

О Т З Ы В

официального оппонента на диссертационную работу Згуральской Е.Н. «Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях», представленной на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность)

Актуальность темы диссертации

В настоящее время в России интенсивно внедряются технологии искусственного интеллекта для решения различного рода практических задач. Одним из приоритетных направлений развития искусственного интеллекта, являются процессы автоматизации извлечения знаний из больших объемов данных. Знания, как правило, содержатся в базах и хранилищах данных в форме скрытых закономерностей. Отсюда возникает задача выявления этих закономерностей.

Большие объемы данных по числу объектов и признаков, описываемых в количественной и номинальных шкалах измерений, наличие пропусков в обучающей выборке значительно ограничивают возможности использования существующих алгоритмов.

В диссертации Е.Н. Згуральской процесс добычи знаний производится через поиск логических закономерностей в форме полуплоскостей, гиперпараллелепипедов и гипершаров с использованием двух критериев. Критерии методов для разбиения упорядоченных значений признака (исходного или латентного) на непересекающиеся интервалы различаются по используемым ограничениям: число интервалов равно числу классов; число интервалов определяется алгоритмическим путём при числе классов равным двум. Для оптимизации критерия с первым ограничением разработан способ предобработки данных, предложена оценка сложности алгоритма метода до и после предобработки.

Алгоритмы на базе интервальных методов позволяют производить анализ данных с пропусками, описываемых в количественной и номинальных шкалах измерений.

Результаты алгоритмов интервальных методов востребованы при отборе информативных наборов признаков, выборе собственного пространства объектов, селекции обучающих выборок объектов, формировании латентного признакового пространства линейными и нелинейными методами, создании if ... then правил для баз знаний.

Таким образом, тема диссертационной работы Згуральской Е.Н., в которой решается задача повышения эффективности интервальных методов выявления скрытых закономерностей в базах данных, является актуальной.

Структура и объём диссертации

Диссертация выполнена в Ульяновском государственном техническом университете. Она состоит из введения, трех глав, заключения, списка литературы из 101 наименования и 2 приложений. Общий объём работы 124 страни-

цы, в том числе 108 страниц основного текста, 31 таблица, 12 рисунков, список литературы на 11 страницах, состоящий из 101 наименования, приложения на 2 страницах, содержащие: скан патента на изобретение и скан акта о внедрения результатов диссертационной работы в гематологическом отделении Государственного учреждения здравоохранения «Ульяновская областная клиническая больница».

Содержание работы

Во введении раскрывается актуальность диссертационной работы, указаны цель и задачи исследования, обозначены научная новизна и практическая ценность основных результатов, сформулированы научные положения, выносимые на защиту.

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 приведён вычислительный эксперимент на данных по сегментации изображений.

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.

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 для селекции обучающих выборок;

– оценкой компактности объектов класса по системе вложенных гипершагов при поиске собственного пространства объекта.

В заключении приведены выводы по работе.

Следует отметить, что работа Згуральской Е.Н. выполнена на высоком научно-техническом уровне, включает необходимые иллюстрации, главы логически согласованы между собой.

Автореферат достаточно полно отражает содержание диссертационной работы.

Научная ценность диссертации

Теоретической основой для обоснования и реализации интервальных методов является гипотеза о компактности классов объектов. Для проверки этой гипотезы используются две меры компактности. Одна в многомерном признаковом пространстве, другая на числовой оси для по оптимальной границе между двумя клас-

сами. При вычислении оптимальной границы не делается никаких предположений о природе среды.

В теорию анализа данных был введён новый показатель «устойчивость признака» в задачах классификации. Устойчивость признака зависит от разбиения его значений на непересекающиеся интервалы и вычисления функции принадлежности по ним. Экспериментально доказано, что значение устойчивости представляет состоятельную оценку на выборках из генеральной совокупности.

Селекция обучающих выборок реализуется через отбор шумовых объектов и неинформативных признаков, минимальное покрытие выборки объектами-эталоном с использованием локальных метрик. Эффективность перечисленных действий подтверждается через повышение обобщающей способности алгоритмов распознавания по мере компактности, вычисляемой через среднее число объектов выборки, притягиваемое одним эталоном минимального покрытия.

Разработаны эвристики двух методов отбора информативных признаков на базе интервальных методов, которые используют:

- правила иерархической агломеративной группировки и принцип динамического программирования;
- рекурсивный алгоритм построения упорядоченной по информативности последовательности разнотипных признаков.

Практическая ценность диссертации

Обнаружение скрытых закономерностей с помощью интервальных методов позволяет повысить обобщающую способность алгоритмов распознавания и обосновывать процесс принятия решений при управлении технических и других систем.

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. Для данных по диагностике неисправностей 8-ми лучевого ультразвукового расходомера значение произведения в $(0;1]$ интерпретировалось как степень неисправности по индивидуальному набору показателей. Выбор от 2 до 6 информативных показателей и оценка по ним степени неисправности устройств сокращают время диагностики, материальные расходы на ремонт и ущерб от неверных значений показателей при измерении.

Разработаны 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. 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, число которых определялось алгоритмами интервальных методов.

Акт внедрения результатов диссертационной работы в гематологическом отделении Государственного учреждения здравоохранения «Ульяновская областная клиническая больница» подтверждает практический опыт полученных результатов.

Оформление диссертации и автореферата

Диссертация и автореферат оформлены аккуратно. Практически нет опечаток в тексте и формулах. Язык изложения ясный и понятный. Автореферат достаточно полно отражает содержание диссертации.

Замечания по диссертации

1. При отборе информативных признаков для диагностики неисправности ультразвуковых расходомеров в качестве меры расстояния использовалась метрика Журавлёва. Так как все признаки в описании объектов количественные, имело смысл сравнить значения меры компактности по числу объектов-эталонов минимального покрытия обучающей выборки и на других метриках.

2. Следовало бы пояснить преимущества, которые даёт свойство инвариантности к масштабам измерений признаков в интервальных методах применительно к задачам управления в технических системах.

3. Не раскрыта связь отбора информативных признаков по иерархическому агломеративному алгоритму с решением проблемы проклятия размерности пространства.

4. Из текста диссертации не ясно, можно ли вычислять устойчивость признаков со значениями в номинальной шкале измерений.

Перечисленные замечания не влияют на общую положительную оценку диссертационной работы.

Заключение

Диссертация Згуральской Е.Н. выполнена на достаточно высоком научно-техническом уровне. Основные научные результаты достаточно полно отражены в 16 публикациях (статьи и тезисы докладов). 4 статьи опубликованы в изданиях, входящих в перечень ВАК, 2 – в изданиях, проиндексированных в Scopus.

Работа по затронутой тематике и используемым методам соответствует следующим пунктам паспорта специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность):

п.5 - разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации;

п.13 - методы получения, анализа и обработки экспертной информации.

Содержание автореферата отражает основные положения работы и главные доказательства их истинности.

Как следует из вышеизложенного, представленная работа Згуральской Е.Н. по актуальности, научно-техническому уровню и практическому значению выполненных исследований, отвечает требованиям п. 9 «Положения о присуждении учёных степеней», утверждённого Постановлением Правительства Российской Федерации от 24.09.2013 № 842 (в ред. Постановления Правительства РФ от 28.08.2017 № 1024) и содержит решение важной научно-технической задачи повышения эффективности методов выявления скрытых закономерностей в базах данных, а ее автор, Згуральская Екатерина Николаевна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01. – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)».

Официальный оппонент

профессор, д.ф-м.н.,

О.Н. Граничин

Подпись О.Н. Граничина заверяю,

И.о. начальника
отдела кадров № 3
И.И. Константинова



24.11.2021

Сведения об оппоненте:

Граничин Олег Николаевич

д.ф-м.н., спец. 05.13.01,

профессор каф. «Системное программирование» СПбГУ,

199034, г. Санкт-Петербург, Университетский проспект, д.28,

тел. +7(812)428-71-09,

e-mail: o.granichin@spbu.ru,

web-сайт: <https://www.math.spbu.ru/user/gran/index.html>

Документ подготовлен
в порядке исполнения
трудовых обязанностей

Текст документа размещен
в открытом доступе
на сайте СПбГУ по адресу
<http://spbu.ru/science/expert.html>