

ОТЗЫВ

официального оппонента Крутикова Владимира Николаевича
на диссертацию Масича Игоря Сергеевича на тему
«Метод оптимальных логических решающих правил
для классификации объектов»,
представленную на соискание ученой степени доктора технических наук
специальности 05.13.01 - Системный анализ, управление
и обработка информации

1. Актуальность темы исследования

В настоящее время при решении практических задач классификации часто требуется знать, почему новое наблюдение принадлежит определенному классу, насколько это решение обоснованно. В таких случаях нужен метод классификации данных, который помимо самого решения представит в явном виде решающее правило, то есть выявляет знания из имеющихся данных. Это справедливо для предлагаемого метода, принцип работы которого состоит в выявлении закономерностей в данных и формализации их в виде набора правил, то есть набора закономерностей, описываемых простой логической формулой. В предлагаемом методе 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 посвящена решению поставленных проблем путем разработки метода оптимальных логических решающих правил, позволяющего повысить интерпретируемость классификатора и качество классификации новых наблюдений.

2. Достоверность полученных результатов, положений и выводов

Автором проведен системный анализ предметной области, изучены существующие подходы к решению задач классификации объектов с помощью логических правил. Все положения, выводы и практические результаты обоснованы достаточным количеством численных экспериментов, как на типовых тестовых задачах, так и на реальных данных. Методика сравнительной оценки эффективности алгоритмов подробно описана в диссертации. Результаты и выводы не противоречат положениям и выводам, известным из работ других российских и зарубежных исследователей.

Результаты диссертационного исследования опубликованы в 37 печатных работах автора в рецензируемых научных изданиях, рекомендованных ВАК РФ для опубликования результатов диссертаций. Работа прошла всестороннюю апробацию на конференциях.

3. Основные научные результаты

Диссертантом решена актуальная проблема создания логических алгоритмов классификации повышенной точности. Новые алгоритмы основаны на формировании оптимальных логических решающих правил – закономерностей в виде конъюнкций значений признаков наблюдений.

Повышение точности достигается за счет новых моделей оптимизации, в соответствие с которыми формируются закономерности, и нового алгоритма комбинаторной оптимизации, позволяющего находить оптимальные закономерности. Рассматриваемые алгоритмы логической классификации обладают важным свойством — они могут быть интерпретированы пользователем, а результат классификации может быть обоснован.

Разработанный диссертантом метод оптимальных логических решающих правил является обобщением проведенных автором и другими исследователями работ по изучению применимости булевых выражений и псевдобулевой оптимизации к решению задачи формирования обобщающих закономерностей и классификации объектов на их основе. Метод позволяет комбинировать различные виды закономерностей (так называемые сильные охватывающие и первичные закономерности), что дает повышение точности и снижение числа нераспознанных случаев. Метод включает в себя также отбор закономерностей, что всё вместе позволяет получать компактный классификатор, состоящий из небольшого числа наиболее информативных правил. Проведенные автором эксперименты на практических данных показывают конкурентоспособность метода среди наиболее популярных методов машинного обучения и превосходство над известными методами классификации, основанными на правилах.

4. Научная новизна, ценность результатов и выводов для теории и практики

Результаты, полученные в диссертации, являются новыми и получены лично автором. Все алгоритмы и модели, предложенные автором в главах 2, 3, 4 и 5 диссертации, являются новыми, имеющими существенные отличия от известных моделей и алгоритмов для рассматриваемых задач.

В главе 2 предложена новая оптимизационная модель для назначения порогов при бинаризации вещественных признаков. При применении предлагаемой модели обеспечивается нахождение оптимального размещения порогов вещественных признаков, при котором наблюдения разных классов наиболее различимы, что позволяет бинаризовать вещественные признаки с помощью наименьшего числа порогов практически без потери информации. Модель оптимизации для поиска сильных охватывающих закономерностей, описанная в главе 3, предлагается впервые. Автором показано, что формируемые с помощью предлагаемой модели закономерности максимальны по отношению доказательности и избирательности. В главе 4 предлагается новая оптимизационная модель для отбора закономерностей на основе максимизации разделяющей полосы. Также впервые предложен подход к решению задачи распознавания, заключающийся в совместном использовании различных видов закономерностей, максимальных по различным отношениям (доказательности, избирательности и простоты), и алгоритмы поиска пары закономерностей. В пятой главе предлагается новый алгоритм псевдобулевой оптимизации, основанный на схеме метода ветвей и границ и поиске по граничным точкам допустимой области, реализующий свойства рассматриваемого класса задач. Предлагаемый алгоритм обеспечивает нахождение оптимальных решений в соответствии с предлагаемыми в других главах оптимизационными моделями.

Главным новым научным результатом следует считать собственно метод поиска оптимальных логических решающих правил, повышающий эффективность решения задач классификации с требованием доказательности результата и интерпретируемости классификатора.

Указанный метод обеспечивает поддержку принятия решений для задач распознавания, в которых требуется объяснение результата и прозрачность решения. Это свойство продемонстрировано автором на реальных задачах, например, на задаче прогнозирования осложнений заболеваний.

Возможность достаточно точно решать задачу классификации с помощью набора логических правил позволило внедрить разработанный метод в процесс проверки качества электрорадиоизделий для космического применения и повысить качество продукции на выходе цикла испытаний, проводимых в АО «Испытательный технический центр - НПО ИМ».

5. Замечания по диссертационной работе

1) Всё внимание в работе уделено решению задачи классификации с двумя классами. Не рассмотрен вопрос решения задачи классификации при числе классов более двух.

2) При решении задачи поиска оптимальных закономерностей не проведено сравнение с универсальными алгоритмами оптимизации, такими как генетический алгоритм или алгоритм локального поиска.

3) Результаты работы метода, приведенные в некоторых таблицах, описаны недостаточно ясно. Не везде четко указано, какие параметры метода настраиваются в конкретном эксперименте.

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

5) 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 таблицах сравнительные результаты говорят, что в отдельных случаях сравниваемые алгоритмы показывают лучшие результаты.

6) В работе описывается выявление и использование закономерностей только в виде конъюнкций и не рассматривается применение закономерностей других типов.

6. Заключение по диссертационной работе

Несмотря на имеющиеся замечания, считаю, что диссертационная работа на соискание ученой степени доктора технических наук Масича И.С. является завершенной научно-квалификационной работой на актуальную тему, выполненной на высоком уровне. Полученные автором результаты достоверны, выводы и заключения обоснованы.

Исследования диссертанта позволяют решить проблему создания более точных логических алгоритмов классификации, предназначенных для поддержки принятия решений при классификации с возможностью интерпретации и обоснования результата на основе логических правил.

Результаты и выводы, представленные в диссертации, являются научным достижением автора, которое вносит существенный вклад в развитие методов машинного обучения и интеллектуального анализа данных.

Диссертационная работа представлена на 261 странице машинописного текста. Состоит из введения, шести глав, заключения, списка литературных источников из 250 наименований и приложения. Диссертация обладает внутренним единством, стиль изложения материала ясный, качество оформления работы хорошее. Диссертация соответствует п.4 паспорта специальности 05.13.01.

Автореферат отражает основное содержание диссертации. В автореферате приведены все основные этапы работы, полученные результаты и выводы. Оформление автореферата и диссертации соответствует требованиям ВАК РФ.

Представленная диссертационная работа полностью отвечает требованиям Положения о порядке присуждения ученых степеней, предъявляемым к докторским диссертациям, а ее автор Масич Игорь Сергеевич заслуживает присуждения ученой степени доктора технических наук по специальности 05.13.01 - Системный анализ, управление и обработка информации (космические и информационные технологии).

Официальный оппонент,
профессор кафедры прикладной математики
ФГБОУ ВО «Кемеровский государственный университет»,
доктор технических наук, доцент



Крутиков Владимир Николаевич

Адрес организации: 650036, г. Кемерово, ул. Терешковой, 40,
Тел.: (384-2) 54-25-09, e-mail: krutikovvn@rambler.ru

Подпись Крутикова Владимира Николаевича заверяю

зав. канцелярией
Иль. В. Кузнецов

