

ОТЗЫВ

официального оппонента доктора технических наук Аршинского Леонида Вадимовича на диссертацию Кузьмича Романа Ивановича на тему «Модифицированный метод логического анализа данных для задач классификации», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.01 – Системный анализ, управление и обработка информации (информатика, вычислительная техника и управление)

Актуальность избранной темы работы. Одной из важных задач обработки и анализа данных является поиск в них скрытых закономерностей. При этом нужно не только их выявить, но и представить (проинтерпретировать) в виде понятных для человека суждений (логических выражений). Получаемые при этом суждения должны обладать достаточным объемом, чтобы претендовать на звание закономерностей и быть обозримыми. Это сложная задача, которой посвящено много работ, но говорить об исчерпывающем ответе на все возникающие вопросы пока рано. Методы и алгоритмы ее решения разрабатываются до сих пор. Получаемые на основе метода решения могут, в частности, служить основой для разработки баз знаний соответствующих экспертных систем, а также для их обучения и самообучения.

В связи с вышеперечисленным считаю диссертационную работу Кузьмича Р.И. актуальной.

Новизна научных результатов. Особенностью представленных в диссертации результатов является разработка метода анализа данных, основанного на решении задачи классификации в ее классическом варианте:

- 1) разработке классификатора;
 - 2) его обучении на специальной выборке с настройкой соответствующих параметров;
 - 3) тестирование и эксплуатации на практических данных;
- с одновременной «вербализацией» критериев классификации в виде логических выражений.

В ходе разработки метода автором предложены следующие новые научные результаты:

1) на основе алгоритма «к-средних» разработана алгоритмическая процедура целенаправленного выбора исходных объектов для выявления закономерностей в данных;

2) разработана алгоритмическая процедура наращивания закономерностей с максимальным покрытием объектов выборки;

3) решена задача формирования закономерностей на основе постановки и решения задачи оптимизации специального вида, отличающейся наличием в целевой функции весового коэффициента покрываемого наблюдения (с возможностью захвата объектов другого класса);

4) разработана алгоритмическая процедура построения классификатора в виде композиции информативных закономерностей за счет совместного использования критерия бустинга для оценки информативности закономерностей и итеративной процедуры выбора порога информативности;

5) развиты методы логического анализа данных на основе разработанных алгоритмических процедур.

Теоретическая и практическая значимость диссертации. На основе полученных результатов автором развит и модифицирован метод логического анализа данных, основанный на использовании оптимизационных моделей для формирования информативных закономерностей, обеспечивающий сокращение количества правил в классификаторе при сохранении высокой точности. Метод реализован в программной системе поддержки принятия решений, которая позволяет эффективно решать практические задачи классификации в различных областях человеческой деятельности. Предложено решение таких практических задач как фильтрация спама, радарное сканирование ионосферы, прогнозирование осложнений инфаркта миокарда.

Степень достоверности и обоснованности научных положений, выводов и рекомендаций. Достоверность и обоснованность результатов диссертации подтверждается: исследованием существующих логических алгоритмов классификации и алгоритмов поиска информативных

закономерностей для них, корректным обоснованием постановок задач, результатами применения предложенных моделей, методов и алгоритмических процедур, сравнительным анализом по точности с существующими алгоритмами классификации на практических задачах.

Все утверждения обоснованы и подтверждены надлежащими аргументами, исходные утверждения подтверждены ссылками на источники.

Апробация. Результаты работы докладывались на конференциях и семинарах различного уровня и с представительной географией. По теме работы опубликовано 15 работ, в том числе 5 – в изданиях из перечня ВАК и программная система, зарегистрированная в Роспатенте.

Структура и содержание. Диссертационная работа состоит из введения, трех глав, заключения, списка литературы из 115 источников и 2 приложений. Основной текст диссертации содержит 120 страниц, 11 рисунков (в автореферате говорится о десяти) и 19 таблиц.

Первая глава диссертации посвящена обзору наиболее распространенных логических алгоритмов классификации, алгоритмов поиска закономерностей в массивах данных, а также обзору основных программных систем, решающих задачи обнаружения закономерностей в данных. Рассматриваются такие алгоритмы как решающие списки, решающие деревья, алгоритмы простого и взвешенного голосования правил. Обсуждаются их преимущества и недостатки. Также в главе приводится анализ программных систем для решения обсуждаемых задач. Отмечены два направления развития соответствующих программных средств: узкоспециализированные пакеты, которые направлены на небольшой круг практических задач, и программные средства широкого назначения, использующие разнообразные методы и подходы.

Вторая глава посвящена описанию основных этапов развиваемого метода логического анализа данных. В его основе лежит бинаризация признаков и сокращение числа анализируемых признаков с помощью решения задачи оптимизации специального вида с последующим преобразованием классифицирующих правил в логические выражения.

Третья глава посвящена программной реализации предложенного метода с демонстрацией его работоспособности на реальных задачах. Исследовались задачи фильтрации спама, радарного сканирования ионосферы, ряд медицинских задач. Проведено сравнение с известными методами и алгоритмами.

Автореферат соответствует содержанию диссертации.

Замечания по диссертационной работе. Несмотря на общее положительное отношение к работе, следует отметить ряд замечаний.

1). Автор сформулировал цель работы как «повышение точности классификации». Однако каких-либо интегральных оценок преимущества в точности по сравнению с известными методами не приводится. Представленные же в таблица гл. 3 сравнительные результаты говорят, что в отдельных случаях некоторые сравниваемые алгоритмы показывают лучшие результаты.

2). Используются некорректные фразы вроде как «разработка алгоритмической процедура улучшения закономерностей» (полагаю, что закономерности, если они имеются, можно только выявлять, но не улучшать), «модель оптимизации для формирования закономерностей...» (модель процесса оптимизации?).

3). Автор утверждает, что им разработана программная система поддержки принятия решений без указания круга решаемых задач. Полагаю, что автору следовало бы конкретизировать его, т.к. этот класс программных продуктов крайне разнообразен по своим методам и задачам.

4). На стр. 43 используется фраза «Компоненты вектора, называемые также признаками, могут быть как численными или номинальными, так и бинарными». Из контекста следует, что автор отождествляет числовую и номинальную шкалы. Но они ПРИНЦИПИАЛЬНО разные.

5). В работе практически не обсуждается проблема минимального объема обучающей выборки, обеспечивающей качественную работу метода, сравнение по этому критерию метода с известными алгоритмами.

6). Некоторые постановки задач недостаточно поясняются, что затрудняет их понимание; также имеются неоправданные на мой взгляд

ссылки, например, набор ссылок [7, 58, 59, 62] при упоминании инструментальной среды Delphi 2009.

Сделанные замечания не снижают ценности выполненной работы.

Заключение. Диссертация Кузьмича Романа Ивановича на тему «Модифицированный метод логического анализа данных для задач классификации» соответствует специальности 05.13.01 – Системный анализ, управление и обработка информации (информатика, вычислительная техника и управление), имеет внутреннее единство и является завершённой научно-квалификационной работой, в которой на основании выполненных автором исследований содержится решение задачи разработки метода логического анализа данных, имеющей существенное значение для теории и практики интеллектуального анализа данных.

Диссертация соответствует требованиям п.9 «Положения о порядке присуждения учёных степеней» постановления Правительства Российской Федерации от 24.09.2013 г. № 842, а ее автор Кузьмич Роман Иванович заслуживает присуждения степени кандидата технических наук.

Официальный оппонент

заведующий кафедрой «Информационные системы
и защита информации» ФГБОУ ВО

«Иркутский государственный университет
путей сообщения»

доктор технических наук, доцент  Аршинский Леонид Вадимович

25 марта 2016 года

Почтовый адрес:

664074, г. Иркутск, ул. Чернышевского, 15.

тел. 8(3952)638359,

e-mail: arsh@iriit.ru

