, i) = \begin{cases} 1, & S \in K_p, \\ 0, & S \notin K_p. \end{cases}$$

Число представителей u_t^p класса $K_p, p=1, \dots, l, t=1, \dots, l$, в интервале $[c_1, c_2]$ при $t=1$ и $(c_{2t-1}, c_{2t}]$ при $t>1$ определим как

$$u_t^p = d_{pv} - d_{p\eta}, \quad (10)$$

где $\eta=a_{t-1}, v=a_t, c_{2t-1}=r_{j\eta}, c_{2t}=r_{jv}$. Предобработка данных по (10) дает сложность

$$R(l, m) = l\psi(l, m) + 2m. \quad (11)$$

Сложность алгоритма с использованием предобработки значительно ниже, чем у алгоритма без использования предобработки. Это позволяет вычислять значение критерия (1) за полиномиальное время.

К числу особенностей реализации алгоритмов поиска оптимального значения критерия (1) можно отнести случай, когда число отличных друг от друга значений равно числу классов и не существует двух классов, объекты которых имеют одно значение признака.

Пусть a_1, a_2, \dots, a_l – упорядоченные в порядке возрастания значения признака для задачи с K_1, \dots, K_l непересекающимися классами объектов. Значения $l-1$ границы $[a_1; c_1] (c_1; a_2) \dots (c_{l-1}; a_l)$ определяются как $c_i = (a_i + a_{i+1})/2, i=1, \dots, l-1$. Сложность алгоритма по (11) для (1) будет минимальной. Для ее вычисления необходимо будет подсчитать число объектов классов K_1, \dots, K_l , принимающих значения a_1, a_2, \dots, a_l .

Скрытые закономерности на многообразии отношений между объектами. Исследуются структуры отношений между описаниями объектов классов на числовой оси. В качестве инструмента для исследования отношений между описаниями объектов классов предлагаются меры компактности, вычисляемые по результатам разбиения значений признаков (исходных и латентных) на непересекающиеся интервалы.

Пусть на наборе признаков $X(h) \subset X(n), 1 \leq h \leq n$ в описании объектов классов определена метрика $\rho(x, y)$. Объект $S \in K_p, p=1, 2$ рассматривается как центр, от которого формируется последовательность вложенных друг в друга гипершаров с радиусами

$$\rho(S, S) < \rho(S, S^1) \leq \rho(S, S^2) \leq \dots \leq \rho(S, S^{m-1}), \quad (12)$$

где $\{S, S^1, \dots, S^{m-1}\} = E_0$. Значения границ интервалов каждого объекта $S \in K_p, p=1, 2$ на E_0 , вычисляемые по критерию (1) на (12), используется для отбора информативного набора разнотипных признаков из $X(n)$.

Пусть по (1) на (12) определены границы интервалов $[c_1; c_2], (c_2; c_3], c_1 = \rho(S, S) = 0$. Оценку компактности объекта $S \in K_p$ по набору признаков $X(h) \subset X(n)$ вычислим как

$$\phi(S, X(h)) = \vartheta_1(1 - \vartheta_2), \quad (13)$$

где

$$\vartheta_1 = \frac{|\{S^i \in K_p \mid \rho(S, S^i) \in [c_1, c_2]\}|}{|K_p|}, \quad \vartheta_2 = \frac{|\{S^i \in K_{3-p} \mid \rho(S, S^i) \in [c_1, c_2]\}|}{|K_{3-p}|}.$$

Другой эвристикой для отбора информативных признаков была взята

$$\varphi(S, X(h)) = \max_{1 \leq i \leq m-1} \left(\frac{d_t(i)}{|K_t|} - \frac{d_{3-t}(i)}{|K_{3-t}|} \right), \quad (14)$$

где $d_t(i) (d_{3-t}(i))$ – частота встречаемости радиусов объектов класса $K_t (K_{3-t})$ в подпоследовательности $\rho(S, S) < \rho(S, S^1) \leq \rho(S, S^2) \leq \dots \leq \rho(S, S^i), i < m$ из (12).

Пусть $\varphi(X(h), S_i)$ – оценка компактности объекта $S_i \in E_0$ на $X(h)$. При формировании набора $X(h+1)$ из $X(h)$ необходимо вычислить

$$R(X(h+1)) = \frac{1}{m} \sum_{i=1}^m \begin{cases} 1, \varphi(X(h+1), S_i) \geq \varphi(X(h), S_i), \\ 0, \varphi(X(h+1), S_i) < \varphi(X(h), S_i). \end{cases} \quad (15)$$

Условием (правилом) иерархической агломеративной группировки для включения признака $x_j \in X(n) \setminus X(h)$ в $X(h+1)$ является:

$$R(X(h) \cup \{x_j\}) > \frac{1}{2} \quad \text{и} \quad R(X(h) \cup \{x_j\}) = \max_{x_j \in X(n) \setminus X(h)} R(X(h) \cup \{x_j\}) \quad (16)$$

Группа (набор) $X(h)$ считается сформированной, если не существует признака $x_j \in X(n) \setminus X(h)$, для которого выполняется (16).

В качестве первого шага при формировании информативного набора признаков предлагается выбирать подмножество $Y \subset X(n)$, состоящее из одного или двух признаков. Подмножество Y должно удовлетворять следующему требованию: при выборе первой пары признаков количество объектов в гипершаре одного класса с его центром в среднем по всей выборке должно быть максимальным.

Оценка обобщающей способности алгоритмов. Для того, чтобы сравнивать наборы признаков, полученные разными способами вычисления оценок компактности $\varphi(S, X(h))$, предлагается использовать значения обобщающей способности алгоритма «ближайший сосед», вычисляемые как

$$\Omega(E_0, \rho) = \frac{|E_{ob}|}{|E_{ed}|}, \quad (17)$$

где E_{ob} – множество объектов обучения без шумовых объектов, $|E_{ed}|$ – число объектов-эталонов минимального покрытия E_{ob} . Шумовые объекты определяются как подмножество граничных объектов классов. Объекты классов разбиваются на непересекающиеся группы, по каждой из которой производится поиск объектов-эталонов минимального покрытия. Разбиение объектов на группы рассматривается как предобработка данных, использование которой позволяет отказаться от перебора всех комбинаций объектов как вариантов минимального покрытия.

В **третьей главе** рассматриваются два подхода к формированию признакового пространства для описания объектов классов, целями реализации которых являются:

- селекция обучающих выборок для повышения обобщающей способности алгоритмов распознавания по прецедентам;
- принятие решения по объекту в рамках его собственного пространства.

Применение интервальных методов связано с:

- упорядочением исходного набора признаков $X(n)$ для селекции обучающих выборок;
- оценкой компактности объектов класса по системе вложенных гипершаров при поиске собственного пространства объекта.

Селекция обучающих выборок через отбор информативных разнотипных признаков и минимальное покрытие объектами-эталонами. Под оценкой (мерой) сложности $R(X(k))$, $k \leq n$ алгоритма принятия решения по

допустимому объекту S будем понимать количество элементарных операций для его распознавания по набору признаков $(X(k))$. Рассматривается следующая задача построения упорядоченной по убыванию сложности алгоритмов последовательности признаков.

Пусть задано множество объектов $E_0 = \{S_1, \dots, S_m\}$, содержащее представителей l непересекающихся классов (подмножеств E_0) K_1, \dots, K_l . Описание объектов произведено с помощью набора из n разнотипных признаков $X(n) = (x_1, \dots, x_n)$, δ из которых измеряются в номинальной шкале, $n - \delta$ в количественных шкалах, допускается отсутствие измеренных значений в данных. Требуется найти последовательность информативных наборов $X(n), X(n-1), \dots, X(k)$, $k \leq n$ на которых значения меры сложности алгоритмов $R(X(n)), R(X(n-1)), \dots, R(X(k))$ образуют невозрастающую последовательность $R(X(n)) \geq R(X(n-1)) \geq \dots \geq R(X(k))$.

Отбор информативных признаков с максимально выраженной независимостью.

Теоретически (в идеале) минимальный набор должны представлять независимые признаки. В практической реализации выбираются исходные (не искусственные) признаки, коррелированность между собой которых минимальна, т.е. признаки с максимально выраженной независимостью. На множестве пар объектов $(S_a = (x_{a1}, \dots, x_{an}), S_b = (x_{b1}, \dots, x_{bn}))$ определим функции:

$$g(a, b, i, j) = \begin{cases} 2, x_{ai} \neq x_{bi} \text{ и } x_{aj} \neq x_{bj}; \\ 1, x_{ai} = x_{bi} \text{ или } x_{aj} = x_{bj}; \\ 0, x_{ai} = x_{bi} \text{ и } x_{aj} = x_{bj}; \end{cases} \quad \alpha(a, b) = \begin{cases} 0, S_a, S_b \in K_i, i = \overline{1, l}; \\ 1, S_a \in K_i, S_b \in K_j, i \neq j. \end{cases}$$

Меру близости между парой номинальных признаков x_i, x_j на E_0 зададим как

$$b_{ij} = \begin{cases} \frac{\sum_{a=1}^m \sum_{b=1}^m \alpha(a, b) g(a, b, i, j)}{2 \sum_{p=1}^l |K_p| (m - |K_p|)}, i \neq j; \\ 0, i = j. \end{cases} \quad (18)$$

Критерием для отбора информативного набора признаков $X(k) = (x_1, \dots, x_k)$, $k \leq n$ служит оценка сложности алгоритма

$$\min_{\{\Pi_j\}} |\Pi_j| k \rightarrow \min_{E_0}, \quad (19)$$

где k – число признаков в наборе, Π_j – множество объектов–эталонов минимального покрытия, используемого в (17).

Для определения информативного набора $X(k) = (x_1, \dots, x_k)$, $k < n$ по (19) предлагается рекурсивная процедура построения упорядоченной по отношению информативности последовательности признаков

$$X_{i_1}, X_{i_2}, \dots, X_{i_n} \quad (20)$$

применялся рекурсивный алгоритм к результатам предобработки данных по (1) и (18).

Формирование собственного пространства объекта. Пусть для объекта $S_d \in K_p$, $p=1, 2$ расстояния до объектов выборки E_0 представлены в виде упорядоченной последовательности (12). Обозначим через P множество

индексов признаков из $X(n)$. Реализация алгоритма пошагового отбора информативных признаков будет следующей.

Шаг 1. $P=\{1, \dots, n\}$. $Y=\{x_{ij}\}_{i \in P}$. Вычислить $\text{crit}=\phi(S_d, Y)$ по (13). $Z=Y$;

Шаг 2. **Цикл** по $i, j \in P$. $b_{ij}=\phi(S_d, Y \setminus \{x_i, x_j\})$. **Конец цикла**;

Шаг 3. $A=0$. **Цикл** по $i, j \in P$. Если $A < b_{ij}$ то $i_1=i, j_1=j, A=b_{ij}$. **Конец цикла**;

Шаг 4. $Y=Y \setminus \{x_{i_1}, x_{j_1}\}$. $P=P \setminus \{i_1, j_1\}$. Если $A > \text{crit}$, то $\text{crit}=A, Z=Y$;

Шаг 5. Если $|P| > 2$, то идти 2;

Шаг 6. Вывод Z ;

Шаг 6. **Конец**.

Обобщенные оценки объектов. Для сокращения размерности признакового пространства использовались латентные признаки, вычисляемые как обобщенные оценки объектов по формуле $Z(S_r) = \sum_{i=1}^n \eta_i(x_{ri})$, где $S_r=(x_{r1}, \dots, x_{rm})$, $\eta_i(x_{ri})$ – вклад признака $x_{ri} \in \{1, 2\}$, значения в $\{1, 2\}$ получены через вычисление функции принадлежности (4) и устойчивости (5).

Эффективность преобразования признаков в $\{1, 2\}$ для сокращения размерности пространства с использованием функций принадлежности (4) и устойчивости (5) доказана через вычисление обобщенных оценок объектов. В ходе вычислительного эксперимента было показано, что компактность объектов классов по обобщенным оценкам была выше на наборах признаков, для которых применялось преобразование в $\{1, 2\}$.

Задача по диагностике неисправностей расходомеров. Расходомеры – это устройства, используемые для измерения объемного или массового расхода жидкости. Существует много причин, приводящих к ошибочным показаниям устройства, для исправления которых необходима повторная калибровка. Задача правильной диагностики состояния расходомера сводится к задаче классификации. Выборка данных состоит из 4-х классов: K_1 (исправный), K_2 (впрыск газа), K_3 (дефекты установки), K_4 (восковая депиляция). В базе данных имеется 180 объектов, описываемых 43 признаками (<http://archive.ics.uci.edu/ml/datasets/Ultrasonic+flowmeter+diagnostics#>).

Необходимо:

- выбрать минимальный набор признаков, при котором точность классификации была бы выше, чем по набору исходных признаков;
- определить набор признаков, значения которых являются причиной неисправности конкретного устройства.

Для решения этой задачи были применены разработанные в диссертации способы выявления скрытых закономерностей и проведены численные эксперименты.

Во второй главе в ходе вычислительных экспериментов информативные наборы признаков были получены по следующим эвристикам:

1. отбор в информативный набор производился из упорядоченной по весам (1) последовательности;
2. использовалось правило иерархической агломеративной группировки для пошагового объединения признаков в информативный набор с использованием мер компактности объекта по (13) и (14).

В качестве меры расстояния между описаниями объектов использовалась метрика Журавлева. Значения количественных признаков были

пронормированы в $[0;1]$. Компактность по (17) и 25 информативных признаках по первому варианту была 7.6055, по второму варианту на 21 признаке была выше и равнялась 17.3841. Значение меры компактности по (17) служит основанием для выбора классификации неисправностей по второму варианту.

В третьей главе по сравнению с главой 2 при планировании эксперимента вносятся следующие элементы анализа:

1. формирование последовательности вида (20), инвариантной к масштабам измерений признаков;
2. на данных не используется нормирование;
3. визуально отслеживается структура отношений объектов по значениям сырых признаков на разных наборах признаков по методу t-SNE;
4. показаны особенности реализации метода минимального покрытия выборки объектами-эталоном при использовании одной метрики для всех объектов и локальной метрики для каждого.

Оказалось, что компактность выборки данных по (17) больше, чем при их нормировании. На 43 признаках это соответственно значения 9.1683 и 7.8948.

Наборы для эксперимента формировались путем отбрасывания (справа–налево) признаков из упорядоченной по отношению информативности последовательности признаков вида (20). Так как обобщающая способность алгоритмов растет с увеличением (17), то рекомендуется для классификации неисправностей использовать набор из 33 признаков со значением меры компактности, равной 10.4116.

Для визуализации результатов вычислительного эксперимента на разных наборах признаков использовался алгоритм нелинейного снижения размерности t-SNE. Визуализация приводилась относительно структуры отношений всех объектов выборки в пространстве $X(k)$, $k \leq n$ и объектов–эталонных минимального покрытия по их локальной метрике. Локальная метрика объекта $S_k \in K_t$, $t=1, \dots, 4$, по формуле $\rho^*(S, S_k) = \rho(S, S_k)/b$, где b – минимальное расстояние от S_k до объекта из дополнения CK_t к классу K_t .

В ходе вычислительного эксперимента была показана эффективность выбора эталонных объектов с локальной метрикой, данный результат также подтверждается при использовании метода t-SNE.

Выбор собственного пространства объекта позволяет определить набор признаков, значения которых являются причиной неисправности конкретного устройства. Интерес для исследования представляет анализ частоты встречаемости отдельных признаков в составе собственного пространства объектов из класса с показателями дефектов установки. Наиболее часто встречаемыми показателями неисправностей в работе ультразвукового расходомера являются значения признака «скорость звука на путях».

Задача анализа причин, влияющих на общую выживаемость больных хроническим лимфолейкозом. В качестве материала для исследования использовались данные 123 пациентов, описываемых 11 признаками с хроническим лимфолейкозом А-С стадии по Binet в возрасте от 47 до 82 лет с известными значениями общей выживаемости, полученные в гематологическом отделении Ульяновской областной клинической больницы.

Выборка данных больных мужского пола состояла из 60 объектов, женского пола из 63 объектов. Объекты каждой из выборок были разделены на два класса K_1 (фактическая выживаемость меньше прогнозируемой общей выживаемости) и K_2 (фактическая выживаемость больше или равна прогнозируемой общей выживаемости).

Необходимо:

Определить причины (медико-биологические показатели), которые влияют на продолжительность жизни у людей хроническим лимфолейкозом, с целью принятия решений при выборе тактики лечения пациентов.

Разработана методика поиска скрытых закономерностей методами интеллектуального анализа данных для больных хроническим лимфолейкозом. При анализе использовались нелинейные преобразования признаков на основе значений функции принадлежности объектов к классам. Описана и обоснована последовательность преобразований признаков от исходного представления до значений в номинальной шкале измерений для вычисления обобщенных оценок объектов. Найдены логические закономерности в форме полуплоскостей, пороговые значения для которых определены как по отдельным признакам, так и обобщенным оценкам объектов. Наиболее практически значимые результаты при анализе сроков фактической выживаемости были получены с учетом гендерной принадлежности пациентов.

В заключении приведены основные результаты исследований, представленные в диссертации.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

В результате проведенных исследований получено несколько способов снижения вычислительных затрат при выявлении скрытых закономерностей в значениях признаков заданного множества объектов. Получаемый эффект от применения этих способов позволяет считать, что поставленная цель диссертации достигнута. Основными результатами диссертации являются следующие.

1. Разработан численный алгоритм разделения значений признаков в описании допустимых объектов классов на непересекающиеся интервалы с использованием предобработки данных при числе интервалов, равном числу классов. Эффект от применения предобработки заключается в сокращении вычислительных затрат, что дало возможность расширения объемов анализируемых данных как по числу признаков, так и по числу объектов. На тестовом примере из 16 объектов показано уменьшение оценки сложности вычислений с 29520 до 1140 соответственно без использования и при использовании предобработки. По экстремальному значению критерия на 43 признаках показаний калибровки 4-х канального ультразвукового расходомера для диагностики 3 видов неисправностей определено, что максимальный вклад в принятие решения о неисправностях дают 2 признака: коэффициенты усиления в начале и конце 4-го пути. Коэффициенты имели одинаковое значение критерия (компактности), равное 0.6464.

2. Получена оценка устойчивости разбиения значений признаков в границах непересекающихся интервалов для выборки данных из двух классов

при числе интервалов больше либо равном двум. Множество допустимых значений оценки принадлежит $(0.5;1]$. Такая оценка является более важным показателем для поиска скрытых закономерностей, чем число интервалов и значения их границ. Устойчивость равна 1, если в каждом интервале представлены объекты только одного из классов. Показано, что устойчивость разбиения признаков на интервалы как при наличии до 30% пропусков в данных, так и при их отсутствии на выборке данных по сегментации изображений отличалась разбросом значений не более 5%.

3. Разработан способ отбора информативных наборов разнотипных признаков по выборке данных с использованием нескольких эвристик. Для сравнения эвристик применялся поиск минимального покрытия выборки объектами-эталоном. Лучшим считался набор признаков, при использовании которого среднее число объектов выборки, притягиваемых одним объектом-эталоном покрытия, было максимальным. Реализован алгоритм отбора информативных признаков объекта по максимальному значению произведения внутриклассового сходства и межклассового различия. Например, для данных по диагностике неисправностей 8-ми лучевого ультразвукового расходомера значение произведения в $(0;1]$ интерпретировалось как индекс тяжести неисправности по индивидуальному набору показателей. Выбор от 2 до 6 информативных показателей и оценка по ним степени тяжести неисправности устройств сокращают время диагностики, материальные расходы на ремонт и ущерб от неверных значений показателей при измерении.

4. Разработаны следующие рекомендации по выбору правил для распознавания объектов с использованием интервальных методов.

– Признаки, значение устойчивости разбиения которых лежит в $[0.9; 1]$ при числе интервалов не больше 4, предложено использовать для формирования *if...then* правил для распознавания. Например, для данных по сегментации изображений было отобрано 19 признаков, в границах разбиения, на интервалы которых сформированы *if...then* правила.

– При вычислении порога логических закономерностей в форме полуплоскостей, при числе интервалов и классов, равным двум, рекомендовано использовать значения границ интервалов.

– Для прогнозирования сроков выживаемости у больных хроническим лимфолейкозом мужского пола рекомендовано использовать обобщенные оценки объектов по унифицированным значениям пяти медико-биологических показателей. Унификация в $\{1,2\}$ проводилась с помощью значений функции принадлежности объектов к классам, вычисляемых в границах интервалов, определяемых по (4).

Перспективы дальнейшей разработки темы:

- В диссертации для оценивания качества разбиения на интервалы использовано два критерия конкретного вида. В дальнейшем предполагается рассмотреть использование других критериев качества при применении и модернизации интервальных методов с целью извлечения скрытых закономерностей в базах данных.

- В дальнейших исследованиях интервальные методы будут использоваться при разработке интеллектуальных встроенных систем, в которых совмещены процессы «добычи знаний» и принятия управленческих решений. Например, при сертификации датчиков авиационной техники.

ОСНОВНЫЕ ПОЛОЖЕНИЯ ДИССЕРТАЦИОННОЙ РАБОТЫ ИЗЛОЖЕНЫ В СЛЕДУЮЩИХ РАБОТАХ

Статьи, опубликованные в журналах, рекомендованных ВАК РФ:

1. Згуральская Е.Н. Алгоритм выбора оптимальных границ интервалов разбиения значений признаков при классификации / Е.Н.Згуральская // Известия Самарского науч. центра Рос. акад. наук. – 2012. – Т. 14, №4(3). – С. 826-829.
2. Згуральская Е.Н. Выбор информативных признаков для решения задач классификации с помощью искусственных нейронных сетей / Е.Н.Згуральская // Нейрокомпьютеры: разработка, применение. – 2012. – № 2. – С. 20-27.
3. Згуральская Е.Н. Устойчивость разбиения данных на интервалы в задачах распознавания и поиск скрытых закономерностей / Е.Н.Згуральская // Известия Самарского науч. центра Рос. акад. наук. – 2018. – Т. 20, № 4(3). – С. 451-455.
4. Згуральская Е.Н. Поиск скрытых закономерностей, влияющих на общую выживаемость больных, методами интеллектуального анализа данных / Н.А.Игнатъев, Е.Н.Згуральская, М.В.Марковцева // Искусственный интеллект и принятие решений. – 2020. – №3. – С. 73-80.

Статьи в зарубежных рецензируемых журналах:

5. Zguralskaya E.N. Analysis of the structure of the relationship between the descriptions of objects of classes and evaluation of their compactness. / E.N.Zguralskaya // Workshop Proceedings Information Technology and Nanotechnology (ITNT-2019). – 2019. – Pp. 283-289.
6. Zguralskaya E.N. Nonlinear transformation of signs and the search for patterns in the data of patients with chronic lymphocytic leukemia. / Ignatiev N.A., Zguralskaya E.N., Markovtseva M.V. // Workshop Proceedings Information Technology and Nanotechnology (ITNT-2020). – 2020. – Pp. 333-336.

Публикации в других изданиях:

7. Игнатъева Е.Н. (Згуральская Е.Н.) Изучение информативности параметров неинвазивных методов диагностики в качестве маркеров старения у больных ишемической болезнью сердца / Г.А.Розыходжаева, Е.Н.Игнатъева // Врач и информационные технологии. – 2006. – №1 – С. 38-44.
8. Згуральская Е.Н. Оценка степени тяжести хронической сердечной недостаточности с позиции биосимметрии / Ф.Т.Адылова, П.П.Черныш, Е.Н.Згуральская // Укр. журн. телемедицины и медицинской информатики. – 2008. – Т.6, №1. – С. 42-47.
9. Згуральская Е.Н. Поиск закономерностей по значениям количественных признаков с помощью детерминистических критериев разбиения на интервалы. / Е.Н Згуральская. // В сборнике: Междисциплинарные исследования в области математического моделирования и информатики труды 3-й научно-практической конференции. – 2014. – С. 199-203.
10. Згуральская Е.Н. Поиск скрытых закономерностей в форме полуплоскостей интервальными методами. / Е.Н.Згуральская // В сборнике: Современные проблемы проектирования, производства и эксплуатации радиотехнических систем труды XI всероссийской научно-практической конференции. – 2019. – С. 249-251.

11. Згуральская Е.Н. Иерархический кластерный анализ данных и снижение размерности признакового пространства / Е.Н.Згуральская // В сборнике: Современные проблемы проектирования, производства и эксплуатации радиотехнических систем. – 2015. – С. 220-222.
12. Згуральская Е.Н. Анализ структур отношений между описаниями объектов классов и оценки их компактности. / Е.Н.Згуральская // В сборнике: Информационные технологии и нанотехнологии (ИТНТ-2019) труды V международной конференции и молодежной школы. – 2019. – С. 166-170.
13. Згуральская Е.Н. Нелинейные преобразования признаков и поиск закономерностей на данных больных хроническим лимфолейкозом / Н.А.Игнат'ев, Е.Н.Згуральская, М.В.Марковцева // В сборнике: Информационные технологии и нанотехнологии (ИТНТ-2020) труды VI международной конференции и молодежной школы. – 2020. – С. 123-128.
14. Згуральская Е.Н. Способ отбора информативных признаков для решения задачи классификации / В.Р.Крашенинников, Е.Н.Згуральская // REDS: Телекоммуникационные устройства и системы. – 2015. – Т.5, № 4. – С. 324-327.
15. Ignat'eva E.N. (Згуральская Е.Н.) New approach neural networks designing: empirical study on acute myocardial infarction predicting / N.A.Ignat'ev, F.T.Adilova, E.N Ignat'eva // В сборнике: Инфокоммуникационные и вычислительные технологии в науке, технике и образовании труды международной научной конференции. – 2004. – С. 451-454.
16. Патент РФ 2725877. Способ прогнозирования общей выживаемости больных хроническим лимфолейкозом мужского пола в стадии А-С/ Марковцева М. В., Згуральская Е. Н.. опубл. 07.07.2020, Бюл. № 19. – 3 с.

Згуральская Екатерина Николаевна

Повышение эффективности поиска скрытых закономерностей в базах данных
применением интервальных методов на примерах в промышленности и других
областях

Автореферат

Подписано в печать __.__.20__. Формат 60×84/20.

Усл. печ. л. 1,40. Тираж 100 экз. Заказ ____.

ИПК «Венец» УлГТУ, 432027, г. Ульяновск, Северный Венец, д. 32.