

## **ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА**

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

### **1. Актуальность темы и соответствие паспорту специальности**

В настоящее время в России и во всём мире интенсивно развиваются информационные технологии, идёт процесс цифровизации всех сфер жизни человечества. Основой информационных технологий являются информационные модели, основанные на знаниях. Но эти знания изначально неявны и содержатся в базах и хранилищах данных в виде скрытых закономерностей. Отсюда возникает проблема выявления этих закономерностей. Несмотря на значительные усилия специалистов, проблема эта далека от полного решения, поэтому идёт интенсивный поиск путей повышения эффективности методов и алгоритмов выявления знаний из баз данных.

Одним из таких методов является рассматриваемый в диссертации интервальный метод, состоящий в разбиении значений признаков объектов на интервалы, по принадлежности к которым и делаются выводы о классе, к которому принадлежит конкретный объект. Использование этого метода сопряжено с несколькими недостаточно проработанными вопросами, среди которых в первую очередь следует отметить выбор объективного критерия качества разбиения на интервалы, учёт возможных пропусков в базах данных и разработка технически реализуемых алгоритмов.

Помимо общих вопросов касающихся интервальных методов, в диссертации рассмотрено четыре примера задачи выявления закономерностей в базах данных, подтверждающих актуальность рассматриваемой проблемы: сертификация расходомеров жидкостей, сегментация изображений, диагностика сердечно-сосудистых заболеваний и анализ причин, влияющих на продолжительность срока выживаемости у больных хроническим лимфолейкозом.

В связи с этим не вызывает сомнений актуальность темы диссертации Е.Н. Згуральской, в которой рассматриваются эти и другие вопросы, связанные с разработкой и повышением эффективности интервальных методов поиска скрытых закономерностей в базах данных.

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.

## 2. Структура и объём

Диссертация состоит из введения, трех глав, заключения, списка литературы (103 наименования) и двух приложений (акт внедрения результатов диссертации и патент на изобретение). Работа изложена на 124 страницах, включающих 12 рисунков и 15 таблиц.

**Во введении** обоснована актуальность выполненного исследования, сформулированы цель и задачи работы, определена научная новизна и практическая значимость полученных результатов, сформулированы положения, выносимые на защиту, отмечены апробация и внедрение результатов исследования, дано краткое содержание диссертации.

**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. Приведены результаты вычислительного эксперимента на данных по сегментации изображений.

**Во второй главе** разрабатываются алгоритмы 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. 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** представлены результаты диссертационного исследования.

**В приложениях** приведены копии акта о внедрении результатов диссертации и патента об изобретении.

### **3. Новизна проведённых исследований и полученных результатов**

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.

2. Разработан рекурсивный алгоритм для упорядочивания разнотипных признаков по отношению информативности с использованием предобработки данных путем формирования матрицы близости по парам признаков.

3. Интервальные методы применяются для синтеза латентных признаков, отбора информативного набора разнотипных признаков для всей выборки в целом и для собственного пространства объекта, выбора границ между классами как логических закономерностей в форме полуплоскостей, формирования if...then правил, вычисления обобщенных оценок объектов по нелинейным преобразованиям признаков.

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

#### **5. Публикация и апробация материалов диссертации**

Основные результаты и положения диссертации достаточно полно опубликованы в 16 печатных работах, из них 4 в изданиях из перечня ВАК, 2 индексированы в базе Scopus и 1 патент на изобретение. Результаты диссертации прошли апробацию на трёх международных и трёх Всероссийских конференциях.

#### **6. Практическая ценность**

В диссертации приведено решение задач обнаружения неисправностей ультразвуковых расходомеров жидкости, классификации изображений, медицинской диагностике сердечно-сосудистых заболеваний и анализе причин, повлиявших на продолжительность срока выживаемости у больных хроническим лимфолейкозом.

Для решения этих задач были применены разработанные в диссертации способы выявления скрытых закономерностей и проведены численные эксперименты, показывающие эффективность разработанных алгоритмов.

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

#### **7. Оформление диссертации и автореферата**

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

## **8. Замечания**

1. Для лучшего понимания разработанных в диссертации алгоритмов следовало бы привести их блок-схемы. А нет ни одной.
2. При селекции обучающих выборок применялся поиск и удаление шумовых объектов из множества граничных объектов классов. Нет обоснования использования только этого множества.
3. В работе используется мера компактности, вычисляемая через покрытие выборки эталонами. Следовало бы показать из чего следует единственность значений меры компактности.
4. Следовало бы провести сравнительный анализ информативного набора признаков для всех объектов выборки и состава из собственных пространств объектов.

## **9. Заключение**

Отмеченные недостатки носят частный характер и легко могут быть устранены, поэтому они не влияют на общую положительную оценку работы. Диссертация выполнена на достаточно высоком математическом уровне, аккуратно оформлена. Автореферат даёт ясное представление о выполненной работе. Основные результаты диссертации опубликованы в рецензируемых научных изданиях, апробированы на множестве научных конференций. Имеется практическое внедрение в практику лечебного учреждения, подтверждённое соответствующим актом, и патент на изобретение. Вычислительные эксперименты, выполненные на реальном материале, подтверждают высокую эффективность разработанных алгоритмов.

На основании сказанного считаю, что диссертация Згуральской Екатерины Николаевны «Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях» удовлетворяет 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 её автор

Згуральская Екатерина Николаевна заслуживает присуждения учёной степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информационные технологии и промышленность).

Официальный оппонент:

доктор технических наук, профессор

А.А. Смагин

Подпись профессора Смагина А.А. заверяю:

Сведения об оппоненте:

Смагин Алексей Аркадьевич

доктор технических наук (спец. 05.13.05 - Элементы и устройства вычислительной техники и систем управления), профессор, заведующий кафедрой «Телекоммуникационные технологии и сети» Ульяновского государственного университета

432017, г. Ульяновск, ул. Набережная реки Свияги, 106, корпус 3,

тел. (8422) 37-24-73(5),

web-сайт: <https://www.ulsu.ru/ru/employees/2544/>,

e-mail: smaginaa1@yandex.ru

