

ЗАСЕДАНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д212.277.04

Повестка дня:

Защита диссертации **Згуральской Екатериной Николаевной**

на соискание ученой степени кандидата технических наук:

**"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).**

Официальные оппоненты:

**Граничин Олег Николаевич, доктор физико-математических наук, профессор, профессор кафедры системного программирования математико-механического факультета, Санкт-Петербургский государственный университет**

**Смагин Алексей Аркадьевич, доктор технических наук, профессор, заведующий кафедрой «Телекоммуникационные технологии и сети», Ульяновский государственный университет**

Ведущая организация - **ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева»**

ЗАСЕДАНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 212.277.04

от 29 декабря 2021 года

на заседании присутствовали члены Совета:

- |     |   |                         |          |                     |
|-----|---|-------------------------|----------|---------------------|
| 1.  | Ярушкина Н.Г.,<br>председатель Со-<br>вета      | д.т.н.,<br>профессор    | 05.13.12 | - технические науки |
| 2.  | Киселев С.К. зам.<br>председателя Со-<br>вета   | д.т.н.,<br>доцент       | 05.13.05 | - технические науки |
| 3.  | Наместников А.М.,<br>ученый секретарь<br>Совета | д.т.н.,<br>доцент       | 05.13.12 | - технические науки |
| 4.  | Браже Р.А.                                      | д.ф.-м.н.,<br>профессор | 05.13.05 | - технические науки |
| 5.  | Васильев К.К.                                   | д.т.н.,<br>профессор    | 05.13.01 | - технические науки |
| 6.  | Гладких А.А.                                    | д.т.н.,<br>профессор    | 05.13.01 | - технические науки |
| 7.  | Дьяков И.Ф.                                     | д.т.н.,<br>профессор    | 05.13.12 | - технические науки |
| 8.  | Епифанов В.В.                                   | д.т.н.,<br>доцент       | 05.13.12 | - технические науки |
| 9.  | Иванов О.В.                                     | д.ф.-м.н.,<br>доцент    | 05.13.05 | - технические науки |
| 10. | Клячкин В.Н.                                    | д.т.н.,<br>профессор    | 05.13.01 | - технические науки |
| 11. | Крашенинников<br>В.Р.                           | д.т.н.,<br>профессор    | 05.13.01 | - технические науки |
| 12. | Курганов С.А.                                   | д.т.н.,<br>доцент       | 05.13.05 | - технические науки |
| 13. | Негода В.Н.                                     | д.т.н.,<br>доцент       | 05.13.12 | - технические науки |
| 14. | Самохвалов М.К.                                 | д.ф.-м.н.,<br>профессор | 05.13.05 | - технические науки |
| 15. | Сергеев В.А.                                    | д.т.н.,<br>профессор    | 05.13.05 | - технические науки |
| 16. | Ташлинский А.Г.                                 | д.т.н.,<br>профессор    | 05.13.01 | - технические науки |

Председатель Совета  
д.т.н., профессор

Н.Г. Ярушкина

Ученый секретарь  
д.т.н., доцент

А.М. Наместников



Председатель

**Уважаемые коллеги!**

На заседании диссертационного Совета Д212.277.04 из **23** членов Совета присутствуют 16 человек. Необходимый кворум имеем.

Членам Совета повестка дня известна. Какие будут суждения по повестке дня? Утвердить? (принято единогласно).

По специальности защищаемой диссертации **05.13.01 «Системный анализ, управление и обработка информации» (информационные технологии и промышленность)** (технические науки) на заседании присутствуют 5 докторов наук.

Наше заседание правомочно.

Председатель

Объявляется защита диссертации на соискание ученой степени кандидата технических наук **Згуральской Екатериной Николаевной** по теме: "Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях".

Работа выполнена в Ульяновском государственном техническом университете

Научный руководитель - д.т.н., профессор Крашенинников В.Р.

**Официальные оппоненты:**

**Граничин Олег Николаевич, доктор физико-математических наук, профессор, профессор кафедры системного программирования математико-механического факультета, Санкт-Петербургский государственный университет**

**Смагин Алексей Аркадьевич, доктор технических наук, профессор, заведующий кафедрой «Телекоммуникационные технологии и сети», Ульяновский государственный университет**

Присутствуют оба оппонента, оппонент Граничин О.Н. подключен удаленно через средства телеконференцсвязи.

Письменные согласия на оппонирование данной работы от них были своевременно получены.

Ведущая организация - **ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева».**

Слово предоставляется **Ученому секретарю** диссертационного Совета д.т.н. **А.М. Наместникову Д212.277.04** для оглашения документов из личного дела соискателя.

Ученый секретарь

Соискателем **Згуральской Екатериной Николаевной** представлены в Совет все необходимые документы для защиты кандидатской диссертации (зачитывает):

- заявление соискателя;
- копия диплома о высшем образовании (заверенная);
- справка об обучении в аспирантуре;
- заключение по диссертации от организации, где выполнялась работа;
- отзыв научного руководителя;
- диссертация и автореферат в требуемом количестве экземпляров.

Все документы личного дела оформлены в соответствии с требованиями Положений ВАК.

Основные положения диссертации отражены **Згуральской Е.Н.** в **16** научных работах, в т.ч. в **4** статьях в изданиях из перечня ВАК, **2** публикациях индексируемых Scopus, получен **1** патент. Соискатель представлен к защите **27.10.2021 г.** (протокол №7). Объявление о защите размещено на сайте ВАК РФ **28.10.2021 г.**

Вся необходимая информация по соискателю внесена в ФИС ГНА.

Председатель

Есть ли вопросы по личному делу соискателя к ученому секретарю Совета? (Нет).

Есть ли вопросы к **Згуральской Е.Н.** по личному делу? (Нет).

**Екатерина Николаевна,** Вам предоставляется слово для изложения основных положений Вашей диссертационной работы.

Уважаемый председатель, уважаемые члены диссертационного совета, присутствующие. Разрешите представить результаты диссертационной работы «Повышение эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях».

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. Одной из основных проблем поиска закономерностей является высокая комбинаторная сложность алгоритмов. Важное значение имеет выбор способов предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных, разработка новых методов оценки обобщающей способности алгоритмов распознавания. Основным инструментарием в диссертации для поиска закономерностей в данных являются интервальные методы. Поиск логических закономерностей производится в форме полуплоскостей, гиперпараллелепипедов и гипершаров с использованием двух критериев для разбиения

упорядоченных значений признака (исходного или латентного) на непересекающиеся интервалы, которые различаются по используемым ограничениям: число интервалов равно числу классов; число интервалов определяется алгоритмическим путём при числе классов равным 2. 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.

Цель диссертационной работы: 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. Задачи диссертационной работы представлены на слайдах. Научная 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.

Основные положения, выносимые на защиту представлены на слайдах.

В первой главе рассматриваются интервальные методы в анализе данных. Приводится описание и различие двух критериев для разбиения значений признаков на непересекающиеся интервалы. Постановка задачи. Пусть задано множество объектов  $E_0$ , содержащее представителей  $l$  непересекающихся классов. Описание объектов производится с помощью набора из  $n$  разнотипных признаков  $X(n)$ ,  $\delta$  из которых измеряются в номинальной,  $n-\delta$  в интервальных шкалах ( $\delta < n$ ). Допускается наличие пропусков и повторяющихся значений в данных. Латентные признаки могут представлять комбинации из номинальных и количественных признаков. Требуется определить значения границ  $l$  интервалов, при которых  $F(*) = \text{extr. } l$  критерий. Критерий для разбиения значений признаков на число интервалов, равное числу непересекающихся классов.

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, отнесённые к разным классам. «Схожесть» и «отличие» при этом понимаются в смысле заданного разбиения.

Этот критерий в диссертации используется для:

- разбиения на интервалы исходных и латентных признаков;
- отбора информативных наборов признаков объектов и выборки в целом;
- преобразования количественных признаков в номинальные.

Оптимальное значение первого критерия для двух классовой задачи будем считать, мерой компактности признака на числовой оси в задачах с пересекающимися классами.

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 классами.

Вычислять значения границ интервалов необходимо за приемлемое время. С этой целью разработан численный алгоритм с применением предобработки.

Оценка сложности 1 критерия разбиения значений признаков на число интервалов, равное числу непересекающихся классов, определяется по формуле, представленной на слайде.

Разработанный численный алгоритм с применением предобработки данных требует значительно меньших вычислительных ресурсов при выборе оптимальных значений границ интервалов, чем алгоритм без использования предобработки. Вычисление значения первого критерия разбиения значений признаков на число интервалов, равное числу непересекающихся классов, является NP – полной задачей без использования предобработки данных. Интерес представляет вычисление значения  $u_t^p$  (число объектов класса  $K_p$  попадающих в интервал  $p$ ).

Разработан численный алгоритм для разбиения значений признаков в описании объектов классов на непересекающиеся интервалы с использованием предобработки данных. Затраты ресурсов на вычисление алгоритмом после предобработки имеют полиномиальную сложность.

Для предобработки данных сформируем целочисленную матрицу  $D$ , значения которой вычисляются по формулам, показанным на слайде.

Тестовый пример разбиения на интервалы 16 несовпадающих значений количественного признака показан на слайде.

Оценка сложности алгоритма с использованием предобработки значительно ниже, чем у алгоритма без использования предобработки. Это позволяет вычислять значение первого критерия за полиномиальное время.

Второй критерий, когда число классов равно двум, количество интервалов изначально неизвестно и определяется рекурсивным алгоритмом. Критерий предназначен для поиска минимального покрытия упорядоченной последовательности значений признака непересекающимися интервалами. В границах каждого интервала разность частот встречаемости значений объектов двух классов максимальна. Частоты встречаемости используются для вычисления значений функции принадлежности к классам по интервалам. Показателем качества разбиения на интервалы является устойчивость, множество допустимых значений которой принадлежит  $(0.5; 1]$ . Устойчивость выражает степень однородности (не перемешанности) значений  $s$ -го признака объектов в границах интервалов доминирования. При максимальном значении ус-

тойчивости каждый интервал содержит представителей (значения объектов) только одного класса.

Устойчивость и компактность – в чём разница? Рассмотрим пример, когда разбиение на интервалы не компактно, но устойчиво. Рисунок – а, разбиение на интервалы проводилось по 1 критерию, рисунок – б, по второму. Как видно на втором рисунке каждый интервал содержит представителей только своего класса, поэтому значение устойчивости равно единице.

В диссертации предложена модификация критерия 2 с учётом наличия пропусков в данных. Разработанный рекурсивный алгоритм позволяет вычислять границы интервалов и их число при отсутствии измеренных значений признаков в описании части объектов классов. При наличии пропусков в данных формула для вычисления устойчивости имеет следующий вид.

Показатели устойчивости разбиения значений признака на интервалы на выборках данных из генеральной совокупности имеют приоритетное значение для обнаружения скрытых закономерностей по сравнению с числом непересекающихся интервалов. Данное утверждение продемонстрировано на выборке объектов по сегментации изображений, состоящей из 2300 объектов. 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.

Как было сказано ранее первый критерий использовался для отбора информативных наборов признаков. Результаты анализа отношений между объектами по значениям на числовой оси используются для отбора информативных признаков. Пусть на множестве объектов по набору признаков определена метрика. Объект  $S$  рассматривается как центр гипершара, от которого по упорядоченному множеству расстояний до объектов формируется последовательность вложенных друг в друга гипершаров.

Значения границ интервалов каждого объекта  $S$ , вычисляемые по первому критерию, используется для оценки наборов разнотипных признаков. В качестве меры расстояния применяется метрика Журавлева.

Геометрическая интерпретация формирования упорядоченного множества гипершаров представлена на слайде.

Для формирования набора информативных признаков для выборки в целом предлагается использовать известное правило иерархической агломеративной группировки.

Условие (правило) иерархической агломеративной группировки для включения признака: «При добавлении нового признака значение компактности у более чем половины объектов выборки больше чем, на предыдущем наборе».

В качестве первого шага при отборе информативного набора признаков предлагается выбирать подмножество, состоящее из 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.

Эвристические методы отбора информативных наборов признаков дают локально-оптимальное решение задачи. Чтобы сравнивать наборы признаков, полученных по разным эвристикам, в диссертации предла-

гается использовать методы оценки обобщающей способности, базирующиеся на вычислении компактности.

Обобщающая способность алгоритма по правилу «ближайший сосед» определяется по формуле, представленной на слайде. Показателем компактности выборки является среднее число объектов, притягиваемых одним эталоном минимального покрытия.

Шумовые объекты удалялись с целью повышения обобщающей способности алгоритма. Геометрическая интерпретация обнаружения шумового объекта представлена на слайде.

В диссертации рассматривалась задача по диагностике неисправностей расходомеров. Выборка данных состоит из 4-х классов: K1 (исправный), K2 (впрыск газа), K3 (дефекты установки), K4 (восковая депиляция). В базе данных имеется 180 объектов, описываемых 43 признаками.

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 мультипликативному и аддитивному критерию.

В качестве меры расстояния между описаниями объектов использовалась метрика Журавлева. Значения количественных признаков были пронормированы в  $[0;1]$ .

В диссертации решалась задача селекции обучающих выборок через отбор информативных разнотипных признаков и минимальное покрытие объектами-эталоном. Требовалось найти последовательность информативных наборов, на которых значения меры сложности алгоритмов образуют невозрастающую последовательность.

Еще одним способом использования интервальных методов (первого критерия) для отбора информативных признаков является предобработка данных путем преобразования количественных признаков в номинальные. Новизна этого метода заключается в применении рекурсивного алгоритма для упорядочивания признаков по отношению информативности. Значения признака разбиваются на непересекающиеся интервалы. Номер интервала, в границах которого лежит значение признака, является градацией в номинальной шкале измерений. Вычисляются элементы матрицы различий  $V=\{b_{ij}\}$  между объектами классов по паре признаков  $(x_i, x_j)$ . Матрица различий используется для формирования упорядоченной по информативности последовательности признаков. По

матрице  $V$  выделяется пара  $(x_i, x_j)$  с наибольшим значением  $b_{ij}$  и включается (слева направо) в последовательность. Каждая следующая пара признаков по аналогичному принципу определяется из матрицы  $V$ , после удаления в ней строк и столбцов с номерами  $i$  и  $j$ . Эффектив-



ность поиска информативного набора обеспечивается за счёт ограниченного перебора признаков-кандидатов (справа налево) на последовательное исключение.

Критерий для отбора информативного набора признаков представлен на слайде.

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

- формирование последовательности признаков, инвариантной к масштабам измерений;
- на данных не используется нормирование;
- визуально отслеживается структура отношений объектов по значениям сырых признаков на разных наборах признаков по методу  $t$ -стохастического вложения соседей;
- показаны особенности реализации метода минимального покрытия выборки объектами-эталоном при использовании одной метрики для всех объектов и локальной метрики для каждого.

Для данных по диагностике расходомеров была сформирована последовательность признаков, упорядоченная по отношению информативности, инвариантная к масштабам измерений. Наборы для эксперимента из формировались путём отбрасывания (справа-налево) по 5 признаков. Значения сложности алгоритмов приводятся в таблице. Как видно из таблицы самое большое значение компактности получено при количестве признаков в наборе равном 33, причем сложность значительно ниже, чем на наборе из 44 признаков. Исходя из вышесказанного рекомендуется для классификации неисправностей использовать набор из 33 признаков.

В ходе вычислительного эксперимента была показана эффективность выбора эталонных объектов с локальной метрикой, данный результат представлен на слайде.

Результаты оценки компактности объектов класса по системе вложенных гипершаров при поиске собственного пространства объекта представлены на слайде. Выбор собственного пространства объекта позволяет определить набор признаков, значения которых являются причиной неисправности конкретного устройства.

Практические результаты диссертационной работы применялись в Ульяновской областной клинической больнице.

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 K1 (фактическая выживаемость меньше прогнозируемой общей выживаемости) и K2 (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 лимфолей-

козом. 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 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. Основные результаты диссертации представлены на слайдах.

По теме диссертационной работы опубликовано 16 печатных работ, из них 5 - в изданиях из перечня ВАК, 2 Scopus, 1 патент на изобретение. Получена справка о внедрении результатов диссертационной работы в гематологическом отделении ГУЗ Ульяновской областной клинической больницы.

Спасибо за внимание!

Председатель

У кого есть вопросы к соискателю?

Вопросы

д.т.н., профессор Васильев К.К.

Екатерина Николаевна, цель работы: повышение эффективности поиска скрытых закономерностей. Что Вы понимаете под эффективностью, которую Вы собираетесь повысить или Вы повысили?

Соискатель

Снижение вычислительной сложности при поиске закономерностей

д.т.н., профессор Васильев К.К.

Т.е. эффективность - вычислительная сложность, так что ли?

Соискатель

Качество при приемлемых вычислительных затратах.

д.т.н., профессор Васильев К.К.

Просто эффективность - это вычислительная сложность. Вы хотите ее улучшить, повысить эффективность, снизить вычислительную сложность? Правильно?

Соискатель

Да.

д.т.н., профессор Васильев К.К.

С учетом современной вычислительной сложности, - это какие времена? Ну вот я возьму компьютер помощнее и побыстрее, все буду вычислять, т.е. с точки зрения другое качество, только время и вычислительная сложность?

Соискатель

Принимать решения необходимо в режиме реального времени, при меньшем количестве измеримых параметров.

д.т.н., профессор Васильев К.К.

Всё-таки качество решения тоже самое будет? Если возьму старые алгоритмы, но более мощный компьютер? Или нет? Во сколько раз Вы уменьшаете время?

Соискатель

Разработанный численный алгоритм является частью других методов.  
д.т.н., профессор Васильев К.К.

Примерно во сколько раз в объёме операций Вы выигрываете?

Соискатель

Задача является Np-полной, полным перебором решить ее невозможно.  
д.т.н., профессор Васильев К.К.

Я посмотрел заключение, практическая значимость? Вы же защищаетесь по техническим наукам? Значит, Вы должны сказать, где в технике, в технических приложениях удастся применить ваши алгоритмы и какое они дают улучшение? Что Вы можете сказать по этому поводу?

Соискатель

В диссертации рассматривалась задача по диагностике расходомера. В режиме реального времени можно диагностировать неисправность.  
д.т.н., профессор Васильев К.К.

Во сколько применительно к этому устройству удастся сократить объёмы и время?

Соискатель

Решение о неисправности устройства принимается в режиме реального времени.

д.т.н., профессор Васильев К.К.

Во сколько раз удастся сократить объёмы и время?

Соискатель

Численной оценки нет.

д.т.н., доцент Киселев С.К.

Как Вы понимаете реальное время? В технике есть достаточно жесткое понятие реального времени. Вы ловко оперируете понятием реального времени. У вас реальное время – это часы, сутки?

Соискатель

Для принятия решения необходимо 12 эталонных объектов и 33 признака. Решение принимается по правилу «ближайший сосед».

д.т.н., доцент Киселев С.К.

А реальное время?

д.т.н., профессор Васильев К.К.

На сколько задержка влияет на ход процесса, на управление? Это секунды?

Соискатель

Несколько секунд.

Киселев С.К.

Можно тогда привести пример такой системы?

Соискатель

В рамках диссертационного исследования данные вопросы не рассматривались

д.т.н., профессор Ташлинский А.Г.

В пункте 3 научной новизны синтезируются латентные признаки, они синтезируются на базе исходных признаков? Число латентных признаков может отличаться от числа исходных?

Соискатель

Латентные признаки получены как комбинации исходных признаков по операциям умножения и деления.

д.т.н., профессор Ташлинский А.Г.

Допустим, у вас 5 исходных признаков? Латентных сколько будет?

Соискатель

Мы перебираем все возможные комбинации.

д.т.н., профессор Ташлинский А.Г.

При синтезе латентных признаков Вы какую-то априорную информацию используете?

Соискатель

Нет.

д.т.н., профессор Ташлинский А.Г.

Т.е. Вы из некоторого набора признаков синтезируете другие? И утверждаете, что у этих синтезированных больше информации чем у исходных? Почему?

Соискатель

Теоретической основой использования интервальных методов является гипотеза о компактности. Если синтезированный латентный признак имеет более высокое значение компактности, принятие решения по данному латентному признаку предпочтительнее.

д.т.н., профессор Ташлинский А.Г.

Для данного конкретного алгоритма. Который из них может быть плохой, кривой и т.д. Но Вы же не так утверждаете, Вы вообще говорите, мы взяли сколько-то исходных признаков, априорную информацию не используем, на их базе скомбинируем каким-то другим образом и информации стало больше. Или это просто такая формулировка?

Соискатель

Такая формулировка.

д.т.н., профессор Ташлинский А.Г.

Вы синтезируете, задаете целевую функцию. Что является целевой функцией?

Соискатель

Критерий.

д.т.н., профессор Ташлинский А.Г.

Вы говорите об информативности постоянно. Т.е. это некая информация подразумевается под этим? Как Вы численно измеряете объем информации?

Соискатель

Уточните вопрос?

д.т.н., профессор Ташлинский А.Г.

Вы говорите «повышается информативность»? Т.е. есть численная оценка? Информативность — это некий объем информации, который Вы извлекаете из этих признаков? Или это что-то другое?

Председатель

Вы, наверное, имели введу значимость?

Соискатель

Значимость.

д.т.н., профессор Ташлинский А.Г.

Спасибо, я понял. И вот Вы говорите о локально-оптимальном покрытии. Эвристически оптимальном?

Соискатель

Эвристические методы отбора информативных наборов признаков дают локально-оптимальное решение задачи.

д.т.н., профессор Ташлинский А.Г.

Оптимальное, значит есть критерий оптимальности? Значит это уже не эвристическое решение?

Соискатель

При отборе информативных наборов признаков мы используем различные эвристики. Мы получаем результат, одно из возможных решений задачи.

д.т.н., профессор Ташлинский А.Г.

Если оптимальный, то есть критерий оптимальности? Оптимальный по какому-то критерию. Или просто оптимальный?

Соискатель

Поскольку решается некорректная задача, полным перебором, решить которую невозможно. Ищем локально-оптимальное решение задачи.

д.т.н., профессор Ташлинский А.Г.

И вот сокращение времени за счет предобработки, а с учетом предобработки время сокращается? А предобработка сколько времени занимает?

Соискатель

Предобработка выполняется в ходе вычисления значения самого критерия.

д.т.н., профессор Ташлинский А.Г.

Быстродействие повышается?

Соискатель

Быстродействие повышается, сложность снижается.

д.т.н., профессор Ташлинский А.Г.

Но опять-таки для конкретного алгоритма?

Соискатель

Предобработка используется для вычисления значений первого критерия

д.т.н., профессор Сергеев В.А.

До этого была формула критерия качества разбиения. Поясните пожалуйста, можно плакат 8.  $K_i$  - число положительное, зачем модуль брать?

Соискатель

Это мощность множества.

д.т.н., профессор Сергеев В.А.

Задача диагностики расходомеров всё-таки ближе к техническим приложениям, поясните мне пожалуйста у вас задавалась таблица неисправностей и таблица признаков?

Соискатель

Да задавалась.

д.т.н., профессор Сергеев В.А.

А она где?

Соискатель

На слайдах ее нет.

д.т.н., профессор Сергеев В.А.

Хорошо, можно тогда еще раз повторить, сколько было неисправностей, сколько дефектов и сколько признаков, которые Вы пытались привязать к этим неисправностям?

Соискатель

Было 180 объектов, 4 класса рассматривались, один класс исправный и три класса с различными видами неисправностей.

д.т.н., профессор Сергеев В.А.

С какими видами то?

Соискатель

Три вида неисправностей впрыск газа, дефекты установки, восковая депиляция.

д.т.н., профессор Сергеев В.А.

И 43 признака?

Соискатель

Да.

д.т.н., профессор Сергеев В.А.

И вам удалось что? Повысить достоверность диагностики?

Соискатель

Сократить набор признаков для принятия решения о неисправности устройства и, второе, определить набор признаков, которые являются причиной неисправности каждого устройства.

д.т.н., профессор Сергеев В.А.

По сравнению с существующими методами диагностики, что это дало, я хочу понять? Вот Вы предложили признаки разделить по интервалам. Это что улучшило, т.е. достоверность диагностики повысилась? Или до сих пор никто не выявлял дефекты по совокупности этих признаков?

Соискатель

Данную задачу рассматривали американские ученые. Эту выборку я брала из machine learning repository. Основным недостатком, как отмечают сами авторы, является время распознавания неисправностей, при достаточно высокой точности распознавания.

д.т.н., доцент Негода В.Н.

Если взять книжку «Техническая диагностика» Маркоменко в 2-х томах, мы можем прийти к методу анализа иерархий, теории вопросников, мы можем прийти к таблице решений, где будут ваши неисправности по горизонтали, а по вертикали эти признаки. Мы можем прийти к тому и к другому, и это используется десятки лет, и это когда мы говорим про один прибор. Теперь у меня вопрос, когда мы говорим, что у нас база данных 180 приборов, что мы делаем такого применительно к этой совокупности, что улучшает техническую диагностику существующую, которую мы применяем, к отдельному прибору, когда мы по признакам его не диагностируем отдельно. Тут два вопроса возникает, что дает эта база данных, когда 180 объектов вместе обрабатываем, и второй вопрос, Вы же их вместе классифицируете, что это дает для диагностики отдельно прибора, и еще вопрос, который возникает, а если у нас будет другая выборка из 180 объектов? Вы говорите, что сократили число признаков, оно по-другому сократится? Или нет?

Соискатель

В диссертации предлагается использовать выбор собственного пространства объекта, который позволяет принимать решение о неисправности каждого конкретного устройства, определить набор признаков – причины неисправности устройства и значения критерия, которое показывает степень неисправности каждого конкретного устройства. В такой постановке эту задачу, насколько мне известно, никто не рассматривал.

д.т.н., доцент Негода В.Н.

В голове возникает, что агрегируют сведения о разных объектах, благодаря тому, что их много – 180. Но вот этот механизм агрегации на основе деления пространства на объекты, расстояния близости, оно кажется немножко схоластическим. Оно не обусловлено смыслом параметров, которые все решает.

Председатель

Коллеги здесь не диагностическая таблица, здесь таблица классификации.

д.ф.-м.н., профессор Браже Р.А.

Вы предлагаете новый способ отыскания скрытых закономерностей путем применения интервальных методов. В чем ограниченность этого способа?

Председатель

Наверное, наоборот, где эти методы не применимы?

Соискатель

Для того, чтобы применять эти методы данные, должны быть формализованы и представлены в виде таблицы объект свойство.

д.ф.-м.н., профессор Браже Р.А.

А Вы сами не можете формализовать, если Вам дали какой-то объект?

Соискатель

Обычно это делают совместно со специалистами в предметной области.

д.ф.-м.н., профессор Браже Р.А.

Т.е. если данные представлены в табличном виде, для всех случаев можно?

Соискатель

Да

д.ф.-м.н., профессор Браже Р.А.

А если у Вас не два класса?

Соискатель

Первый критерий работает для количества классов больше двух.

д.ф.-м.н., профессор Браже Р.А.

В презентации я нашел научную новизну, а практическая значимость не обозначена. Скажите, пожалуйста, где это может использоваться?

Соискатель

Разработанные алгоритмы являются универсальными, и могут быть использованы в слабо формализуемых областях. Могут применяться и в технике, и в медицине.

д.ф.-м.н., профессор Браже Р.А.

Хотелось бы больше примеров.

Соискатель

В диссертации представлено 4 примера: по диагностике расходомера, по сегментации изображений и два из области медицины.

д.т.н., профессор Дьяков И.Ф.

У вас решалась задача классификации по выживаемости пациентов?

Соискатель

Нет, производился анализ данных с целью выявления закономерностей, которые использовались для принятий решений при выборе тактики лечения пациентов.

д.т.н., доцент Наместников А.М.

Поясните степень компактности у Вас в одной из таблиц.

Соискатель

Используется гипотеза о компактности. Для проверки этой гипотезы используются две меры компактности. Одна в многомерном признаковом пространстве, другая на числовой оси.

Председатель

Есть еще вопросы? (Нет).

**Согласны ли члены Совета сделать технический перерыв? (Нет).**

Тогда продолжаем работу.

Слово предоставляется научному руководителю работы **профессору Крашенинникову В.Р.**

Я должен охарактеризовать соискателя. Насколько соискатель соответствует нашему представлению о научной работе. Должен вас заверить, Екатерина Николаевна выполнила эту работу самостоятельно. Очень большая доля ее собственного труда. Даже в публикации мы только в одной поучаствовали вместе. Она способна сама ставить задачи, сама их решать. Многие из защитившихся у меня аспирантов и аспирантов у сидящих здесь, после защиты пропадали из поля зрения. Екатерина Николаевна, я убежден, будет дальше продолжать заниматься этой своей тематикой. Мы уже наметили, обозначили, то что можно еще сделать. Я очень доволен настоящим и надеюсь будущим научным сотрудником. Мы не ошибемся, если присудим ей учёную степень, и она будет работать и дальше.

*(Отзыв прилагается).*

Председатель

**Ученому секретарю Совета** предоставляется слово для оглашения заключения организации, где выполнялась работа и отзыва ведущей организации.

**Ученый секретарь** оглашает заключение организации, где выполнялась работа. Затем зачитывает отзыв ведущей организации.

(Заключение и отзыв прилагаются).

Председатель

На автореферат диссертации поступило 6 отзывов, все они положительные. Согласны ли члены Совета заслушать обзор отзывов или зачитать их полный текст?

Слово для обзора отзывов, поступивших на диссертацию, предоставляется **Ученому секретарю Совета.**

**Ученый секретарь зачитывает обзор отзывов.**

*(Отзывы прилагаются).*

**1. Сибирский государственный университет науки и технологий им. акад. М.Ф. Решетнева.**

Отзыв подписан заведующим кафедрой «Информатика и компьютерная техника», д.т.н., профессором М.Н. Фаворской.

Замечания: 1) В примере по сегментации изображений исследуется устойчивость к пропускам данных, которая по показателю «устойчивость разбиения» представляется высокой (при 30% пропусков снижение этого показателя не более 5%). Но остается не ясным, на сколько при этом может снизиться точность классификации, например, процент правильных решений; 2) Предлагаются алгоритмы поиска закономерностей интервальными методами с возможными пропусками в данных, однако ничего не говорится о выбросах в данных.

**2. Московский государственный университет им. М.В. Ломоносова**

Отзыв подписан профессором кафедры «Автоматизация научных исследований», д.ф.-м.н., профессором А.Г. Шишкиным.



Замечание: В критерии (формула (1) на стр. 8) применены термины «мера» внутриклассового сходства, межклассового различия и компактности. Однако эти функции не являются мерами в точном математическом смысле этого термина (счетно-аддитивная функция множества). Следовало бы использовать другое название, например, «показатель».

### **3. Удмуртский федеральный исследовательский центр Уральского отделения Российской академии наук**

Отзыв подписан к.т.н., научный сотрудник А. И. Баженовой  
 Замечание: В критерии (формула (1)) меры внутриклассового сходства и межклассового различия перемножаются, то есть рассчитываются одинаково. Но одна из них может считаться более важной, чем другая. Можно ли видоизменить этот критерий, чтобы в случае необходимости была возможность учесть такую разницу?

### **4. Чувашский государственный университет им. И.Н. Ульянова**

Отзыв подписан к.т.н., доцент, доцент кафедры «Общая физика» В.И. Семеновым

Замечание: На стр. 17 автореферата написано: «Признаки, значение устойчивости которых лежат в  $[0.9, 1]$  при числе интервалов не больше 4, предложено использовать для формирования if...then правил». Но нет объяснения этой рекомендации. Почему именно от 0.9? Может быть, можно считать значение этого критерия чем-то в роде функции принадлежности?

### **5. Национальный Университет Узбекистана им. М. Улугбека**

Отзыв подписан заведующим кафедрой «Алгоритмы и технологии программирования», д.т.н. Ш.Ф. Мадрахимовым

Замечания: 1) Утверждение, что в каждом интервале должно находиться как можно больше объектов верно лишь для одного критерия. Для другого критерия доминирование представителей класса определяется через частоту их встречаемости в интервале; 2) В работе доказывается, что число интервалов доминирования на выборках из генеральной совокупности не является строго фиксированным. Необходимо было показать, как это свойство интервального метода использовать при формировании if...then правил.

### **6. АО Ульяновское конструкторское бюро приборостроения**

Отзыв подписан главным конструктором направления «Радиоэлектронная аппаратура», к.т.н., С.В. Черкашиным

Замечание: В диссертации рассматривается задача по диагностике устройства (расходомер), однако не понятно, как будут выявляться неисправности в режиме реального времени.

Председатель

Слово для ответа на замечания по заключению и отзывам предоставляется соискателю.

Соискатель

С замечаниями согласна.

Председатель

Слово для отзыва предоставляется официальному оппоненту **-д.ф.-**  
**м.н. Граничину Олегу Николаевичу.**

(Отзыв прилагается).

Председатель

Соискателю предоставляется слово для ответа на замечания оппонента.

Соискатель

С замечаниями согласна.

Председатель

Слово для отзыва предоставляется официальному оппоненту **-**  
**д.т.н. Смагину Алексею Аркадьевичу.**

(Отзыв прилагается).

Председатель

Слово для ответа на замечания оппонента предоставляется соискателю.

Соискатель

С замечаниями Алексея Аркадьевича согласна, хотелось бы дать пояснения по 4 замечанию. Цели отбора информативного набора признаков и собственного пространства объекта разные. В первом случае - это селекция обучающей выборки для повышения обобщающей способности алгоритмов. Для второго случая особенности принятия решений по конкретному объекту.

Председатель

Кто хочет выступить?

д.т.н., профессор Васильев К.К.

Мне два вопроса не понравились. Первое, это приложения. То, что я думаю, совершенно упущена возможность по классификации объектов. На изображениях лес, реки природа. Это можно привязать к задачам мониторинга Земли, как-то вот это, наоборот, подать больше, потому что вот эти два вопроса расходомер и изображение - два доказанных способа применения. Конечно, здесь можно было бы поработать, потому что здесь эти методы могут применяться гораздо шире, т.е. получился стиль физмат почти. Второе, это если мы что-то улучшаем, то что-то ухудшается, т.е. если мы сокращаем вычислительные затраты, то я уверен, что существует набор таких предметов с такими признаками, с которыми вычислительные затраты, наоборот, будут больше или качество распознавания будет гораздо хуже, т.е. вот выявить бы объекты, на которых эти методы не работают - это тоже задача соискателя вместе с руководителем. Все остальное, в целом, - это очень большая работа интересная, но за исключением этих двух моментов мне все остальное очень понравилось. Я буду голосовать «за».

д.т.н., профессор Ташлинский А.Г.

Тематика работы такова, что каждый ее может интерпретировать на свою область. Отсюда и множество вопросов разнородных. Работа не мертвая и в ней ошибки есть. Есть недочеты, мне, в частности, не все понравилось, если говорить о синтезе, это не перебор вариантов. Второе, решаются задачи кластеризации и классификации, хотелось бы как-то четко их разделить переход от одной к другой. Но опять-таки, если кластеризация, классификация, тут напрашивается нейронная сеть, там тоже обучение, обучающая выборка, контрольная выборка, тем более использует стандартные data set. Но этого вот не было, как бы сравнения такого сделано, хотя есть какие-то результаты стандартные, которые можно было бы сравнивать. С другой стороны, те же результаты они вполне могут быть применимы в тех же глубоких нейронных сетях и создания декодеров, где бы учитывались те же не меры, но хотя бы критерии эффективности, которые вы предлагали. И самое главное эффективность и устойчивость, как-то задействовать и, наверное, для расширения базы данных генерация условий, поскольку они дают определенное правило близости. А в целом я слышал эту работу года два назад, наверное, она сильно выигрывает сейчас по содержанию, хотя основа осталась. Я считаю, что стоит проголосовать за соискателя.

д.т.н., профессор Негода В.Н.

Ну вот, раз сказано, что речь идет о скрытых закономерностях в базах данных, хотелось бы четче видеть, вообще-то, во что материализуется представление закономерностей, скрытых потому что ведь на самом деле формула – это зависимость какая-то, графика, это сигнатура какая-то, функциональная зависимость. Что от чего зависит? Сигнатуру-то я увидел, когда она заговорила о множестве признаков, а вот, что еще, если бы это четче было бы представлено, то конечно бы она выиграла, но во время защиты мы вот тут втроем так дискутировали бурно, и перспективы у этой работы вдруг в ходе этих дискуссий возникали и работа эта та, внутри которой много таких перспектив. Эта работа заслуживает поддержки всяческой. Я буду голосовать в связи с этим «за». В рамках того, что защищалось, была очень устойчиво представлена.

д.т.н., профессор Клячкин В.Н.

Я слушал эту работу, наверное, больше всех из членов совета потому, что Екатерина Николаевна дважды докладывала ее на кафедре, когда первый раз она представлялась, и уже перед защитой еще раз, потом на НТС и вот сейчас. Я должен сказать, что я очень удовлетворен этой работой. Я не задавал вопросов потому что во время заседания кафедры я ей задал их, на все получил свои ответы. Второе, что я хотел бы подчеркнуть, вот у меня защитилось 8 аспирантов и не один из них не работает в университете, к сожалению, а вот Екатерина Николаевна уже давно работает. Мало того, что она работает в университете, она в авиационном филиале у нас работает, она умудряется нашу кафедру каждые два года привлекать к участию в конференции, которую проводит авиационный филиал, хотя вроде кафедра от авиации достаточно далеко, но вот в этой конференции мы благодаря Екатерине Николаевне принимаем участие, поэтому безусловно буду поддерживать работу.

Председатель

Кто еще хочет выступить? Нет желающих?

Тогда я вот тоже не удержусь и скажу буквально немного. Мне эта работа очень интересна была всегда, и вот как бы подхватить мысль Александра Григорьевича, что у каждого к этой работе возникают свои ассоциации. Я считаю, что вообще работа в этом направлении имеет важное значение, она идет вот такой линией развития изучения интеллектуальных систем, как экспертные системы, а экспертные системы предполагают классификационные таблицы, а классификационные таблицы строятся на основе признаков, и как человек, который строил не просто экспертные системы, а нечеткие экспертные системы, я отлично знаю, что такое редукция признаков из которых потом возникает редукция правил «если ... то» и даже понимаю, каким образом это попадает в контур реального времени. Потому, как только у нас есть редуцированные правила «если ... то», экспертная система и мы зашиваем их в нечеткий аппаратный контроллер, то он начинает работать в контуре с обратной связью, управлять может чем угодно впрыском топлива в автомобиле, он может управлять системой кондиционирования и там после происшедшей редукции и после происшедшей классификации на их основе набора состояний вот это все осуществляется. Поэтому вопросы, которые задавались я знаю, что исторически ответы есть. Второе, я считаю, что очень важная работа если исследователи работают на уровне явной работы со знаниями. Почему? Потому что очень большой успех глубокого обучения приводит всех к тому, что работать нужно со знаниями неявными, нужно использовать сети глубокого обучения, но, например, сверточные, которые многослойные. А что делают сверточные сети кумулятивные? Давайте задумаемся от одного до другого свертывают определенный ряд пикселей, есть аналогия к тому, что они отыскивают скрытый информационный признак, по которому потом идет некое распознавание. Есть, но это делается совершенно не осознанно. Эта работа лежит в рамках другого направления работы по инженерии знаний, где эта работа происходит с помощью явно прописанного алгоритма. На мой взгляд, стратегически важно, чтобы такие работы ни в коем случае не исчезали, параллельно велись и появлялись. Во-вторых, Александр Григорьевич упомянул инкодеры и автокодеры, а что это такое, как не редукция пространства признаков? Это только редукция пространства признаков. Но как осуществляют их соответствующие модные сейчас виды глубоких нейронных сетей – неосознанно, в рамках поиска нового представления объектов. А вот здесь это происходит по другой технологии осознанно, когда признак потом вытаскивается на поверхность и такой признак стал сигнатурой, он явный. Вы же говорили Виктор Николаевич, а где же там найденные скрытые закономерности, функции, я вижу только одну сигнатуру? К сожалению, наш соискатель не упомянул, то что написано в автореферате и диссертации, генерируется набор «если ... то». Если признаки имеют такие значения, и они стоят в левой части, то в выходной части стоит то-то и то-то, и мы получаем табличное задание функции, т.е. выявленную скрытую закономерность. Вот из методологических соображений, которые я вам сейчас изложила и своей точки зрения на развитие нашей интеллектуальной науки, я считаю, что поддержка таких работ важна. Кроме того, задачи, которые соискатель поставил перед собой, решены, поэтому я считаю, что это очень важная работа, буду голосовать

«за», тем более что Екатерина Николаевна работает в нашем университете и я думаю, что впереди в нашем университете у нее будут замечательные перспективы.

**Соискателю предоставляется заключительное слово.**

Председатель

Переходим к голосованию. Какие будут предложения по составу счетной комиссии? Поступили предложения включить в состав счетной комиссии Ташлинского А.Г., Сергеева В.А., Негоду В.Н.

Прошу голосовать. Возражений нет.

Председатель

Прошу счетную комиссию приступить к работе.

(Счетная комиссия организует тайное голосование.)

Председатель

Коллеги! Продолжаем нашу работу. Слово предоставляется председателю счетной комиссии Ташлинскому А.Г.

Оглашается протокол счетной комиссии.  
(Протокол счетной комиссии прилагается).

Кто против? (Нет).

Кто воздержался? (Нет).

Протокол счетной комиссии утверждается.

Таким образом, на основании результатов тайного голосования (за - 16 , против - нет , недействительных бюллетеней - нет) диссертационный совет Д212.277.04 при Ульяновском государственном техническом университете признает, что диссертация **Згуральской Е.Н.** содержит новые решения по повышению эффективности поиска скрытых закономерностей в базах данных применением интервальных методов на примерах в промышленности и других областях, соответствует требованиям, предъявляемым к кандидатским диссертациям (п.9 "Положения" ВАК), и присуждает **Згуральской Екатерине Николаевне** ученую степень кандидата технических наук по специальности **05.13.01.**

Председатель

У членов Совета имеется проект заключения по диссертации **Згуральской Е.Н.** Есть предложение принять его за основу. Нет возражений? (Нет). Принимается.

Какие будут замечания, дополнения к проекту заключения?

**(Обсуждение проекта) .**

Председатель

Есть предложение принять заключение в целом с учетом редакционных замечаний. Нет возражений? Принимается единогласно.

**Заключение объявляется соискателю.**

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д212.277.04, СОЗДАННОГО НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ «УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»  
ПО ДИССЕРТАЦИИ

НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ КАНДИДАТА НАУК

аттестационное дело № \_\_\_\_\_  
решение диссертационного совета от 29.12.2021 № 11

О присуждении Згуральской Екатерине Николаевне (Российская Федерация) ученой степени кандидата технических наук.

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

по специальности 05.13.01 – «Системный анализ, управление и обработка информации (информационные технологии и промышленность)» принята к защите 27.10.2021 (протокол заседания № 7) диссертационным советом Д212.277.04, созданным на базе ФГБОУ ВО «УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ» (432027, г. Ульяновск ул. Северный Венец, 32) №678/НК от 18.11.2020.

**Соискатель** Згуральская Екатерина Николаевна «11» июня 1981 года рождения.

В 2015 году соискатель окончила очную аспирантуру в ФГБОУ ВО «Ульяновский государственный технический университет», работает старшим преподавателем на кафедре «Самолетостроение» в ФГБОУ ВО «Ульяновский государственный технический университет».

Диссертация выполнена на кафедре «Самолетостроение» в ФГБОУ ВО «Ульяновский государственный технический университет».

**Научный руководитель** – доктор технических наук (05.12.14 Радиолокация и радионавигация) Крашенинников Виктор Ростиславович, ФГБОУ ВО «Ульяновский государственный технический университет», профессор кафедры «Прикладная математика и информатика», профессор.

**Официальные оппоненты:**

**Граничин Олег Николаевич**, доктор физико-математических наук, профессор, профессор кафедры «Системное программирование», Санкт-Петербургский государственный университет.

**Смагин Алексей Аркадьевич**, доктор технических наук, профессор, заведующий кафедрой «Телекоммуникационных технологий и сетей», Ульяновский государственный университет.

дали положительные отзывы на диссертацию.

Ведущая организация Федеральное государственное автономное образовательное учреждение высшего образования «Самарский националь-

ный исследовательский университет имени академика С.П. Королева» в своем положительном отзыве, подписанном Сергеевым Владиславом Викторовичем, доктором технических наук, профессором, заведующим кафедрой геоинформатики и

информационной безопасности, указала, что представленная диссертация 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)».

Соискатель имеет 22 опубликованных работы, в том числе по теме диссертации опубликовано 16 работ, из них 4 работы в изданиях из перечня ВАК. Общий объем опубликованного материала составляет 5,3 п.л., авторский вклад 80%. Недостоверные сведения об опубликованных соискателем работах отсутствуют.

Наиболее значимые работы по теме диссертации:

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. (лично соискателем 40%)

На диссертацию и автореферат поступило 6 отзывов, все отзывы положительные. Отзывы поступили из:

**1. Московский государственный университет им. М.В. Ломоносова (г. Москва).** Отзыв подписан профессором кафедры «Автоматизация научных исследований», д.ф.-м.н., профессором **А.Г. Шишкиным**. **Замечание:** В критерии (формула (1) на стр. 8) применены термины «мера» внутриклассового сходства, межклассового различия и компактности. Однако эти функции не являются мерами в точном математическом смысле этого термина (счетно-аддитивная функция множества). Следовало бы использовать другое название, например, «показатель».

**2. Сибирский государственный университет науки и технологий им. акад. М.Ф. Решетнева (г. Красноярск).** Отзыв подписан заведующей кафедрой «Информатика и компьютерная техника», д.т.н., профессором **М.Н. Фаворской**. **Замечания:** 1) В примере по сегментации изображений исследуется устойчивость к пропускам данных, которая по показателю «устойчивость разбиения» представляется высокой (при

30% пропусков снижение этого показателя не более 5%). Но остается не ясным, на сколько при этом может снизиться точность классификации, например, процент правильных решений. 2) Предлагаются алгоритмы поиска закономерностей интервальными методами с возможными пропусками в данных, однако ничего не говорится о выбросах в данных.

**3. Удмуртский федеральный исследовательский центр Уральского отделения Российской академии наук (г. Ижевск).** Отзыв подписан к.т.н., научный сотрудник **А. И. Баженовой**. **Замечание:** В критерии (формула (1)) меры внутриклассового сходства и межклассового различия перемножаются, то есть рассчитываются одинаково. Но одна из них может считаться более важной, чем другая. Можно ли видоизменить этот критерий, чтобы в случае необходимости была возможность учесть такую разницу?

**4. Чувашский государственный университет им. И.Н. Ульянова (г. Чебоксары).** Отзыв подписан к.т.н., доцент, доцент кафедры «Общая физика» **В.И. Семеновым**. **Замечание:** На стр. 17 автореферата написано: «Признаки, значение устойчивости которых лежат в  $[0.9, 1]$  при числе интервалов не больше 4, предложено использовать для формирования if...then правил». Но нет объяснения этой рекомендации. Почему именно от 0.9? Может быть, можно считать значение этого критерия чем-то в роде функции принадлежности?

**5. Национальный Университет Узбекистана им. М. Улугбека (г. Ташкент).** Отзыв подписан заведующим кафедрой «Алгоритмы и технологии программирования», д.т.н. **Ш.Ф. Мадрахимовым**. **Замечания:** 1) Утверждение, что в каждом интервале должно находиться как можно больше объектов верно лишь для одного критерия. Для другого критерия доминирование представителей класса определяется через частоту их встречаемости в интервале. 2) В работе доказывается, что число интервалов доминирования на выборках из генеральной совокупности не является строго фиксированным. Необходимо было показать, как это свойство интервального метода использовать при формировании if...then правил.

**6. АО Ульяновское конструкторское бюро приборостроения (г. Ульяновск).** Отзыв подписан главным конструктором направления «Радиоэлектронная аппаратура», к.т.н., **С.В. Черкашиным**. **Замечание:** В диссертации рассматривается задача по диагностике устройства (расходомер), однако не понятно, как будут выявляться неисправности в режиме реального времени.

Выбор официальных оппонентов и ведущей организации обосновывается их высокой компетенцией, научными достижениями и наличием публикаций в соответствующей отрасли науки.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

**разработаны:**

- численный алгоритм вычисления экстремума критерия качества разбиения значений признака на непересекающиеся интервалы с использованием предобработки данных.

- способы использования интервальных методов в рамках информационных моделей, основанных на знаниях.

**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.

**доказано** экспериментально, что показатели устойчивости разбиения значений признака на интервалы на выборках данных из генеральной совокупности имеют приоритетное значение для выявления скрытых закономерностей по сравнению с числом непересекающихся интервалов.

Теоретическая значимость исследования обоснована тем, что:

**доказаны** положения, вносящие вклад в развитие методов предобработки данных для уменьшения комбинаторной сложности алгоритмов интеллектуального анализа данных.

**применительно к проблематике диссертации результативно** (эффективно, то есть с получением обладающих новизной результатов) использованы

методы интеллектуального анализа данных, нечеткой логики, дискретной оптимизации.

**изложены** рекомендации, позволяющие повысить эффективность обнаружения скрытых закономерностей в базах и хранилищах данных.

**раскрыты** 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.

**проведена модернизация** алгоритма вычисления границ интервалов и их числа с учетом частичного отсутствия измеренных значений признаков в описании части объектов.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что:

**определены** в задаче по диагностике состояния ультразвукового расходомера жидкости:

- минимальный набор признаков, при котором точность классификации не ниже, чем по набору исходных признаков;

- набор признаков, значения которых являются показателем неисправности расходомера;

**представлены** рекомендации по выбору правил для распознавания объектов в данных по сегментации изображений с использованием интервальных методов;

**разработаны и внедрены** (указать степень внедрения) в Ульяновской областной клинической больнице алгоритмы поиска скрытых закономерностей методами интеллектуального анализа данных с целью принятия решений при выборе тактики лечения пациентов. Найдены логические закономерности в форме полуплоскостей, пороговые значения

для которых определены как по отдельным признакам, так и по обобщенным оценкам объектов.

Оценка достоверности результатов исследования выявила:

**для экспериментальных работ** результаты обосновываются корректным применением математического аппарата, подтверждаются вычислительными экспериментами и результатами практического использования;

**теория** построена на известных проверенных научных данных, которые в полной мере согласуются с ранее опубликованными данными по теме диссертационного исследования;

**идея базируется** на обобщении передового опыта и практических исследованиях ряда российских и зарубежных ученых по теме диссертации;

**установлено**, что результаты вычислительных экспериментов согласуются с предположениями, выдвигаемыми специалистами-экспертами по предметной области;

**использованы** представительные выборочные совокупности с обоснованием подбора объектов наблюдения.

**Личный вклад** соискателя состоит в: анализе научных источников по теме диссертации, разработке численного алгоритма разделения значений признаков в описании допустимых объектов классов на непесекающиеся интервалы с использованием предобработки данных при числе интервалов, равном числу классов; разработке алгоритма для упорядочивания признаков по отношению информативности с использованием предобработки данных путем формирования матрицы близости по парам признаков; планировании экспериментов; формулировке выводов; апробации и внедрении результатов; патентовании.

В ходе защиты диссертации были высказаны следующие критические замечания:

1. 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.

2. В работе следовало бы большее внимание уделить анализу результатов зарубежных авторов (в списке литературы менее 10 источников)

3. Отсутствуют публикации в зарубежных изданиях

Соискатель Згуральская Е. Н. ответила на заданные ей в ходе заседания вопросы и дала необходимые пояснения.

На заседании 29.12.2021 диссертационный совет принял решение за решение научной задачи, имеющей важное значение для развития методов системного анализа, управления и обработки экспертной информации, в частности, повышения эффективности поиска скрытых закономерностей по базам и хранилищам данных за счет применения интервальных методов,

присудить Згуральской Е. Н. ученую степень кандидата технических наук.

При проведении тайного голосования диссертационный совет в количестве 16 человек, из них 5 докторов наук по специальности рассматриваемой диссертации, участвовавших в заседании, из 23 человек, входящих в состав совета, дополнительно введены на разовую защиту 0 человек, проголосовали: за 16, против нет, недействительных бюллетеней нет.

Защита окончена. Есть ли замечания по процедуре защиты? (Нет).

Поздравляет соискателя с успешной защитой. Благодарит членов совета и всех участников за внимание.

**Заседание объявляется закрытым.**

Председатель Совета Д212.277.04  
профессор

  
Н.Г. Ярушкина

Ученый секретарь Совета Д212.277.04,  
доцент



  
А.М. Наместников

20.01.2022