

ОТЗЫВ

официального оппонента Дулесова Александра Сергеевича
на диссертацию Масича Игоря Сергеевича
на тему «Метод оптимальных логических решающих правил
для классификации объектов»

по специальности 05.13.01 - системный анализ, управление и обработка
информации на соискание ученой степени доктора технических наук

Классификация объектов – наиболее распространенные технологии реализации задач машинного обучения с учителем. Для их решения известно немало различных методов, в их числе присутствуют те, которые позволяют получать конечные результаты с довольно незначительной ошибкой.

Тем не менее, в ряде случаев, относящихся к решению практических задач, востребованы методы классификации, которые требуют не только точного решения, но и выявления возможностей обоснования результатов. Такие требования обусловлены необходимостью строгой регламентации процесса решения в областях, где цена ошибки и мера ответственности весьма высоки. В качестве примера можно привести задачи медицинской диагностики, предсказания месторождений полезных ископаемых, оценивания кредитоспособности заемщиков, покупательной способности граждан и др. Их решение будет иметь наибольшую ценность, наделенную наличием алгоритмов классификации, основанных на правилах. К этим, алгоритмам предъявляются вышеупомянутые требования. Тем не менее, остаются неразрешенными вопросы повышения точности или снижения высоких вычислительных затрат.

В этой связи диссертация посвящена актуальной проблеме создания метода для нахождения логических правил с наилучшей обобщающей способностью и его применения при решении практических задач классификации объектов при условии соблюдения высокой производительности и достижения низкого уровня ошибки.

Структура работы

Диссертация с приложениями изложена на 261 странице, основной текст состоит из шести глав (разделов).

Следуя структуре и содержанию рукописи диссертации, далее представлено обоснование содержательности работы, как результата научной деятельности соискателя.

Во введении имеются необходимые краткие материалы, отражающие значимость, функциональное и структурное содержание выполненных исследований.

Тем не менее было бы целесообразным отметить актуальность работы в области перспектив развития таких систем как Big Data и Data Mining, которые востребованы Национальной программой «Цифровая экономика Российской Федерации».

В первой главе проведен обзор спектра существующих методов классификации. Автором рассмотрен широкий спектр методов, основанных на правилах и обоснованы проблемы реализации существующих методов. Обозначены и сформулированы требования к новому методу решения задач классификации и сформулирована идея нового метода. Данная идея имеет четкое обоснование, связанное с выявлением логических закономерностей, которые можно построить на основе комбинации оптимизационных моделей с применением алгоритмов оптимизации, реализующих свойства классов задач.

В целом обзор основателен, прослеживается понимание автором

фундаментальных принципов и свойств, присущих построению классификаторов, основанных на правилах.

Вторая глава посвящена вопросу построения логических правил по данным с разнотипными признаками. Последовательно, логично и с наличием примеров, автор излагает терминологию и предлагаемые обозначения, уделяя серьезное внимание проблеме бинаризации количественных признаков.

Рассматривая согласованность булева отображения (стр. 50), дискриминант количественного признака a создается для каждого значения как средняя величина, что в некоторых случаях расчетов является упрощенным решением. На стр. 51 рисунки 2.2б и 2.2в полностью совпадают.

Кроме этого, автор рассмотрел и выявил недостатки известного способа выбора порогов при бинаризации действительных признаков в решении задач оптимизации. Далее он предлагает новую модель оптимизации для нахождения оптимального назначения порогов количественных признаков. Предлагаемая модель предназначена для вычисления таких порогов количественных признаков, при которых наблюдения разных классов наиболее различимы. Применение такого подхода ведет к повышению качества отдельных закономерностей и, следовательно, к повышению точности классификатора.

На стр. 63 автор предлагает определять величину порога как среднее значение между двумя выборками. Однако такое решение справедливо для выборки с малой разницей между соседними значениями выборки.

В третьей главе предлагаются к рассмотрению модели оптимизации для выявления закономерностей. Проведен подробный анализ существующих способов выявления закономерностей в данных, подкреплённый математическим описанием и примерами. Исследованы свойства моделей оптимизации для нахождения закономерностей. Показано, что существующие алгоритмы не гарантируют получения сильных закономерностей с максимальным покрытием. Устранение данного недостатка предлагается осуществить путем разработки новой модели оптимизации для нахождения сильных охватывающих закономерностей в данных, что и представлено в п.3.4. Такая математическая модель задачи оптимизации включает в себя целевую строго монотонную функцию при отсутствии множества постоянства целевой функции. Важным качеством построенной модели можно считать следующий факт: крайние точки допустимой области соответствуют сильным охватывающим закономерностям.

На стр. 68 имеется опечатка (третий абзац, последняя строка) «... которая **не** пересекается с K^* ». Было бы целесообразным представить пример в п.3.4, что позволило бы подтвердить важность предлагаемой модели.

В четвертой главе рассматриваются вопросы построения решающего правила для классификации объектов с помощью логических закономерностей.

Рассматривается важная задача, связанная с рассмотрением способов отбора закономерностей для формирования решающего правила. В преддверии выбора решающего правила автор дает обоснование вопросам анализа наблюдений, которые можно классифицировать. Далее предлагаются к рассмотрению способы отбора закономерностей:

1. Путем выделения подмножества закономерностей для покрытия всех объектов обучающей выборки. Обоснованно предлагается включить в данный способ задачу оптимизации, которая относилась бы к задаче условной псевдобулевой оптимизации и в которой целевая функция и функции ограничений являлись бы унимодальными монотонными псевдобулевыми функциями. В этом случае можно говорить о возможном поиске решения.

2. Посредством отбора закономерностей, которые при совместном использовании увеличат разделяющую способность решающего правила. Вполне

обоснованно при формировании решающего правила предлагается критерий - ширина «разделяющей полосы».

Соискатель показал, что на основе предлагаемых способов можно значительно снизить число логических закономерностей и, тем самым, упростить решающее правило, практически не снижая точность распознавания. Становится очевидным: решающее правило будет получено прозрачным, что необходимо для поддержки принятия решений при распознавании.

Дополнительно к вышеизложенному отметим следующее:

1). На стр. 91 (п.4.2) предложена решающая функция $D(a)$ как среднее расстояние между двумя противоположными закономерностями (ширина «разделяющей полосы»). Фактически она определяется исходя из нормальности закона распределения случайного наблюдения. Возможно ли использовать формулы, которые ориентированы на получение расстояния как медиану или моду?

2). На стр. 92 представлена модель минимизации числа используемых в решающем правиле закономерностей. Было бы уместным показать эту модель в графической интерпретации с целью обозначения области допустимых решений.

3). В табл. 4.1 даны результаты. Отображение результатов набора закономерностей (при его достаточности) графически, могло бы наглядно обосновать роль разных задач распознавания в формировании набора закономерностей.

4). В п.4.4 в алгоритм поиска пары закономерностей предлагается добавить дополнительные ограничения. Не создаст ли такое решение затруднений в поиске оптимального решения, поскольку ограничения могут привести к созданию дополнительных областей допустимых решений и появлению множества экстремальных точек.

Пятая глава посвящена алгоритмам оптимизации для решения тех задач, которые рассмотрены в предыдущих главах.

Подробно исследованы свойства имеющихся классов задач, которые являются задачами условной псевдоболевой оптимизации. Описывается новый алгоритм оптимизации, базирующийся на схеме ветвей и границ, который лежит в основе предлагаемого в диссертации метода и предназначен для нахождения оптимальных логических правил. Имеются необходимые поясняющие примеры.

Следует признать: предлагаемый автором алгоритм повышает эффективность решения задач условной псевдоболевой оптимизации с алгоритмически заданными функциями. Не исключено, что разработанный алгоритм оптимизации применим для любых задач условной псевдоболевой оптимизации, возникающих в процессе управления сложными техническими объектами, а также в областях экономики. Алгоритм апробирован при решении задачи оптимизации загрузки технологического оборудования предприятия.

Стремление автора получать быстрое, наиболее точное решение направлено на создание нового алгоритма условной псевдоболевой оптимизации на основе схемы метода ветвей и границ и поиска среди граничных точек допустимой области. Предлагаемый алгоритм ориентирован на нахождение сильных первичных закономерностей за приемлемое время: достаточно произвести лишь несколько итераций алгоритма, чтобы получить решение, лучшее, чем получаемое жадным алгоритмом.

К содержанию главы имеются некоторые вопросы:

1). На стр. 118 выделены полиномиально разрешимые NP -трудные задачи и отмечается невозможность их реализации для решения задач условной псевдоболевой оптимизации. Однако автор предлагает новый алгоритм, в основе которого лежит метод ветвей и границ, который входит в класс решения NP -задач. Верно ли утверждение о несостоятельности такого рода задач?

2). В пп.5.1.2 и 5.1.3 автор ссылается на работу Антамошкина А.Н. [183]. Не означает ли наличие в этих пунктах повторения результатов работы данного автора?

3). Рассматривается задача (5.5) (она же (5.7) и на стр. 144, 154, 162 и 166), которая представлена в обобщенной форме. Из неё конкретно не видно наличие и условия множества локальных максимумов и какой из них следует признать лучшим.

4). Введение на 2/3 представляет собой материал учебно-методического содержания.

Шестая глава посвящена решению практических задач с применением нового метода. Его использование посредством классификации электрорадиоизделий космического применения позволило эффективно решить задачу формирования однородных партий при отбраковочных испытаниях. Фактически решается задача выявления закономерностей для классификации электрорадиоизделий с целью дальнейшего прогнозирования показателей безотказности электронной компонентной базы.

Предлагаемый метод был применен для решения задачи прогнозирования осложнений инфаркта миокарда. Здесь явно прослеживаются преимущества (по сравнению с другими методами: искусственные нейронные сети), заключающиеся в возможности объяснения результатов прогнозирования.

Автор обоснованно приводит доводы в пользу дальнейшего развития метода оптимизации для поиска логических решающих правил.

В «*заключении*» последовательно изложены основные решения, выработанные в результате выполненной научно-исследовательской работы, а также их возможности в достижении поставленной цели.

Достоверность и обоснованность научных положений, выводов и рекомендаций

Автором выполнен подробнейший анализ зарубежной и отечественной литературы по существующим методам классификации (основанным на правилах), алгоритмам комбинаторной оптимизации и решенным ранее практическим задачам.

Все основные положения, выводы и практические рекомендации строго обоснованы численными экспериментами, проведенными в соответствии с подробно описанной методикой на достаточном количестве тестовых примеров. Эксперименты подтверждают заявленные преимущества новых алгоритмов перед известными методами.

Достоверность результатов и обоснованность выводов подтверждается также высоким научным уровнем российских и зарубежных рецензируемых изданий, в которых опубликованы результаты диссертации. Ссылки на литературные источники обоснованы необходимостью преемственности в научных изысканиях. Результаты прошли всестороннюю апробацию на конференциях. Обоснованность подтверждается и тем, что результаты внедрены в промышленную эксплуатацию.

Научная новизна, ценность для науки

Научная ценность работы состоит в следующем, а именно: автор предлагает новый метод решения задачи классификации объектов на основе выявления в данных оптимальных логических правил.

Метод можно считать комплексным, поскольку автор разработал и включил в него ряд новых моделей оптимизации, обеспечивающих получение логических правил с высокой обобщающей способностью. Эти оптимизационные модели, с целью получения высокой результативности, включают в себя модули: определения порогов при бинаризации количественных признаков; поиска закономерностей, оптимальных по отношениям доказательности и избирательности; поиска пары сильных закономерностей (первичной и охватывающей); отбора закономерностей

для построения решающего правила.

Исследованы свойства созданных моделей, на основе которых разработаны алгоритмы поиска оптимальных решений.

Разработан новый алгоритм для решения класса задач оптимизации монотонных псевдобулевых функций с ограничениями. Алгоритм основан на известной и надежной схеме ветвей и границ, но при этом использует новый способ ветвления и оценки верхней границы, основанный на использовании свойств рассматриваемого класса задач. При этом положительным моментом можно считать то, что алгоритм не требует явного задания функций и не использует их релаксацию, в отличие от ранее известных способов применения схемы для подобных задач. Тем самым целевая функция и ограничения могут быть заданы алгоритмически, в виде «черного ящика». Предлагаемый алгоритм представляет самостоятельную ценность для развития методов комбинаторной оптимизации.

В диссертационном исследовании впервые для классификации объектов с помощью правил предложено использовать комбинации различных видов закономерностей, обладающих разными свойствами и дополняющих друг друга. Это расширяет возможности классификатора, основанного на правилах, и повышает его эффективность.

Результаты научной деятельности диссертанта позволили создать пригодные для работы в интерактивном режиме стабильные алгоритмы нахождения сильных закономерностей для решения задачи классификации с большим объемом входных данных, что является значительным вкладом в теорию машинного обучения, классификации и комбинаторной оптимизации.

Полученные научные результаты лично автором являются новыми, что подтверждается наличием различного рода публикаций.

Наиболее весомым новым результатом является создание метода нахождения сильных первичных и сильных охватывающих закономерностей, базирующегося на основе решения задачи комбинаторной оптимизации. Метод позволяет строить классификаторы, обеспечивающие высокую точность и при этом способные интерпретировать и обосновывать результаты. Автор доказал, что получаемые с использованием нового метода классификаторы дают в большинстве случаев результаты, не уступающие по точности известным методам, основанным на правилах.

Практическое значение результатов

Алгоритмы, разработанные в соответствии с предложенным методом, полезны для широкого круга производственных задач, где требуется классификация объектов для обоснования результатов и интерпретации (в виде логических правил) с повышенной точностью и стабильностью результата.

Метод классификации продукции по производственным партиям по данным проведенных испытаний может быть рекомендован для предприятий, выпускающих продукцию самого высокого класса качества (выше действующих стандартов) с требованиями однородности в пределах партии изделий в целях повышения надежности продукции.

Внедрение результатов диссертационного исследования в АО «Испытательный технический центр - НПО ПМ» позволяет предприятию совместно с заводами-изготовителями начать производство нового класса изделий повышенного качества и надежности для космической промышленности.

Акт о внедрении и государственная регистрация программ ЭВМ подтверждают практическую ценность метода оптимальных логических решающих правил.

Замечания

1. Было бы целесообразным отметить актуальность выполненной работы в области перспектив развития таких систем как Big Data и Data Mining, которые востребованы Национальной программой «Цифровая экономика Российской Федерации».
2. На стр. 91 (п.4.2) представлена решающая функция $D(a)$ как среднее расстояние между двумя противоположными закономерностями (ширина «разделяющей полосы»). Фактически она определяется исходя из нормальности закона распределения случайного наблюдения. Возможно ли использовать формулы, которые ориентированы на получение расстояния как медиану или моду?
3. В алгоритм поиска пары закономерностей (п.4.4) предлагается добавить дополнительные ограничения. Не создаст ли такое решение затруднений в поиске оптимального решения, поскольку ограничения могут привести к созданию дополнительных областей допустимых решений и появлению множества экстремальных точек.
4. Рассматривается задача (5.5) (она же (5.7) и на стр. 144, 154, 162 и 166), которая представлена в обобщенной форме. Из неё конкретно не видно наличие и условия множества локальных максимумов и какой из них следует признать лучшим.
5. Выводы в 5-ой главе на 2/3 представляет собой материал учебно-методического содержания.
6. Нет указаний или рекомендаций к тому, как выбрать параметр h (наибольшее заданное число порогов для каждого численного признака) в оптимизационной модели поиска порогов (стр. 65-66).
7. Свойство множества допустимых решений (стр. 78 диссертации и стр. 19 автореферата) касается строго монотонных унимодальных псевдобулевых функций. Но рассматриваемые далее функции являются не строго монотонными.
8. Для оценки качества закономерности рассматривается лишь один критерий – покрытие закономерностью наблюдений обучающей выборки. Не рассмотрены другие критерии, например, основанные на неопределенности информации или учете закономерностей распределения случайной выборки.

Общая оценка и заключение по диссертационной работе

Несмотря на имеющиеся замечания, технические недочеты, работа Масича И.С. является завершённой структурированной научно-квалификационной работой на актуальную тему, обладающей смысловым и содержательным единством формулировок темы, цели, поставленных задач, результатов и выводов, теоретические положения которой можно охарактеризовать как крупное научное достижение в области методов оптимизации и классификации.

В работе представлены обоснования поставленных научно-исследовательских задач, методы их решения подробно описаны и обоснованы, выполнен анализ полученных научных и практических результатов, сформулированы выводы и обобщения.

Изложение материала ясное, стилистика соответствует уровню докторского исследования, используется терминология и методика описания результатов, принятая в областях исследования операций, системного анализа, оптимизации и интеллектуального анализа данных.

Автореферат отражает основное содержание и результаты диссертации.

Диссертационная работа отвечает требованиям действующего «Положения о

порядке присуждения ученых степеней», введенного в действие постановлением Правительства Российской Федерации от 24.09.2013, предъявляемым к докторским диссертациям, в том числе, требованиям, установленным пунктом 9 «Положения». Тема, содержание и результаты диссертации соответствуют п.4 паспорта заявленной специальности. Таким образом, автор диссертации – 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:

доктор технических наук, доцент, профессор кафедры
«Информационные технологии и системы»,
Федеральное государственное бюджетное
образовательное учреждение высшего
образования «Хакасский государственный
университет им. Н.Ф. Катанова»

А.С. Дулесов

Научная специальность – 05.13.01 «Системный анализ, управление и обработка информации (по отраслям)

ФГБОУ ВО «Хакасский государственный
университет им. Н.Ф. Катанова»,
655017, Республика Хакасия, г. Абакан, ул. Ленина, д. 92/1, каб. 620.
Тел.: 8(390)2-222-432
E-mail: dulesov@khsu.ru

Подпись официального оппонента Александра Сергеевича Дулесова заверяю:

Ученый секретарь Хакасского государственного
университета им. Н.Ф. Катанова,
доктор исторических наук, профессор

Н.Я. Артамонова

«30» 09 2019 г.

