

## УТВЕРЖДАЮ

И.о. проректора по науке и инновациям  
Воронежского государственного технического  
университета д-р техн. наук, профессор



И.Г. Дроздов

2016 г.

## ОТЗЫВ

**ведущей организации на диссертацию Кузьмича Романа Ивановича  
«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)

### **1. Актуальность темы диссертации**

При решении практических задач классификации, в которых потери от принятия неверных решений могут быть велики, требуются алгоритмы классификации данных, которые помимо самого решения предоставляют в явном виде решающее правило, то есть выявляют знания из имеющихся данных. Это справедливо для логических алгоритмов классификации, принцип работы которых состоит в выявлении закономерностей в данных и формализации их в виде набора правил, т.е. набора закономерностей, описываемых простой логической формулой.

В предлагаемом методе логического анализа данных несомненную важность и актуальность имеют проблемы, связанные с построением оптимизационных моделей для формирования информативных закономерностей и получением интерпретируемого классификатора с высокой обобщающей способностью. Представленная диссертационная работа является актуальной, так как посвящена решению поставленных проблем путем разработки модификаций для метода логического анализа данных, позволяющих повысить интерпретируемость классификатора и качество классификации новых наблюдений.

## **2. Цель и задачи диссертационной работы**

Целью диссертационной работы Кузьмича Р.И. является повышение точности решения задач классификации и улучшение интерпретируемости классификатора, основанного на логических закономерностях. Для достижения поставленной цели решается ряд задач, связанных с модификацией метода логического анализа данных на основе разрабатываемых алгоритмических процедур.

Следует отметить, что цель и задачи диссертации сформулированы корректно и достаточно полно отражают содержание проведенных исследований.

## **3. Структура и объем диссертации**

Общий объем работы 121 страница. Диссертационная работа состоит из введения, трех глав, заключения, списка использованных источников из 115 наименований и 2 приложений.

Во **введении** обоснована актуальность работы, сформулирована цель и поставлены задачи исследования, рассмотрены вопросы теоретической и практической ценности проведенных исследований, изложены новые научные результаты, выносимые на защиту.

В **первой главе** приведен обзор наиболее распространенных логических алгоритмов классификации, алгоритмов поиска закономерностей в форме конъюнкций для них, а также анализ основных программных систем, решающих задачи классификации. Выяснено, что для классификации наиболее приемлем алгоритм, основанный на голосовании правил. Также важную роль играет построение классификатора, обладающего хорошей обобщающей способностью.

**Вторая глава** посвящена описанию основных этапов метода логического анализа данных, созданию оптимизационных моделей для формирования закономерностей и разработке алгоритмических процедур, позволяющих улучшить интерпретируемость классификатора.

Разработаны алгоритмические процедуры наращивания закономерностей, выбора базовых наблюдений для формирования закономерностей, построения классификатора как композиции информативных закономерностей.

На основе разработанных алгоритмических процедур предложены модификации для метода логического анализа данных, позволяющие повысить интерпретируемость классификатора за счет сокращения числа



правил в нем, сохраняя при этом высокую точность классификации при решении практических задач.

В **третьей главе** представлена программная реализация метода логического анализа данных и проведены экспериментальные исследования на практических задачах.

С помощью метода логического анализа данных в диссертационной работе решены следующие задачи классификации:

- выявление спама;
- классификация результатов радарного сканирования ионосферы;
- прогнозирование осложнений инфаркта миокарда (фибрилляция предсердий, фибрилляция желудочков, разрыв сердца, отек легких, летальный исход).

Метод логического анализа данных является достаточно гибким инструментом анализа данных, позволяющим учитывать специфику конкретной задачи классификации и требования заказчика (исследователя) путем целенаправленной настройки параметров метода.

Проведено сравнение по точности различных алгоритмов классификации с методом логического анализа данных. В результате данный метод показал наилучшие результаты по точности классификации, кроме того, он обладает возможностью соблюдения баланса между различными критериями сравнения алгоритмов классификации.

В **заключении** диссертационного исследования сформулированы основные результаты и выводы.

В **приложении** приведены названия полей базы данных и расшифровка их значений для задачи прогнозирования осложнений инфаркта миокарда, а также признаки с нулевой и максимальной важностью для задачи прогнозирования инфаркта миокарда, полученные в результате апробации алгоритмической процедуры получения усеченного набора признаков.

#### **4. Достоверность и обоснованность полученных результатов**

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.

## **5. Научная и практическая значимость основных положений диссертации**

Научная ценность результатов диссертационного исследования состоит в разработке и исследовании модификаций для метода логического анализа данных, основанных на создании оптимизационных моделей для формирования информативных закономерностей и алгоритмических процедур сокращения количества правил в классификаторе, что является существенным вкладом в теорию интеллектуальных технологий и представления знаний, практики их применения в системах обработки информации и интеллектуального анализа данных.

Практическая ценность результатов диссертационной работы заключается в следующем:

- разработана программная система поддержки принятия решений, которая позволяет эффективно решать практические задачи классификации в различных областях человеческой деятельности;
- в ходе выполнения работы успешно решены следующие практические задачи: классификация результатов радарного сканирования, выявление спама, прогнозирование осложнений инфаркта миокарда.

## **6. Рекомендации по возможности использования результатов и выводов диссертации**

Результаты диссертационной работы Кузьмича Р.И. рекомендуются к использованию при разработке программных систем поддержки принятия решений, эффективно решающих практические задачи классификации в различных областях человеческой деятельности (геологии, медицине, информационных технологиях и т. д.). Особенно необходимость в применении результатов диссертации возникает при решении тех практических задач классификации, где потери от принятия неверных решений могут быть велики, поэтому каждое принятое решение должно быть обосновано.

## **7. Замечания по диссертационной работе**

По представленной диссертации Кузьмича Р.И. имеются следующие замечания:



1. При проверке алгоритмической процедуры выбора базовых наблюдений для формирования закономерностей не приведено правило выбора количества центроидов для каждого класса.

2. В работе рассмотрены закономерности только в форме конъюнкций. Автор не указал, что существуют и другие формы закономерностей.

3. В параграфе 3.4 с излишней детализацией описаны известные алгоритмы.

4. На стр. 58 приведено правило изменения веса наблюдения, используемого в целевой функции (2.5) оптимизационной модели для формирования закономерностей. Отсутствуют рекомендации по выбору параметра  $N_{\max}$ , означающего максимальное количество закономерностей, которыми может покрываться наблюдение обучающей выборки в классификаторе.

5. На стр. 104 при сравнении результатов классификации по точности для метода логического анализа данных не указана используемая оптимизационная модель формирования закономерностей.

Сделанные замечания не снижают научной и практической ценности диссертации и не влияют на общую положительную оценку результатов.

## **8. Общее заключение о соответствии выполненной работы требованиям, предъявляемым к кандидатским диссертациям**

Диссертация Кузьмича Р.И. представляет собой завершённое научное исследование, содержит подходы к решению важной научной задачи, имеющей большую практическую значимость. Представленные в работе исследования обладают научной новизной и достоверностью, все полученные выводы научно обоснованы. Основные положения диссертационной работы достаточно полно освещены в научных публикациях автора. Содержание автореферата соответствует основным положениям диссертации.

Структура диссертации соответствует всем нормам и требованиям, принятым в современной науке и пунктам 3, 4, 13 паспорта специальности 05.13.01. Полученные результаты значимы, отличаются новизной и оригинальностью, вносят существенный вклад в развитие интеллектуальных технологий и представления знаний, интеллектуального анализа данных.

Все вышеизложенное даёт основание считать, что диссертация «Модифицированный метод логического анализа данных для задач классификации» Кузьмича Романа Ивановича является научно-

квалификационной работой, соответствует специальности 05.13.01 – Системный анализ, управление и обработка информации (информатика, вычислительная техника и управление) и удовлетворяет требованиям пунктов 9-14 «Положения о присуждении ученых степеней» от 24 сентября 2013 года за № 842 ВАК Минобрнауки РФ, предъявляемым к кандидатским диссертациям, а диссертант заслуживает присуждения ему ученой степени кандидата технических наук по специальности 05.13.01.

Отзыв обсужден и одобрен на заседании кафедры «Автоматизированные и вычислительные системы» ВГТУ (протокол № 9 от «29» марта 2016 г.), присутствовали 19 человек.

Заведующий кафедрой «Автоматизированные и вычислительные системы» Воронежского государственного технического университета  
Заслуженный деятель науки Российской Федерации,  
доктор технических наук, профессор

Подвальный Семен Леонидович

ФГБОУ ВО «Воронежский государственный технический университет»,  
394026, Воронежская область, г. Воронеж, Московский проспект, д. 14,  
Телефон +7(473)221-09-19, веб-сайт <http://www.vorstu.ru>, адрес электронной почты [rector@vorstu.ru](mailto:rector@vorstu.ru)

