



САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

федеральное государственное автономное
образовательное учреждение высшего образования
«Самарский национальный исследовательский университет
имени академика С.П. Королева»

ул. Московское шоссе, д. 34, г. Самара, 443086
Тел.: +7 (846) 335-18-26, факс: +7 (846) 335-18-36
Сайт: www.ssau.ru, e-mail: ssau@ssau.ru
ОКПО 02068410, ОГРН 1026301168310,
ИНН 6316000632, КПП 631601001

УТВЕРЖДАЮ

Первый проректор –
проректор по научно-

исследовательской работе
доктор технических наук, доцент



А. Б. Прокофьев

« 02 » декабря 2021 г.

02 ДЕК 2021

№ 104-6253

На № _____ от _____

ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

о диссертационной работе Згуральской Екатерины Николаевны
на тему «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

Одним из главных факторов инновационного развития современного общества является использование цифровых технологий в научной и производственной сферах. Важную роль в совершенствовании цифровых технологий играют информационные модели, основанные на знаниях. Неявные знания содержатся в базах и хранилищах данных в форме скрытых закономерностей. Поиск скрытых закономерностей в базах и хранилищах данных является важной задачей разработки и реализации методов интеллектуального анализа данных.

Для решения этой задачи, в частности, могут быть применены интервальные методы. Однако при их использовании возникает ряд проблем. Основной проблемой является высокая комбинаторная сложность алгоритмов, наличие пропусков в данных. Возможности для анализа данных

в предметных областях существенно ограничены из-за отсутствия теоретического обоснования на использование средств контроля за отношениями между объектами, связанными с изменением размерности признакового пространства. Таким образом, актуальным направлением исследований является разработка методов поиска информативных наборов признаков и подмножества объектов обучения, которые обладают лучшим качеством в смысле решения задач интеллектуального анализа, чем исходные множества признаков и объектов.

На решение отмеченных проблем направлена диссертация Е.Н. Згуральской, в которой рассматриваются вопросы выбора способов предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных, предлагаются новые способы оценки обобщающей способности алгоритмов и способы отбора информативных наборов признаков в описании допустимых объектов.

Применение полученных в диссертации результатов позволяет находить закономерности интервальными методами в данных (с возможными пропусками) при приемлемых вычислительных затратах.

Эффективность разработанных алгоритмов проиллюстрирована примерами обработки данных на примерах в промышленности (задача поиска неисправностей расходомера) и других областях (задачи медицины и сегментации изображений).

Таким образом, тема диссертационной работы Згуральской Е.Н., посвященная повышению эффективности поиска скрытых закономерностей в базах данных, является актуальной.

Структура и содержание диссертационной работы

Диссертация состоит из введения, трех глав, заключения, списка литературы из 101 наименования и 2 приложений. Общий объём работы 124 страницы, в том числе 108 страниц основного текста, 31 таблица, 12 рисунков, список литературы на 11 страницах, состоящий из 101 наименования, двух приложений на 2 страницах, содержащих: скан патента на изобретение и скан акта о внедрении результатов диссертационной работы в гематологическом отделении Государственного учреждения здравоохранения «Ульяновская областная клиническая больница».

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 гипершаров при поиске собственного пространства объекта.

В заключении отражены основные результаты диссертации.

Научная новизна и достоверность полученных результатов

Научной новизной обладают следующие результаты диссертационной работы:

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.

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

Ценность результатов диссертационной работы для науки и практики

Научную ценность представляют разработанные в диссертации методы предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных.

Практическое значение результатов работы заключается в том, что обнаруженные скрытые закономерности с использованием интервальных методов позволяет повысить обобщающую способность алгоритмов распознавания и обосновывать процесс принятия решений в технических и других системах.

Рекомендации по использованию результатов и выводов диссертации

Полученные в диссертации результаты могут найти применение в организациях, связанных с разработкой и эксплуатацией встроенных интеллектуальных систем, в которых совмещены процессы «добычи знаний» и принятия управленческих решений. Например, на промышленных предприятиях Ульяновской области АО «Авиастар-СП» и УФ КБ ПАО «Туполев». Кроме того, результаты диссертации рекомендуются использовать в учебном процессе вузов в модуле «Системы искусственного интеллекта» для бакалавриата и специалитета.

Соответствие требованиям по выполнению и оформлению диссертационной работы

Диссертация и автореферат написаны грамотным техническим языком, все положения аргументированы. 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.

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.).

Автореферат полностью отражает содержание диссертационной работы.

Замечания по диссертационной работе

По работе имеются следующие замечания.

1. В работе не приведено описания способа синтеза латентных признаков из наборов номинальных и количественных признаков.

2. Для вычислительных экспериментов, проводимых в диссертационной работе, использовались относительно небольшие наборы данных (например, в задаче по диагностике расходомеров выборка состоит из 180 объектов, описываемых 43 признаками), хотя утверждается, что разработанные алгоритмы работают с Big Data.

3. Для сравнения обобщающей способности алгоритмов на разных наборах признаков автор применяет меру компактности, при вычислении которой используются число шумовых объектов и число объектов–эталонов минимального покрытия выборки. Отсутствует мотивированное обоснование отказа от метода скользящего экзамена для вычисления обобщающей способности.

4. Нет объяснения связи процедур отбора информативных признаков с решением проблемы переобучения алгоритмов.

5. В работе чрезвычайно мало внимание уделено результатам зарубежных авторов (в списке литературы менее 10 источников).

6. Отсутствие публикаций в зарубежных изданиях.

Отмеченные недостатки носят частный характер и не влияют на общую положительную оценку диссертационной работы.

Заключение

Диссертация является законченной научно-квалификационной работой, выполненной на актуальную тему. Новые научные результаты, полученные в диссертации, направлены на решение научной задачи, имеющей важное значение для развития методов системного анализа и обработки экспертной информации. Работа Згуральской Екатерины Николаевны является самостоятельным научно-исследовательским трудом для областей исследований, перечисленных в паспорте 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).

Представленная диссертация на тему «Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях» отвечает требованиям п. 9 Положения о присуждении учёных степеней утверждённого Постановлением Правительства Российской Федерации от 24.09.2013 № 842 (в ред. на 20 марта 2021 года) и содержит решение научной задачи повышения эффективности методов выявления скрытых закономерностей в базах данных, имеющей значение для развития информационных технологий и промышленности, 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).

Отзыв подготовлен профессором кафедры геоинформатики и информационной безопасности Самарского университета доктором физико-математических наук Мясниковым Владиславом Валерьевичем.

Диссертационная работа Згуральской Екатерины Николаевны на тему «Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях» на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность) рассмотрена и обсуждена на НТС кафедры геоинформатики и информационной безопасности Самарского университета, протокол № 2 от 29.11.2021 г.

Заведующий кафедрой геоинформатики
и информационной безопасности
Самарского университета,
д.т.н., профессор

Сергеев В. В.

Профессор кафедры
геоинформатики и информационной
безопасности Самарского университета,
д.ф.-м.н., доцент

Мясников В. В.

Информация об организации:

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

Сокращенное название: Самарский университет.

Почтовый адрес: 443086, Приволжский федеральный округ, Самарская область,
г. Самара, Московское шоссе, 34

телефон +7 (846) 335-18-26, 267 -43 -01,

e -mail: ssau@ssau.ru, Web -сайт: www.ssau.ru