

Отзыв

научного руководителя д.ф.-м.н. Дюковой Елены Всеволодовны
на диссертацию П.А. Прокофьева «Корректное распознавание по прецедентам:
построение логических корректоров общего вида и вычислительные аспекты»
на соискание учёной степени кандидата физико-математических наук по
специальности 05.13.17 – Теоретические основы информатики.

П.А. Прокофьев в 2005 г. закончил с отличием Институт криптографии связи и информатики Академии ФСБ России по специальности «Организация и технология защиты информации». Являлся соискателем в Вычислительном центре им. А.А. Дородницына РАН. В период подготовки диссертации принимал участие в выполнении грантов РФФИ №10-01-00770-а «Развитие методов логического и алгебро-логического синтеза корректных процедур распознавания», №13-01-00787-а «Исследование в области логического и алгебро-логического анализа данных в распознавании» и №16-01-00445-а «Логический анализ данных в распознавании: обобщение классических подходов и вычислительные аспекты»..

Для обработки данных в плохо формализованных областях успешно применяются методы распознавания образов, в частности методы, основанные на распознавании по прецедентам. Примерами являются задачи медицинской диагностики и прогнозирования, обработки социологической информации, технического и геологического прогнозирования, анализа банковской деятельности, анализа пользовательской среды интернета и т.д. В диссертации П.А. Прокофьева рассмотрены вопросы применения аппарата дискретной математики (логических методов анализа данных) для решения одной из центральных задач распознавания образов – задачи распознавания по прецедентам. Указанный аппарат играет особую роль для анализа малых обучающих выборок в условиях, когда методы прикладной статистики и математического моделирования практически не применимы. Особое внимание уделяется построению корректных распознающих алгоритмов, то есть алгоритмов, безошибочно классифицирующих прецеденты. В случае больших данных (при большой размерности признакового пространства) возникают сложности вычислительного характера, обусловленные необходимостью рассматривать труднорешаемые дискретные задачи, среди которых центральное место занимает перечислительная задача, называемая дуализацией. Построение логических распознающих процедур основано на поиске в данных конъюнктивных закономерностей или элементарных классификаторов (эл.кл.). Корректность классических моделей обеспечивается корректностью используемых эл.кл., каждый из которых правильно классифицирует обучающие объекты. Современные модели (модели логических корректоров) строятся на базе произвольных (не обязательно корректных) эл.кл. с использованием идей алгебраического подхода. Эти модели показывают лучшее качество по сравнению с классическими моделями логических процедур распознавания на сложных для логического подхода задачах. Такими задачами, в частности, являются задачи с целочисленными признаками высокой значности.

При построении логического корректора на этапе обучения ставится трудная в вычислительном плане задача построения семейства корректных наборов эл.кл., в котором каждый набор обладает хорошей распознающей способностью. Для ускорения поиска применяются генетические алгоритмы, а также итерационные и стохастические методы обработки обучающей информации. В ряде моделей для сокращения вычислительных затрат вводится ограничение на ранг используемых эл.кл. Поэтому для расширения области применимости логических корректоров актуальны вопросы разработки новых, более совершенных моделей. Эти вопросы напрямую связаны с вопросами обобщения понятия корректного набора эл.кл. и использования этого понятия для разработки общей схемы синтеза корректных логических алгоритмов распознавания.

Дуализации уделяется большое внимание в работах отечественных и зарубежных авторов, поскольку дуализация полиномиально или квазиполиномиально эквивалентна многим прикладным задачам в различных областях, таких, как искусственный интеллект, теория баз данных, распознавание образов, математическое программирование, теория матроидов, теория надежности и т.д. Вопрос о полиномиальной разрешимости дуализации поставлен более 40 лет назад, однако ответ на этот вопрос до сих пор не найден. При решении прикладных задач большой размерности для сокращения вычислительных затрат приходится отказываться от перечисления всех решений дуализации. Тем не менее, на точных алгоритмах базируются многие приближенные алгоритмы дуализации, например стохастические. Последние исследования свидетельствуют о том, что среди известных алгоритмов дуализации наиболее быстрыми являются асимптотически оптимальные алгоритмы (АО-алгоритмы). АО-алгоритмы, в отличие от других практически применимых алгоритмов дуализации, имеют теоретическое обоснование эффективности в типичном случае (в среднем). Подход впервые предложен в 1977 году Е.В. Дюковой.

В диссертации рассмотрены две группы вопросов. Во-первых, это вопросы разработки общих подходов к синтезу корректных распознающих алгоритмов с использованием логических методов анализа данных и построения на этой основе новых более совершенных моделей логических процедур распознавания. Во-вторых, это вопросы построения и экспериментального исследования вычислительной сложности алгоритмов решения дискретных перечислительных задач, возникающих при построении логических процедур распознавания. Получены следующие результаты.

1. Построена общая схема синтеза корректных логических процедур распознавания. Фактически обобщено понятие корректного набора эл.кл. и на этой основе построена модель логического корректора общего вида. В рамках построенной общей схемы описаны классические логические алгоритмы распознавания и ранее построенные логические корректоры. При построении логического корректора общего вида в качестве корректирующих предложены функции более широкого класса, чем ранее используемые. Усовершенствованы методики повышения качества распознавания и скорости обучения логических корректоров. Для итеративного формирования семейств корректных наборов эл.кл. с хорошей распознающей способностью и высокой диверсификацией адаптирован метод бустинга. На реальных задачах проведено тестирование новых методик формирования семейств корректных наборов эл.кл. и созданных на их основе моделей логических корректоров.

2. Проведено существенное усовершенствование АО-алгоритмов дуализации с целью уменьшения временных затрат. Работу АО-алгоритма дуализации можно представить в виде обхода дерева решений (ДР) в глубину. На каждом шаге за полиномиальное время строится в точности одна висячая вершина ДР, которая либо является найденным впервые решением задачи, либо является лишним шагом. Число лишних шагов АО-алгоритма для почти всех задач данного размера имеет более низкий порядок роста по сравнению с числом всех решений задачи с ростом размера её входа. Для ряда отечественных АО-алгоритмов в результате применения «жадной» стратегии при построении вершины в ДР удалось значительно уменьшить общее число вершин в ДР, что привело к существенному сокращению временных затрат. В результате удалось обогнать наиболее быстрые из существовавших алгоритмов дуализации, среди которых АО-алгоритмы, разработанные учеными из Японии К. Murakami и Т. Uno. В экспериментах были использованы модельные данные и данные прикладных задач большого размера. Всего было обчислено около 1000 задач. Таким образом, подтверждена практическая применимость АО-алгоритмов дуализации. Следует отметить, что более ранние отечественные исследования по усовершенствованию АО-алгоритмов были в основном направлены на сокращение числа лишних шагов АО-алгоритма за счёт усложнения его шага, поэтому не удавалось значительно уменьшить временные затраты.

В целом диссертация является законченной научно-исследовательской работой, имеющей как теоретическое, так и практическое значение. Результаты диссертационной работы важны для развития распознавания образов, дискретной математики и теории сложности дискретных задач и могут быть использованы в работе институтов и учреждений, ведущих исследования в указанных областях. Эти результаты включены в спецкурс «Логический анализ данных в распознавании», читаемый на факультете ВМиК МГУ.

Основные результаты диссертации опубликованы. Всего 15 публикаций, из них пять статей в журналах, рекомендованных ВАК. Автореферат правильно отражает содержание диссертации. Результаты диссертации были представлены автором на нескольких международных и всероссийских конференциях, вошли в отчеты по грантам РФФИ, упомянутым выше.

За время работы над диссертацией П.А. Прокофьев проявил большую самостоятельность при проведении исследований, продемонстрировал способность решать сложные задачи и обосновывать предлагаемые им решения, а также способность анализировать полученные результаты и делать необходимые выводы. Он показал высокую квалификацию, изучил большое число литературных источников и получил существенные результаты в области логического анализа данных и построения логических процедур распознавания. Результаты диссертационной работы важны для развития распознавания образов, дискретной математики и теории сложности дискретных задач.

Работа полностью соответствует требованиям, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук по специальности 05.13.17 – «Теоретические основы информатики», а ее автор П.А. Прокофьев заслуживает присуждения ему учёной степени кандидата физико-математических наук по данной специальности.

Научный руководитель
доктор физико-математических наук
доцент
7 июля 2015 г.

Е. В. Дюкова

Дюкова Елена Всеволодовна, edjukova@mail.ru

Главный научный сотрудник Вычислительного центра им. А.А. Дородницына
Федерального государственного учреждения «Федеральный исследовательский центр
«Информатика и управление» Российской академии наук», Россия, 119333, г. Москва,
ул. Вавилова, д. 44 корп. 2, www.ccas.ru, телефон: +7(499)135-41-63.

