

ОТЗЫВ

официального оппонента Гергет Ольги Михайловны 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

Актуальность работы

Диссертационная работа Бардамовой Марины Борисовны направлена на решение научно-технической задачи повышения точности классификации несбалансированных данных при использовании систем, основанных на нечетких продукционных правилах.

Классификация данных актуальна во многих предметных областях, таких как медицина, информационная безопасность, экономика. Построение систем автоматического анализа данных представляет собой нетривиальную задачу, так как эффективность анализа зависит не только от решающих алгоритмов, но и от особенностей накопленных данных. Нельзя не согласиться с автором работы, что нечеткие системы являются одним из наиболее подходящих инструментов при необходимости получения интерпретируемого классификатора. Решающие алгоритмы, в том числе нечеткие классификаторы, склонны переобучаться на классах, представленных наибольшим объемом экземпляров. Актуальна разработка алгоритмов, позволяющих сместить фокус обучения с наибольших классов на наименьшие, чтобы увеличить процент их правильного распознавания.

Цель работы сформулирована как повышение точности нечетких классификаторов несбалансированных данных. В качестве метрики качества автором рассматривается средняя геометрическая точность. Автор предлагает использовать алгоритмы оптимизации на различных этапах построения системы классификации. Направленность на решение задачи, имеющей существенное значение в области анализа данных, позволяет классифицировать тематику рассматриваемой диссертации как актуальную.

Структура и содержание диссертации

Диссертационная работа Бардамовой М.Б. представляет собой 126 страниц рукописи, среди которых 116 страниц занимает основная часть, включающая 39 таблиц и 14 иллюстраций. Работа содержит введение, четыре главы и заключение. В четырех приложениях приведены дополнительные таблицы с результатами эксперимента по третьему защищаемому положению, акты внедрения в учебный процесс и в медицинскую организацию города Томска, свидетельства о государственной регистрации программ. Список источников насчитывает 155 наименований.

Автореферат Бардамовой М.Б. занимает 24 страницы текста. Содержание автореферата соответствует диссертации. Оформление диссертации и автореферата соответствует требованиям ГОСТ Р 7.0.11-2011.

Во введении диссертантом обоснована актуальность темы, определена цель и перечислены задачи работы. Приведены основные результаты проведенного

исследования и изложены научная новизна, теоретическая и практическая значимость полученных результатов, сформулированы положения, выносимые на защиту.

В первой главе представлен полноценный обзор научных работ, соответствующих теме исследования. Автором проанализирована специфика работы с несбалансированными данными, изложены основные способы работы с такими данными и их недостатки. Далее приведено подробное описание процесса построения нечетких систем, в частности нечетких классификаторов. Охарактеризованы два метаэвристических алгоритма оптимизации, используемых для оптимизации нечетких классификаторов.

Вторая глава посвящена детальному описанию разработанных алгоритмов. Каждый из них предназначен для отдельного этапа построения нечеткого классификатора – формирования структуры, оптимизации параметров термов, настройки весовых коэффициентов признаков.

В третьей главе содержится экспериментальная часть исследования. Автором кратко описываются использованные для тестирования наборы данных, далее последовательно приводятся итоги построения классификаторов с каждым из разработанных алгоритмов и их сравнение с результатами аналогичных алгоритмов.

В четвертой главе изложена актуальность использования нечетких классификаторов в задаче оценки состояния свертывающей системы крови у беременных женщин. Стоит отметить грамотное владение медицинской терминологией без наличия соответствующего образования. Далее описаны имеющиеся входные данные для построения системы классификации и процесс построения и обучения классификаторов. Приводятся результаты обучения, валидации и тестирования созданных моделей. Приложенный акт внедрения подтверждает, что разработанные нечеткие классификаторы, используемые в составе программного обеспечения для оценки системы коагуляции, позволили упростить труд сотрудников лаборатории медицинского учреждения города Томска.

В заключении сформулированы выводы по проделанной работе и перечислены основные результаты исследования.

Достоверность и новизна полученных результатов

Научная новизна диссертационной работы заключается в создании трех новых инструментов оптимизации нечетких классификаторов несбалансированных данных, которые позволяют улучшать и среднюю геометрическую точность построенных классификаторов, и точность целевого – положительного – класса:

- алгоритма формирования базы правил, добавляющего правила для классов с наименьшей точностью метаэвристикой "прыгающие лягушки";
- алгоритма оптимизации параметров термов, состоящего из метаэвристик «гравитационный поиск» и «прыгающие лягушки»;
- алгоритма настройки весов признаков на основе гибрида из упомянутых выше метаэвристик.

Предложена оригинальная фитнес-функция, сочетающая среднюю геометрическую и общую точность через коэффициент приоритета. Доказано, что при наличии дисбаланса экземпляров такая функция позволяет достигать лучшего

соотношения между процентами правильной классификации двух классов, чем отдельные и аналогичные метрики качества.

Впервые к классификатору, основанному на правилах и функционирующему в офлайн-режиме, применен алгоритм взвешивания признаков. Предложен оригинальный способ расчета нечеткого вывода при наличии весов признаков.

Достоверность и обоснованность научных результатов в достаточной мере подтверждается грамотно проведенными экспериментами и корректными сравнениями с аналогами, значительным числом публикаций в различных российских и международных журналах, представлением результатов на конференциях международного и всероссийского уровня, четырьмя свидетельствами о государственной регистрации программ.

Теоретическая и практическая значимость полученных соискателем результатов

Теоретическая значимость результатов, полученных в процессе выполнения диссертационного исследования автором работы, состоит в развитии теоретических основ создания классификаторов, основанных на нечетких правилах, при анализе несбалансированных данных для различных предметных областей. Проведен анализ метрик, подходящих для оценки качества данного типа классификаторов в условиях наличия дисбаланса.

Практическая значимость научной работы подтверждается положительной оценкой и использованием разработанных алгоритмов в деятельности ОГАУЗ «Родильный дом №1» города Томска. Результаты диссертационной работы нашли применение при выполнении двух проектов РФФИ: №16-07-00034-а (2016-2018) и №19-37-90064-аспиранты (2019-2021); в проекте № 8.9628.2017/8.9 в рамках государственного задания Министерства науки и высшего образования РФ (2017–2019); в проекте № 2.8172.2017/8.9 в процессе исполнения государственного задания в ФГБОУ ВО «ГУСУР» (2017–2019).

Результаты, полученные в диссертации, могут быть использованы для построения систем автоматического анализа данных в рамках любых предметных областей.

Замечания по диссертационной работе

Имеются следующие вопросы и замечания по содержанию диссертационной работы.

1) Не приведены сравнения точности классификации на несбалансированных данных с классификаторами других типов, таких как нейронные сети, деревья решений и т.д. Данное сравнение позволило бы оценить точность предлагаемого автором решения среди инструментов анализа данных.

2) Не указаны доводы по выбору в качестве целевой функции средней геометрической точности, хотя для работы с несбалансированными данными существуют и другие метрики, например, F-мера.

3) На стр. 4 указано, что «генерация данных затруднительна при количестве классов, большем двух». Само по себе утверждение "затруднительно", весьма спорно, так как в настоящее время алгоритмы SMOTE, ADASYN и их модификации при выполнении добавления данных для множества классов используются вместе с подходом one-vs-rest.

4) Не приведено сравнение и отсутствует обоснование выбора вида функций принадлежности. На стр. 66 указано, что применялись функции Гаусса, а на рисунке 4.2 в главе 4 показано, что в рамках внедрения использовались термы в форме трапеций.

Приведенные замечания не снижают общей положительной оценки диссертационной работы и не ставят под сомнение значимость полученных результатов.

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

Диссертационная работа Бардамовой Марины Борисовны является самостоятельно выполненной завершенной научно-квалификационной работой, характеризующейся актуальностью, новизной и достоверностью результатов, обоснованностью выводов. Диссертация и автореферат написаны грамотным и понятным языком, имеют логичную структуру. Основные положения диссертационного исследования изложены в достаточном количестве публикаций, среди которых 6 опубликованы в научных изданиях из перечня ВАК и 10 проиндексировано международными базами Scopus и Web of Science. Защищаемые положения соответствуют пунктам 5 и 13 паспорта специальности 05.13.17 - Теоретические основы информатики.

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

Официальный оппонент:

Профессор отделения информационных технологий ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», доктор технических наук, доцент

Гергет Ольга Михайловна


16.11.2021

Докторская диссертация защищена по специальности 05.13.17 – Теоретические основы информатики.

Адрес: 634050, г. Томск, пр. Ленина, д. 30.

Телефон (приемная): (3822) 70-17-79.

E-mail: gerget@tpu.ru.

Веб-сайт: tpu.ru.

Подпись Гергет О.М. удостоверяю

Ученый секретарь





Кулинич Е.А.