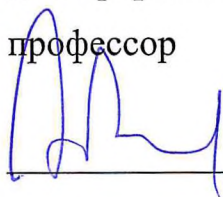


УТВЕРЖДАЮ:

Проректор по научной и инновационной
деятельности Национального исследовательского
Томского государственного университета,
доктор физико-математических наук,
профессор



Ворожиков Александр Борисович

« 24 » мая 2019 г.



ОТЗЫВ

ведущей организации на диссертацию Катаевой Алины Владимировны
на тему «Извлечение и избыточное представление закономерностей
в многомерных данных»

по специальности 05.13.17 – Теоретические основы информатики
на соискание учёной степени кандидата физико-математических наук

Актуальность исследований и степень разработанности темы.

Интеллектуальный анализ данных (Data Mining), как совокупность математических моделей, численных методов, программных средств и информационных технологий, обеспечивающих обнаружение в эмпирических данных доступной для интерпретации информации и синтез на основе этой информации ранее неизвестных, нетривиальных и практически полезных для достижения определенных целей знаний, в настоящее время является одной из актуальных направлений современных научных исследований в области искусственного интеллекта.

Современные методы интеллектуального анализа данных ориентированы на исследование многомерных и разнотипных данных с целью выявления знаний в виде закономерностей. Они основаны на мощном математическом и статистическом аппарате, корректное применение которого позволяет достичь высоких результатов.

Ассоциация является одним из наиболее известных математических инструментов (методов) интеллектуального анализа, применяемых для определения закономерностей в событиях или процессах.

В настоящее время практическое применение ассоциативных правил во многом ограничивается проблемой размерности, обусловленной тем, что число ассоциативных правил, извлекаемых современными методами анализа данных, часто достигает несколько десятков и сотен тысяч. Большое число искомым ассоциативных правил существенно усложняет их интерпретацию и снижает степень доверия пользователя к полученным результатам. Это особенно касается, когда на базе извлеченных закономерностей из накопленных медицинских данных решаются задачи медицинской диагностики.

Все вышесказанное подтверждает актуальность исследований Катаевой Алины Владимировны, диссертация которой посвящена проблеме размерности, возникающей при анализе многомерных данных. Для решения данной проблемы Катаевой А.В. исследуется когнитивный подход к анализу данных: уменьшение размера входа и выхода алгоритмов извлечения ассоциативных правил. Сокращение входных данных в диссертации Катаевой А.В. производится путем применения современных известных статистических методов по установлению наиболее информативных признаков и фильтрации ассоциативных правил по пороговым значениям основных мер значимости (поддержки и достоверности).

Для сокращения выходных данных в диссертации предложены новый метод и алгоритм построения избыточного минимаксного базиса строгих ассоциативных правил. Когнитивный подход предполагает создание базисов ассоциативных правил как «сжатых» форм представления множества искомым ассоциативных правил и является современным и актуальным направлением анализа данных.

Характеристика диссертационной работы по главам. Диссертация состоит из введения, четырех глав, заключения, списка литературы. Текст диссертации содержит 102 страницы, иллюстрирован 13 рисунками и 10 таблицами. Список литературы включает 134 источника.

Во введении описана актуальность работы, определены цель и задачи исследования.

Целью диссертационной работы Катаевой А.В. является повышение эффективности анализа данных при решении задач клинической диагностики путем установления для строгих ассоциативных правил набора выводимостей, гарантирующих сохранение поддержки, и разработка на их основе математического и программного обеспечения.

В первой главе рассматриваются особенности медицинских данных, исследуются различные виды зависимостей между данными и существующие методы их извлечения.

Во второй главе содержатся основные результаты диссертационного исследования, связанные с установлением для строгих ассоциативных правил набора выводимостей, гарантирующих сохранение поддержки найденных правил, разработкой метода и алгоритмов построения избыточного минимаксного базиса. Приведен анализ алгоритмов для решения задачи нахождения множества всех ассоциативных правил для заданной предметной области с указанием проблем, возникающих при извлечении ассоциативных правил. Были установлены свойства строгих ассоциативных правил и получен набор выводимостей, гарантирующих сохранение поддержки ассоциативных правил. На их основе разработан и теоретически обоснованы метод и алгоритм построения избыточного минимаксного базиса строгих ассоциативных правил.

Предложен оригинальный прием, позволяющий существенно снизить количество сгенерированных ассоциативных правил с помощью экспертной группировки.

Третья глава посвящена формированию набора эффективных для клинической диагностики средств снижения размерности матрицы «объект-признак», которые целесообразны для медицинских аналитических систем клинической диагностики (МАСКД). Сформирован набор средств снижения размерности матрицы «объект-признак», позволяющий уменьшить число искомых ассоциативных правил и обладающих хорошей объяснительной способностью для практикующих врачей.

В четвертой главе представлен комплекс программ, реализующий алгоритмы извлечения строгих ассоциативных правил и их «сжатого» представления в виде избыточного минимаксного базиса, снижения размерности матрицы «объект–признак» и результаты экспериментов, выполненных на реальных медицинских данных. Численные эксперименты, выполненные Катаевой А.В. на реальных базах медицинских данных (по наркозависимости, множественной лекарственной устойчивости возбудителя туберкулеза легких, сепсису), показали высокую результативность разработанных методов, алгоритмов и программ.

В заключении представлены основные результаты исследований.

Научная новизна результатов работы. Диссертация содержит новые научные результаты в области разработки математического и программного обеспечения для интеллектуального анализа данных:

- разработан и теоретически обоснован новый метод построения избыточного минимаксного базиса строгих ассоциативных правил, в отличие от существующих методов позволяющий устранять ту избыточность в минимаксном базисе, которые не способны удалять другие методы, сохраняя при этом поддержку строгих ассоциативных правил.

- разработан новый алгоритм извлечения строгих ассоциативных правил и представления их в форме избыточного минимаксного базиса. Алгоритм расширяет возможности известного алгоритма Close путем включения в него процедур по удалению из искомого множества зависимостей тех ассоциативных правил, которые распознаны как избыточные, без дополнительного обращения к анализируемому набору данных.

Достоверность основных научных положений, выводов и рекомендаций подтверждается математически корректными выводами и доказательствами теорем, представленных в работе, и экспериментами с применением методов численного анализа.

Представленные в диссертационной работе результаты широко и полно апробированы на всероссийских и международных конференциях, что также свидетельствует о достоверности и обоснованности изложенных в диссертации результатов.

Теоретическая и практическая значимость результатов работы не вызывает сомнений. Полученные и математически доказанные в ней научные результаты способствуют развитию новых технологий моделирования и обработки больших данных. Предложенный в работе метод построения неизбыточного минимаксного базиса может быть использован для дальнейшего развития раздела интеллектуального анализа данных, связанного с извлечением закономерностей в данных и устранением избыточности в их представлении.

Результаты диссертационной работы Катаевой А.В. рекомендуется использовать в практическом здравоохранении для решения различных диагностических задач. Полагаем, что возможно их применение в тех приложениях, где требуется высокая степень достоверности установленных ассоциативных правил и важна их «сжатая» форма представления, например, в информационной безопасности и анализе компьютерных сетей.

Соответствие паспорту специальности. Представленные в диссертации Катаевой А.В. результаты научных исследований полностью соответствуют паспорту научной специальности 05.13.17 – Теоретические основы информатики по области исследования «Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений» (п. 5 паспорта специальности).

Публикации. По результатам диссертационного исследования Катаевой А.В. опубликовано 12 печатных работ, из них 5 статей в журналах, включенных в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, на соискание ученой степени доктора наук, 5 – в прочих научных изданиях; получено 2 свидетельства о государственной регистрации программ для ЭВМ.

Все материалы диссертационного исследования достаточно полно отражены в опубликованных работах.

Соответствие содержания автореферата основным положениям диссертации

Автореферат логически выстроен, хорошо структурирован, правильно и в достаточной мере отражает содержание диссертационной работы. Как часть выполненной научной работы автореферат отражает компетентность автора в области проводимых исследований и хорошее владение используемыми математическими методами.

Замечания по диссертационной работе

1. В диссертации Катаевой А.В. не рассмотрен вопрос устойчивости извлеченных ассоциативных правил.

2. Отличительной особенностью диссертации Катаевой А.В. является её весьма успешное применение к анализу медицинских данных. В то же время, хотелось бы уточнить, каким требованиям должны удовлетворять данные из других предметных областей для корректного применения разработанного метода и алгоритма.

3. Учитывая высокий прикладной потенциал исследовательских результатов, в диссертационной работе отсутствуют сведения о трудоемкости разработанных алгоритмов.

4. Область применения полученных алгоритмов из задачи 3 на странице 7 могла бы быть шире, если эти результаты были бы адаптированы на случай данных.

5. В тексте диссертации на стр. 72 представлено сравнение алгоритмов по числу результирующих ассоциативных правил (таблица 4.4) и по времени работы (таблица 4.5). Из анализа таблиц следует, что увеличение объёма выборки уменьшает число сгенерированных строгих ассоциативных правил и время работы алгоритмов. Автору следовало бы указать, чем обусловлен данный парадокс.

6. Для большей наглядности описание алгоритмов следовало бы представить в виде блок-схем.

Общее заключение. Представленная на отзыв диссертация Катаевой Алины Владимировны на тему «Извлечение и избыточное представление закономерностей в многомерных данных» имеет внутреннее единство, обладает новизной, является завершённой научно-квалификационной работой, в которой на основании выполненных автором исследований содержится решение (через построение избыточного минимаксного базиса) научной задачи большого числа строгих ассоциативных правил, возникающей при работе с многомерными данными и имеющей значение для развития информационных технологий интеллектуального анализа данных.

Достоверность результатов работы не вызывает сомнений, сами результаты достаточно полно опубликованы и известны специалистам в области обработки и передачи данных. Автореферат правильно и достаточно полно отражает содержание диссертации.

Таким образом, считаем, что диссертация Алины Владимировны Катаевой «Извлечение и избыточное представление закономерностей в многомерных данных» удовлетворяет требованиям п.9 «Положения о присуждении ученых степеней», утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842 (в редакции от 01 октября 2018 г.), и удовлетворяет всем требованиям, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук по специальности 05.13.17 - Теоретические основы информатики, а её автор, Катаева Алина Владимировна, заслуживает присуждения ей учёной степени кандидата физико-математических наук по указанной специальности.

Отзыв составила профессор кафедры теории вероятностей и математической статистики федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский Томский государственный университет», доктор физико-математических наук (05.13.18 - Математическое моделирование, численные методы и комплексы программ), профессор Моисеева Светлана Петровна.

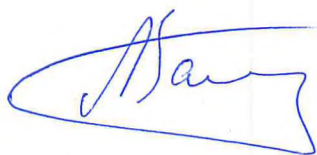
Отзыв на диссертацию А. В. Катаевой обсужден на заседании кафедры теории вероятностей и математической статистики Национального исследовательского Томского государственного университета 10 мая 2019 г., протокол № 7, одобрен на заседании научного-технического совета Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета 20 мая 2019 г., протокол № 6.

Профессор кафедры теории вероятностей и математической статистики Национального исследовательского Томского государственного университета, доктор физико-математических наук (05.13.18 – математическое моделирование, численные методы и комплексы программ), профессор



Моисеева Светлана Петровна

Директор Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета, доктор технических наук (05.13.11 – Системный анализ, управление и обработка информации), доцент



Замятин Александр Владимирович

21.05.2019

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский Томский государственный университет»

Адрес: 634050, Российская Федерация, г. Томск, пр. Ленина, 36.

Тел.: +7 (3822) 52-98-52, факс: +7 (3822) 52-95-85.

E-mail: rector@tsu.ru.

Страница в интернете: <http://www.tsu.ru>.