

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. В.Г. ШУХОВА»**
(БГТУ им. В.Г. Шухова)

Костюкова ул., д.46, Белгород, 308012, тел.(4722)54-20-87, факс (4722)55-71-
39 E-mail: rector@intbel.ru, <http://www.bstu.ru>

«УТВЕРЖДАЮ»



И.о. ректора Федерального государственного
бюджетного образовательного учреждения
высшего образования «Белгородский
государственный технологический
университет им. В.Г. Шухова», доктор
технических наук, профессор

Е. И. Евтушенко

«19» ноября 2021 года

ОТЗЫВ

ведущей организации на диссертацию
Бардамовой Марины Борисовны
на тему: «Алгоритмы построения нечетких классификаторов
несбалансированных данных на основе метаэвристик «гравитационный
поиск» и «прыгающие лягушки»,
по специальности 05.13.17 - Теоретические основы информатики
на соискание ученой степени кандидата технических наук

Актуальность темы диссертационного исследования

Диссертация Бардамовой Марины Борисовны посвящена проблеме построения интерпретируемых нечетких классификаторов несбалансированных данных. Классификация – одна из важнейших задач машинного обучения и распознавания образов, для решения которой разработано достаточно большое количество методов и алгоритмов, неплохо справляющихся с указанной задачей. Однако эффективность классификации данных с несбалансированным распределением классов существенно ухудшается для большинства стандартных алгоритмов обучения, которые предполагают относительно сбалансированное

распределение классов. Проблема в том, что экземпляры, принадлежащие к редким классам, чаще классифицируются ошибочно, чем образцы, принадлежащие к классам большинства. Поскольку существует ряд проблем, таких как медицинская и техническая диагностика, для которых правильная классификация образцов из редких классов имеет более существенное значение по сравнению с правильной классификацией образцов классов большинства, актуальной является задача разработки методов и средств решения проблемы дисбаланса в данных при машинном обучении. Классификаторы предсказывают метки классов с помощью обучающего набора данных и алгоритма обучения. При разработке классификаторов акцент делается как на их точности, так и на интерпретируемости. Преимущество нечеткой логики состоит в том, что она позволяет понятным для человека образом описывать решаемую задачу в терминах отношений между переменными.

Исходя из выше сказанного актуальность данного исследования не вызывает сомнений.

Степень обоснованности научных положений, выводов и рекомендаций, сформулированных в диссертации

Основные научные положения и выводы, представленные в диссертационной работе, в достаточной степени обоснованы и интерпретированы. Автор корректно использует научные методы обоснования полученных результатов, выводов и рекомендаций. В работе проанализированы современные подходы и методы, применяемые в исследованиях дисбаланса в наборах классифицируемых данных. Список использованной литературы содержит 155 наименования. Выводы логически вытекают из материалов исследований и в полном объеме отражают поставленные задачи.

Достоверность и новизна, полученных результатов

Все полученные в работе результаты обоснованы и достоверны. Идея работы базируется на анализе практики и обобщении передового опыта решения проблем дисбаланса в классифицируемых данных. Степень достоверности результатов обеспечивается использованием современных математических методов и статистических критериев, а также подтверждаются актом внедрения результатов диссертационной работы.

Разработанные алгоритмы и программы для решения задач классификации несбалансированных данных на основе нечетких классификаторов исследованы на примере 36 наборов несбалансированных данных из репозитория KEEL.

Научную новизну диссертационной работы определяют следующие результаты исследования, полученные лично соискателем.

1. Предложен алгоритм генерации базы правил нечеткого классификатора несбалансированных данных, отличительной особенностью которого является применение метаэвристики "прыгающие лягушки" для создания нового правила, описывающего класс с наименьшей долей правильной классификации на предыдущей итерации.

Алгоритм позволяет создавать нечеткий классификатор, у которого средняя геометрическая точность либо выше, либо статистически неразличима с точностью алгоритмов-аналогов при существенно меньшем количестве правил.

2. Разработан новый гибридный алгоритм для оптимизации параметров функций принадлежности нечетких классификаторов несбалансированных данных; особенностью алгоритма является использование метаэвристики «прыгающие лягушки» на этапе локального поиска в сочетании с метаэвристикой «гравитационный поиск».

Алгоритм позволил существенно увеличить среднюю геометрическую точность классификации на исследуемых несбалансированных наборах данных по сравнению с точностью до оптимизации. Нечеткие классификаторы, созданные на основе указанного алгоритма, имеют среднюю геометрическую точность либо выше, либо статистически неразличимую с точностью алгоритмов-аналогов при существенно меньшем количестве правил.

3. Разработан новый алгоритм настройки весовых коэффициентов признаков для классификации несбалансированных данных.

Алгоритм позволил увеличить среднюю геометрическую точность классификации относительно точности до введения весов. При существенно меньшем количестве используемых правил алгоритм позволил продемонстрировать сопоставимую точность или большую точность по сравнению с алгоритмами-аналогами.

Теоретическая и практическая значимость полученных автором результатов

Теоретическая значимость результатов диссертационной работы заключается в развитии технологии построения нечетких классификаторов для анализа несбалансированных данных. Разработанные алгоритмы расширяют перечень возможных инструментов для оптимизации моделей классификации. Предложенная в первом положении метрика для оценки качества построенного классификатора полезна для работы с несбалансированными данными при использовании любых решающих алгоритмов.

Практическая значимость итогов исследования подтверждается внедрением в рабочий процесс ОГАУЗ «Родильный дом №1» города Томска. Разработанные алгоритмы 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.; «Методы построения нечетких классификаторов несбалансированных данных на основе алгоритма гравитационного поиска» (№19-37-90064-аспиранты), 2019-2021 гг., а также при выполнении государственных заданий Министерства образования и науки Российской Федерации на 2017–2019 гг., проект № 2.8172.2017/БЧ «Методы и модели определения уровня защищенности информационных систем»; 2017-2019 гг., проект № 8.9628.2017/8.9 «Теоретические основы человеко-машинных интерфейсов».

Результаты, полученные в диссертации, могут быть использованы в научных и проектных организациях, занимающихся исследованиями в области классификации несбалансированных данных.

Представляется, что результаты, полученные автором диссертации, послужат хорошей почвой для анализа последующих исследователей классификации несбалансированных данных и помогут верификации моделей обучения классификаторов таких данных.

Оценка содержания диссертации, её завершенность

Диссертационная работа состоит из введения, четырех глав, выводов к каждой главе, заключения, списка использованной литературы, состоящего из 155 источников и четырех приложений.

Первая глава диссертационного исследования посвящена анализу проблем машинного обучения моделей на несбалансированных данных. Соискатель анализирует предметные области, в которых наиболее часто встречаются задачи классификации несбалансированных данных,

рассматривает особенности классификации несбалансированных данных и методы решения проблемы дисбаланса данных (с. 11-22). В этой же главе приведен обзор существующих типов нечетких систем, этапов и методов построения нечетких классификаторов (с. 22-43). По результатам проведенного анализа выявлена необходимость совершенствования имеющихся методов и алгоритмов классификации несбалансированных данных.

Во второй главе диссертации представлены описания алгоритма генерации базы правил нечеткого классификатора на основе метаэвристики «прыгающие лягушки» (с. 47-50), гибридного алгоритма тонкой настройки параметров нечеткого классификатора несбалансированных данных (с. 50-54), а также алгоритма настройки весовых коэффициентов признаков (с. 54-58). Сильной стороной исследования является наличие возможности получить информацию о численно выраженной важности признаков при формировании выходной метки класса.

Третья глава диссертационного исследования посвящена экспериментальному исследованию разработанных алгоритмов. Для этих целей соискатель использовал общедоступный репозиторий KEEL, на 15 несбалансированных наборах данных которого были исследованы восемь метрик эффективности классификаторов (59-64), а также проверена работоспособность алгоритма генерации базы правил нечеткого классификатора несбалансированных данных, алгоритма оптимизации параметров функций принадлежности и алгоритма настройки весов признаков нечеткого классификатора (64-88). Несомненным достоинством исследования является применение непараметрических критериев Уилкоксона и Фридмана для сравнения полученных результатов с общеизвестными алгоритмами классификации несбалансированных данных.

В четвертой главе описано практическое применение результатов диссертационного исследования для выявления патологий свертывания крови у беременных женщин.

В заключении представлены выводы по проделанной работе.

Список литературы содержит достаточную библиографию по тематике диссертации.

Защищаемые положения соответствуют полученным выводам.

Автореферат соответствует основным положениям диссертации.

Замечания и недостатки диссертационной работы

1. Нечеткий вывод, представленный на стр. 45 диссертации и стр. 9 автореферата, сформулирован без должного обоснования. Дефаззификация в традиционной нечеткой системе типа Мамдани имеет несколько другой смысл и выполняется по другому алгоритму.

2. Недостаточно полно раскрыто решение проблемы дисбаланса классов при наличии нескольких классов, хотя в четвертой главе эта проблема решается.

3. «Качество решения», «метрика качества», «фитнес-функция», «эффективность алгоритма», «эффективность оптимизации», просто «эффективность» – термины, которые встречаются практически на каждой странице текста диссертации. Отсутствие рабочего определения этих терминов затрудняет понимание сути исследования.

4. В обзорной главе для решения проблемы дисбаланса данных упоминается применение различных классификаторов: деревья решений, машина опорных векторов, нейронные сети, ансамбли классификаторов. Однако в основной части диссертационного исследования отсутствует сравнение с указанными классификаторами.

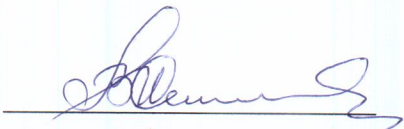
5. На стр. 23 автор диссертации пишет, что для классификации стоит воспользоваться нечеткой системой типа Мамдани, для аппроксимации – Такаги-Сугено. Утверждение является спорным, так как для классификации применяются и системы типа Такаги-Сугено.

Заключение

В целом диссертация представляет собой законченное научное исследование, выполненное на высоком научном уровне. Основные положения диссертации являются новыми и изложены в 28 научных работах, в том числе в шести работах, опубликованных в журналах, входящих в список ВАК РФ, и десяти работах, опубликованных в журналах, индексируемых Scopus и/или Web of Science. Получены четыре свидетельства о государственной регистрации программ для ЭВМ.


Таким образом, диссертация Бардамовой М.Б. является научно-квалификационной работой, в которой содержится решение задачи классификации несбалансированных данных на основе интерпретируемых нечетких систем, имеющей значение для развития теории и практики искусственного интеллекта и машинного обучения, что соответствует требованиям п. 9 «Положения о порядке присуждения ученых степеней»,

утвержденного постановлением Правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор заслуживает присуждения искомой ученой степени.

Отзыв составил  В.Г. Синюк, кандидат технических наук, профессор кафедры программного обеспечения вычислительной техники и автоматизированных систем ФГБОУ ВО «Белгородский государственный технологический университет им. В.Г. Шухова».

Отзыв рассмотрен и утвержден на заседании кафедры программного обеспечения вычислительной техники и автоматизированных систем ФГБОУ ВО «Белгородский государственный технологический университет им. В.Г. Шухова» протокол № 3 от «28» октября 2021 г.

Заведующий кафедрой,
к.т.н., доцент


В.М. Поляков

Подпись Синюка Василия Гоигорьевича и Полякова Владимира Михайловича «**заверено**»

Проректор по научной и инновационной деятельности,
д.п.н., профессор  Т.М. Давыденко

19.11.2021 г.

