

ОТЗЫВ

официального оппонента Дмитриева Михаила Геннадьевича
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

Актуальность темы диссертации. Работа посвящена созданию подхода к решению задач классификации объектов, обеспечивающего определенное обоснование результатов классификации и интерпретацию в виде логических правил, позволяющего рассчитывать на повышенную точность и работу с большими объемами входных данных, что полезно для многих приложений. При этом рассматривается задача поиска оптимальных закономерностей в данных, которая является NP-трудной задачей глобальной оптимизации с большим числом экстремумов целевой функции. Применяемые при этом эвристические жадные алгоритмы, как правило, приводят к локальному экстремуму, часто далекому от глобального. Для повышения точности в подходе автора и вводится поиск рациональных закономерностей в данных, что, в свою очередь, приводит к необходимости решения задач псевдодулевой оптимизации. Разработка метода, позволяющего формировать рациональные логические правила и использовать их для поддержки принятия решения в задачах классификации с большим объемом входных данных, является сложной научной проблемой, и создание алгоритмов, повышающих точность решения, является актуальной задачей, важность которой особенно возрастает в тех случаях, когда цена ошибки высока.

Структура работы. Диссертация изложена на 261 странице, содержит введение, шесть глав, заключение, список литературы и приложение.

Первая глава посвящена обзору существующих моделей, алгоритмов и методов классификации, основанных на правилах. Сформулированы основные проблемы существующих подходов, а также требования к новому подходу и его основная идея.

Во второй главе рассматривается вопрос преобразования данных с разнотипными признаками в бинарный вид для целей логического анализа данных и применения моделей и алгоритмов выявления логических закономерностей. Рассматриваются способы кодирования разнотипных, в том числе количественных признаков и предлагается оптимизационная модель для поиска пороговых значений.

В третьей главе задача поиска закономерностей в данных рассматривается как задача оптимизации. Рассмотрены критерии оценки качества закономерностей. Исследована модель оптимизации П. Хаммера, для которой показано, что она обеспечивает нахождение сильной (максимальной по отношению доказательности) и первичной (наиболее короткой)

закономерности. При этом показана недостаточность точности жадного алгоритма и потребность в более точном решении. Также предлагается новая модель оптимизации для поиска сильных охватывающих закономерностей, которая является необходимым дополнением к модели П. Хаммера и позволяет формировать правила с дополнительными условиями, что является особенностью предлагаемого в диссертации метода.

В четвертой главе рассматриваются вопросы отбора и применения закономерностей для построения решающего правила распознавания. Предлагается техника совместного использования закономерностей различных типов. Разработаны алгоритмы для поиска пары закономерностей и модель оптимизации для выявления пары закономерностей.

Пятая глава посвящена алгоритмам оптимизации. Рассмотрены классы задач псевдодобулевой оптимизации и их свойства. Излагается новый алгоритм условной псевдодобулевой оптимизации, основанный на схеме метода ветвей и границ и поиске среди граничных точек допустимой области. Предлагаемый алгоритм основывается на новой схеме ветвления и поиска верхней границы, которые опираются на свойства рассматриваемого класса задач. Применение новой схемы ветвления позволяет быстро исключать неперспективные подобласти пространства оптимизации. Показано, что предлагаемый автором алгоритм позволяет улучшить решение, получаемое жадным алгоритмом, при этом дополнительные затраты по трудоемкости допустимы с практической точки зрения. Такие приемы улучшения начального приближения при решении задач оптимизации, используя специфику постановки, сейчас активно развиваются в литературе.

В шестой главе рассматривается применение разработанных алгоритмов к решению практических задач классификации. В частности, описывается решение задачи классификации радиодеталей космического применения.

Приложение содержит сведения о разработках и внедрениях результатов диссертационного исследования в производство.

Достоверность и обоснованность научных положений, выводов и рекомендаций. Все рассуждения и выводы о эффективности новых алгоритмов и нового метода оптимальных логических решающих правил подтверждаются корректными рассуждениями и численными экспериментами.

Научная новизна, ценность результатов и выводов для науки и практики. В диссертации предложен целый ряд новых моделей и алгоритмов оптимизации, предназначенных для выявления логических закономерностей в массивах данных и их использования для построения решающего правила классификации объектов. Совокупность разработанных моделей оптимизации и алгоритмов их решения составляет новый метод – метод оптимальных логических решающих правил. Впервые предлагается подход к решению задачи распознавания, заключающийся в использовании одновременно различных видов закономерностей, обладающих различными свойствами, которые ранее использовались отдельно. Совместное использование сильных первичных и охватывающих закономерностей попарно в соответствии с предлагаемой схемой позволяет проявить преимущества тех и других. Новый

алгоритм псевдоболевой оптимизации, описанный в главе 5, обеспечивает нахождение закономерностей, как показано автором на численных экспериментах, с существенно более высоким значением целевой функции, чем жадный алгоритм, что повышает точность решения задач классификации и позволяет использовать в классификаторе меньшее число логических правил. Этим достигается компактность классификатора и повышение интерпретируемости. Разработанный алгоритм условной псевдоболевой оптимизации имеет самостоятельную ценность и предназначен для решения задач условной оптимизации монотонных псевдоболевых функций, в том числе при алгоритмическом их задании. Ценность результатов для практики подтверждается актами внедрения. Новые научные результаты, выносимые на защиту, принадлежат лично автору диссертации. Заимствования без ссылки на источник в диссертации отсутствуют.

Замечания.

1. Работа несколько словесно перегружена и часто напоминает собой некоторый обзор.

2. В работе получено достаточно много новых результатов на стыке теории систем, теории оптимизации и искусственного интеллекта, но, на наш взгляд, в работе недостаточно внимания уделено вопросам математической формализации и технической реализации.

3. Встречаются повторы текста. Так, например, на стр.11-13 и стр.227-228 повторяются основные результаты работы.

4. В работе рассматриваются и задачи с нечеткими данными, и при этом часто встречаются выражения типа «лучший», «оптимальный», но не всегда рядом есть соответствующие критерии оптимальности, оценки близости. Естественно, возникают вопросы, как эти слова понимать. Автору следовало дать более точные определения.

5. К недостаткам работы можно отнести и отсутствие простых тестовых примеров для иллюстрации предложенных результатов, рассуждения по этому поводу описательные и не всегда позволяют быстро представить пошаговый алгоритм.

6. Автор знаком с целым рядом подходов к построению классификаторов и к логическому анализу данных. В главе 1 приводится сравнительный анализ эффективности этих подходов и алгоритмов с алгоритмами, предлагаемыми в работе автора. На наш взгляд, анализ не во всем убедительный.

Общая оценка и заключение. Имеющиеся замечания не снижают общей положительной оценки работы. 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 научном уровне. Полученные автором результаты достоверны, выводы и заключения являются обоснованными. Научные результаты и выводы, представленные в диссертации, являются личным научным достижением автора в области методов классификации объектов и комбинаторной оптимизации. Результаты диссертации достаточно полно опубликованы в 37 публикациях в изданиях, рекомендованных ВАК РФ, и

докладывались на отечественных и международных конференциях. Содержание диссертации и полученные результаты соответствуют п.4 паспорта специальности 05.13.01 — системный анализ, управление и обработка информации.

Автореферат соответствует основному содержанию диссертации, содержит описание основных этапов исследования, полученные результаты и выводы. Оформление автореферата и диссертации соответствует требованиям ВАК РФ.

Представленная диссертационная работа отвечает требованиям действующего 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, главный научный сотрудник
Института системного анализа Федерального исследовательского центра
«Информатика и управление» РАН
доктор физико-математических наук,
профессор

М.Г.Дмитриев

Адрес организации:

117312, Москва, проспект 60-летия Октября, 9, комната №1214

Институт системного анализа Федерального исследовательского центра

«Информатика и управление» Российской академии наук

тел. 8-499-135-43-32,

адрес электронной почты: mdmitriev@mail.ru

