

ОТЗЫВ

официального оппонента Пимонова Александра Григорьевича
на диссертацию Бардамовой Марины Борисовны

«Алгоритмы построения нечетких классификаторов несбалансированных данных на основе метаэвристик «гравитационный поиск» и «прыгающие лягушки», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики

Для анализа представлены диссертация общим объемом 125 страниц и автореферат на 24 страницах текста.

Актуальность темы исследования

В последние годы наблюдается беспрецедентный рост объема данных. Большие данные наряду с такими характеристиками как большой объем, большая размерность, потоковая передача обладают еще и такой особенностью, как несбалансированность, которая означает, что существует неравномерное распределение образцов в каждой категории набора данных. Однако классы с меньшим количеством образцов часто содержат более важную информацию. Примерами могут служить классификация медицинских данных, обнаружение мошенничества с кредитными картами, распознавание сетевых вторжений и спама. Если при построении классификаторов не учитывать факт несбалансированности данных по классам, то информация о классе меньшинства не будет представлена в результатах классификации. Проблема несбалансированности данных обычно решается на уровне алгоритмов или данных. В последние годы стали уделять пристальное внимание проблеме интерпретируемости и выразительности моделей машинного обучения. Эта проблема может быть решена с помощью нечетких систем, поскольку нечеткие множества были предложены для обработки плохо формализуемой информации аналогично тому, как это делает человек.

Поэтому тема диссертационного исследования Бардамовой М.Б., посвященная разработке алгоритмов построения интерпретируемых нечетких классификаторов несбалансированных данных, является актуальной.

Анализ содержания диссертационной работы

Диссертационная работа Бардамовой М.Б. состоит из введения, четырех глав основного содержания, списка литературы из 155 наименований и 4 приложений. Основная часть работы содержит 116 страниц, в том числе 14 рисунков и 39 таблиц.

Структура диссертации и ее оформление соответствуют ГОСТ Р 7.0.11-2011. Содержание автореферата соответствует основным идеям, результатам, выводам и положениям диссертации.

Во **введении** соискателем обоснована актуальность темы диссертации, обозначена решаемая проблема, поставлены цель и задачи исследования, определены объект, предмет, методы исследования, научная новизна, теоретическая

и практическая значимость работы, приведены сведения об основных положениях, выносимых на защиту, и апробации результатов диссертационного исследования, определен личный вклад автора.

Первая глава «Задача построения нечетких классификаторов несбалансированных данных» посвящена описанию задач и решений, связанных с дисбалансом данных, идентификацией структуры и оптимизацией параметров нечетких классификаторов. В рамках данной главы диссертантом проведен анализ мер качества классификации несбалансированных данных как на конечном этапе построения классификатора, так и на этапе его обучения; рассмотрены два подхода к классификации несбалансированных данных: манипулирование исходными данными и модификация алгоритмов классификации; выполнен анализ концепций проектирования нечетких систем и кратко проанализированы метаэвристические алгоритмы оптимизации. Проведенный анализ позволил автору выдвинуть гипотезу о том, что применение метаэвристических алгоритмов на различных этапах построения нечетких классификаторов вместе с изменением способа оценки качества модели позволит повысить точность разрабатываемых систем.

Во второй главе «Алгоритмы построения нечетких классификаторов несбалансированных данных» автором решаются выявленные в первой главе актуальные проблемы, возникающие при решении задачи классификации несбалансированных данных. Для решения этих проблем соискателем разработаны следующие алгоритмы:

1) алгоритм формирования структуры классификатора, позволивший создавать классификаторы с меньшим числом правил и большей средней геометрической точностью по сравнению с общеизвестными алгоритмами;

2) гибридный алгоритм оптимизации параметров нечетких термов на основе комбинации метаэвристик «гравитационный поиск» и «прыгающие лягушки», позволивший увеличить среднюю геометрическую точность классификации на исследуемых несбалансированных наборах данных;

3) алгоритм настройки весовых коэффициентов признаков, позволивший увеличить среднюю геометрическую точность классификации по сравнению с точностью, получаемой до введения весов.

Каждый из алгоритмов представлен двумя способами – пошаговым описанием на естественном языке и текстом в формате псевдокода.

В третьей главе «Исследование эффективности разработанных алгоритмов» представлены результаты экспериментального исследования предложенных алгоритмов. Соискатель использует для экспериментов 36 наборов данных из различных предметных областей и сравнивает полученные результаты с результатами работы пяти аналогичных алгоритмов построения нечетких классификаторов, отличающихся большей сложностью и другим подходом к работе с несбалансированными данными. Результаты экспериментов показали, что, несмотря на существенно более низкую сложность построенных автором класси-

фикаторов, они могут показывать статистически сопоставимые или превосходящие по точности результаты по сравнению с другими алгоритмами, выполняющие аналогичные функции.

В четвертой главе «Практическое применение результатов диссертационного исследования» диссертантом описаны результаты применения построенных нечетких классификаторов для решения задачи автоматической оценки системы свертывания крови у беременных женщин на основе данных, отличающихся несбалансированным характером.

В заключении соискателем изложены основные полученные результаты диссертационного исследования.

Научная новизна проведенных исследований и полученных результатов

Автором предложено решение актуальной в научном и прикладном отношении задачи классификации несбалансированных данных, что позволяет повысить точность предсказания класса меньшинства в наборе классифицируемых данных. Решение задачи диссертантом достигнуто за счет отличающихся научной новизной следующих результатов исследования:

1) алгоритм формирования базы правил нечеткого классификатора, отличающийся использованием метаэвристики «прыгающие лягушки» для итеративной процедуры генерации и настройки дополнительного правила для класса с наименьшим процентом правильной классификации;

2) гибридный алгоритм оптимизации параметров нечетких классификаторов, особенность которого заключается в соединении метаэвристики «гравитационный поиск» и метаэвристики «прыгающие лягушки»;

3) алгоритм настройки весовых коэффициентов признаков для базы правил нечеткого классификатора, отличающийся применением метаэвристического алгоритма из метаэвристик «гравитационный поиск» и «прыгающие лягушки» для нахождения наилучшего вектора весов признаков.

Обоснованность и достоверность научных положений, выводов и рекомендаций, сформулированных в диссертации

Выводы и основные положения, выносимые соискателем на защиту, логично вытекают из содержания диссертационной работы.

Достоверность и обоснованность исследований и полученных в диссертации результатов не вызывают сомнений и в достаточной мере подтверждаются корректным применением методов исследований, обоснованными доказательствами основных положений, практическим использованием в ОГАУЗ «Родильный дом №1» и в учебном процессе, что подтверждается соответствующими актами о внедрении. Разработанные алгоритмы были использованы при выполнении проекта № 8.9628.2017/8.9 «Теоретические основы человеко-машинных интерфейсов» в рамках государственного задания Министерства науки и высшего образования РФ и при выполнении проекта № 2.8172.2017/8.9 «Методы и модели определения уровня защищенности информационных систем» в процессе исполнения государственного задания в ФГБОУ ВО «ТУ-

СУР». Разработанные алгоритмы были использованы при выполнении проектов, поддержанных грантами РФФИ: «Методы и инструментальные средства построения самообучающихся систем, основанных на нечетких правилах» (№16-07-00034-а, 2016-2018 гг.) и «Методы построения нечетких классификаторов несбалансированных данных на основе алгоритма гравитационного поиска» (№19-37-90064-аспиранты, 2019-2021 гг.).

Необходимо отметить, что соискателем подготовлены и опубликованы пять статей в журналах, рекомендованных ВАК Минобрнауки РФ, получены четыре свидетельства об официальной регистрации программы для ЭВМ, а диссертационная работа достаточно апробирована на международных и всероссийских научно-практических конференциях.

Значимость результатов, полученных в диссертационной работе

Теоретическая значимость результатов диссертационного исследования Бардамовой М.Б. заключается в развитии технологии построения нечетких классификаторов для проведения автоматического анализа данных.

Практическая значимость полученных автором результатов заключается в достижении улучшения средней геометрической точности и точности класса меньшинства на большинстве исследуемых данных по сравнению с исходными классификаторами при меньшем ухудшении интерпретируемости по сравнению с аналогичными алгоритмами.

Замечания по диссертационной работе

По содержанию диссертационной работы имеются следующие вопросы и замечания.

1) В тексте диссертации автор приводит внушительный список российских ученых, достигших (с. 5) «*наиболее значимых результатов в изучении нечетких систем*», но в работе отсутствуют анализ и ссылки на эти результаты, а в списке литературы – библиографические описания соответствующих публикаций.

2) В тексте диссертации отсутствует аналитическое описание существующих (с. 23) и использованных (с. 44, 93) в авторских экспериментах функций принадлежности.

3) Из описания гибридного алгоритма настройки параметров нечеткого классификатора несбалансированных данных (с. 50-53) невозможно понять, из каких соображений задается константа *const*, позволяющая сгенерировать новый агент *new*.

4) Из приведенного в тексте диссертации описания (с. 50-53) алгоритма оптимизации термов не понятно, удалось ли автору «*повысить эффективность оптимизации*»? Что диссертант понимает под (с. 50) «*эффективностью оптимизации*», и как ее измерить?

5) Излагая результаты исследования гибридного алгоритма оптимизации параметров нечеткого классификатора (с. 74-78), диссертант описывает условия проведения эксперимента, параметры сравниваемых алгоритмов оптимизации,

приводит результаты эксперимента, однако не приводит описания термов, которые используются при построении классификаторов.

6) В главе 4 диссертант пишет (с. 91), что *«Результаты диссертационной работы были использованы при разработке программного обеспечения...»*. Библиографические описания разработанных программ почему-то в списке литературы диссертации отсутствуют, хотя в списке основных публикаций, приведенном в автореферате они есть. Однако составить какое-либо представление об этом программном обеспечении на основании содержания главы не представляется возможным.

7) В п. 4.2. (с. 93) автор пишет *«В нечетких правилах были использованы функции принадлежности трапецевидной формы, показавшие лучшие результаты на этапе генерации структуры с помощью АЭПК»*. Однако ни описание базы правил, ни результаты, на основе которых выбрана функции принадлежности, диссертантом не приводятся.

Содержание автореферата полностью соответствует тексту диссертационной работы. По нему можно сделать следующие замечания.

1) На с. 10 за формулой (4) следует формула (7). А где формулы (5) и (6)?

2) При описании *«гибридного алгоритма оптимизации параметров нечетких термов»* (с. 12), который в диссертации называется (с. 50) *«гибридным алгоритмом настройки параметров нечеткого классификатора несбалансированных данных»* в явном виде не указан критерий оптимизации. Это затрудняет понимание сути алгоритма без обращения к тексту диссертации.

3) При описании *«алгоритма настройки весовых коэффициентов признаков»* (с. 14) не показано, каким образом проводится нормализация векторов популяции и как обновляется коэффициент *const*.

4) Текст заключения (с. 20-21) несколько отличается от приведенного в диссертации (с. 100-102).

Сделанные замечания не снижают научной и практической ценности проведенного диссертационного исследования и не влияют на общую положительную оценку.

Заключение о соответствии диссертации критериям, установленным Положением о присуждении ученых степеней

Диссертационная работа Бардамовой М.Б. представляет собой законченную научно-квалификационную работу, написана на актуальную тему, отличается научной новизной и практической значимостью, имеет завершенный характер, выполнена на высоком научно-техническом уровне.

Автором диссертации самостоятельно получено решение научно-технической проблемы классификации несбалансированных данных, имеющей существенное значение для развития теории и практики машинного моделирования и искусственного интеллекта. Содержание диссертации соответствует пунктам 5 и 13 паспорта специальности 05.13.17 – Теоретические основы ин-

форматики. Автореферат соответствует содержанию диссертации, а ее основные положения опубликованы в научных работах.

Считаю, что диссертационная работа Бардамовой М.Б. «Алгоритмы построения нечетких классификаторов несбалансированных данных на основе метаэвристик «гравитационный поиск» и «прыгающие лягушки» удовлетворяет требованиям п. 9 «Положения о присуждении ученых степеней», предъявляемым к кандидатским диссертациям, а ее автор, Бардамова Марина Борисовна, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики.

Заведующий кафедрой прикладных информационных технологий
ФГБОУ ВО «Кузбасский государственный технический университет им. Т. Ф. Горбачёва», доктор технических наук, профессор

Пимонов Александр Григорьевич

Докторская диссертация защищена по специальности 05.13.16 – Применение вычислительной техники, математического моделирования и математических методов в научных исследованиях (по отраслям наук).

Адрес: 650000, г. Кемерово, ул. Весенняя, д. 28.

Телефон (приемная): (3842) 58-30-14, факс: (3842) 58-33-80.

E-mail: pag_vt@kuzstu.ru.

Веб-сайт: kuzstu.ru.

Подпись Пименова А. Г.

ЗАВЕРЯЮ
ученый секретарь совета
Э.В. Хеймиш
« 13 » 11 2021 г.

