

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертационную работу Маслякова Глеба Олеговича на тему
«Корректная классификация над произведением частичных порядков»,
представленную на соискание учёной степени кандидата физико-
математических наук по специальности

1.2.3 – «Теоретическая информатика, кибернетика»

Оценка актуальности темы диссертационной работы

Классификация объектов является наиболее востребованной задачей интеллектуальных систем. Традиционно для решения этой задачи используются разные подходы, в том числе с привлечением нейронных сетей. Однако подобные алгоритмы в основном требуют обучения на большом массиве верифицированных данных. При этом для достаточно большого круга прикладных задач классификации собрать достаточный корпус данных не представляется возможным, что требует применения нестандартных алгоритмов классификации. В своей диссертации Масляков Г. О. рассматривает иной класс алгоритмов классификации по прецедентам, а именно класс логических алгоритмов, и обобщает основные конструкции таких алгоритмов на случай, когда информация представлена в виде декартова произведения конечных частично упорядоченных множеств. Во многих прикладных задачах необходимость рассматривать частичные порядки на множествах значений признаков возникает даже в случае бинарных данных. Подобные обобщения активно исследуются, например, в задачах теории графов и гиперграфов.

Для успешной классификации каждого прецедента автором осуществляется поиск специальных фрагментов описаний этого прецедента. Основная трудность при этом состоит в том, что выбор подходящих фрагментов описания прецедентов является в целом переборной задачей, где число вариантов экспоненциально связано с количеством параметров задачи. При этом востребованность в быстрых и эффективных классификаторах на малых данных постоянно растёт. Таким образом, созданные автором в результате подготовки диссертации процедуры классификации над произведением частичных порядков предназначены для решения прикладных задач, для которых не применим традиционный логический подход.

Структура и содержание работы

В состав диссертации на соискание учёной степени кандидата физико-математических наук вошли: введение, четыре главы, заключение и список литературы.

Введение написано грамотно и позволяет оценить общую проблематику логического подхода, а также историю его развития.

В первой главе приведено описание единой схемы синтеза алгоритмов логической классификации, включающей направления Correct Voting Procedures (CVP), Logical Analysis of Data (LAD) и Formal Concept Analysis (FCA). Данные направления разработаны различными школами независимо, однако автором доказано, что обучение каждого логического классификатора сводится к поиску так называемых корректных элементарных классификаторов (ЭК), максимальных с точки зрения некоторого частичного предпорядка. Каждое направление устанавливает свой предпорядок. Описание схемы даётся с использованием терминологии CVP.

Во второй главе предложены обобщённые процедуры CVP для случая частично упорядоченных данных. Показывается, что поиск множества тупиковых (безизбыточных) корректных ЭК общего вида сводится к известной труднорешаемой задаче дуализации над произведением частичных порядков. Создаются практические модели логической классификации над произведением частичных порядков, основанные на стохастической композиции над обобщёнными логическими классификаторами. Приводится подробное экспериментальное сравнение предлагаемых моделей.

В третьей главе разработаны эффективные методы задания частичных порядков на множествах значений признаков, обеспечивающих высокое качество классификации. Приведён критерий корректности обобщённой процедуры CVP, описанной во второй главе. На основании данного критерия сформулирована теорема, позволяющая свести задачу поиска частичных порядков, обеспечивающих корректную классификацию, к задаче монотонной дуализации. Предложена процедура выбора частичного порядка, включающая использование генетического алгоритма для быстрого приближённого поиска минимального покрытия булевой матрицы, обеспечивающая корректность классификации. Рассмотрена процедура выбора частичных порядков, основанная на оценке информативности значений отдельных признаков, которая не гарантирует корректность классификации, но обеспечивает высокое качество классификации при меньшем времени счёта. Приведены результаты экспериментального исследования предложенных процедур.

В четвертой главе поставлена задача дуализации над произведением частичных порядков в матричном виде. Дан вывод асимптотических оценок типичных значений количественных характеристик множества решений задачи дуализации над произведением цепей. Построен асимптотически оптимальный алгоритм дуализации над произведением цепей – RUNC-M+, теоретическое обоснование которого базируется на полученных в этой главе оценках. Описана реализация алгоритма RUNC-M+ с использованием битовых операций. Создан алгоритм дуализации над произведением частично упорядоченных множеств, каждое из которых может являться как цепью, так и антицепью – RUNC-M++.

Основные результаты и их научная новизна

В диссертационной работе Маслякова Г.О. получены следующие основные результаты.

1. Разработан общий подход к синтезу логических процедур классификации по прецедентам, включающий три направления: Correct Voting Procedures, Logical Analysis of Data и Formal Concept Analysis в рамках единого понятийного инструментария CVP.
2. На основе обобщения классических понятий направления CVP построены и исследованы корректные модели классификаторов над декартовым произведением частично упорядоченных множеств.
3. Для множеств значений признаков описана процедура установления частичных порядков и сформулирована задача выбора таких частичных порядков, которые обеспечивают корректность процедуры классификации. Рассмотрены некоторые подходы к решению данной задачи.
4. Описана «ускоренная» процедура упорядочения множеств допустимых значений признаков линейным образом. Данная процедура, имея достаточную точность классификации, работает быстрее, но не гарантирует корректность классификации.
5. Для задачи дуализации над произведением цепей предложен асимптотически оптимальный алгоритм с теоретически доказанной асимптотикой.
6. Созданы и исследованы методики повышения эффективности логических классификаторов, на основе стохастических композиций.
7. Для подтверждения эффективности теоретически обоснованных в рамках данной работы процедур проведены эксперименты, подтверждающие корректность результатов.

Несмотря на схожесть направлений CVP, LAD и FCA, ранее не исследовалась возможность создания единой схемы синтеза логического классификатора, объединяющей все направления. Впервые был предложен подход к построению корректных процедур CVP над произведением частичных порядков. Продемонстрирована возможность применения асимптотически оптимального подхода к решению сложных перечислительных задач, возникающих при обучении обобщённых классификаторов. Сформулирована сложная задача выбора «корректных» частичных порядков на множествах значений признаков для обеспечения корректности классификации и предложены методы ее решения. Разработаны практические модели стохастической композиции логических классификаторов над произведением частичных порядков и проведено их экспериментальное исследование. Все полученные результаты являются новыми!

Достоверность полученных результатов подтверждается публикациями представляемых результатов в рецензируемых научных изданиях, а также докладами на российских и международных конференциях по анализу данных. Проведённые эксперименты подтверждают логичность и корректность теоретических оценок. Научные положения, выводы и рекомендации, сформулированные в диссертации, являются обоснованными и подкреплены необходимыми математическими утверждениями.

Замечания

1. Работа оформлена согласно всем требованиям, предъявляемым к кандидатским диссертациям, однако в ней содержатся ряд опечаток, в целом не несущих существенных проблем в понимании результатов работы. Например, на странице 58 в первой строке формула записана не корректно, хотя смысл её очевиден. На странице 72 ссылка указана некорректно, в связи с чем необходимо приложить усилия для понимания, какую задачу автор имеет ввиду.
2. Иногда в работе автор ссылается на термины, не введённые ранее, например «кроссэнтропия» на стр. 73, что в целом не вызывает больших сложностей, но не желательно.
3. Небольшое количество примеров и малая размерность рассматриваемых в примерах объектов затрудняют восприятие работы. При этом один из примеров присутствует в работе дважды.
4. В диссертации не приведены критерии, по которым проводился выбор библиотек бустинга, и конкретные методы бустинга выбирались также произвольно.

Приведённые замечания не снижают общей высокой оценки диссертационной работы Г. О. Маслякова.

Результаты диссертации изложены в четырёх работах, опубликованных в изданиях, рецензируемых ВАК, а также в двух и четырёх работах, опубликованных в изданиях, индексируемых в Web Of Science Core Collection и Scopus, соответственно. Автореферат составлен корректно, достаточно полно и правильно отражает содержание диссертации, что позволяет судить о результатах работы.

Диссертация Маслякова представляет собой законченное научное исследование, содержит решение задачи, имеющей значение для развития методов анализа данных, она написана автором самостоятельно, обладает внутренним единством, содержит новые научные результаты, полученные автором лично. В диссертации приведены рекомендации по использованию научных выводов.

Работа полностью соответствует требованиям ВАК при Минобрнауки России, предъявляемым к диссертациям на соискание учёной степени кандидата физико-математических наук по специальности 1.2.3 – «Теоретическая информатика, кибернетика», – а её автор, Г. О. Масляков, заслуживает присуждения ему учёной степени кандидата наук по данной специальности.

Официальный оппонент,
кандидат физико-математических наук,
заведующий кафедрой
«Прикладная математика и программирование»
федерального государственного бюджетного
образовательного учреждения высшего образования
«Российский государственный университет
им. А.Н. Косыгина (Технологии. Дизайн. Искусство)»
Адрес: 119071, г. Москва, ул. Малая Калужская, д.1, каб. 1710
Телефон: +7 (495) 951-31-48
Email: mokryakov-av@rguk.ru

 Мокряков Алексей Викторович

Дата

04 декабря 2023 г.

Подпись руки и сведения заверяю
Проректор по науке и инновациям
д. э. н., доцент



 Силаков Алексей Викторович