

УТВЕРЖДАЮ

Директор ФИЦ ИУ РАН,
академик РАН



Соколов И.А.

2023 г.

ЗАКЛЮЧЕНИЕ

Федерального государственного учреждения
«Федеральный исследовательский центр «Информатика и управление»
Российской академии наук»

Диссертационная работа Маслякова Глеба Олеговича «Корректная классификация над производением частичных порядков», представленная на соискание учёной степени кандидата физико-математических наук по специальности 1.2.3 – «Теоретическая информатика, кибернетика», выполнена в отделе №12 Федерального исследовательского центра «Информатика и управление» Российской академии наук.

В 2020 г. Масляков Глеб Олегович окончил факультет вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова по специальности «Прикладная математика и информатика». В 2020 г. поступил в очную аспирантуру ФИЦ ИУ РАН. Справка об обучении в аспирантуре и сдаче кандидатских экзаменов выдана в 2023 году в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук».

В период подготовки диссертации с 2019 г. по 2023 г. Масляков Глеб Олегович работал в компании «Яндекс» в должности разработчика программного обеспечения.

Научный руководитель – д.ф.-м.н. Дюкова Елена Всеволодовна – работает в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» в должности главного научного сотрудника.

По итогам обсуждения было принято следующее заключение.

Актуальность темы исследования.

В диссертации рассматривается одна из центральных задач машинного обучения – задача классификации на основе прецедентов. Основное достоинство логического подхода к задаче классификации – возможность получения результата при отсутствии дополнительных предположений вероятностного характера и при небольшом числе прецедентов.

Актуальным является создание процедур логической классификации, ориентированных на решение задач классификации по прецедентам в случае, когда на множествах значений признаков заданы частичные порядки (информация представлена в виде произведения конечных частично упорядоченных множеств). Создание подобных моделей классификаторов обусловлено существованием прикладных задач классификации, качественное решение которых невозможно, когда для сравнения целочисленных признаков объектов используется простое отношение равенства.

В ходе работы были поставлены и достигнуты следующие **цели**:

1. Обзор основных направлений логической классификации целочисленных данных и создание единой схемы синтеза алгоритмов логической классификации с использованием терминологии Correct Voting Procedures (CVP).
2. Обобщение классических понятий CVP и описание схемы классификации для CVP в случае, когда описания объектов представляют собой элементы декартова произведения конечных частично упорядоченных множеств.
3. Постановка задачи дуализации над произведением частичных порядков в матричном виде. Построение асимптотически оптимальных алгоритмов дуализации над произведением цепей.
4. Разработка и исследование эффективных методов задания частичных порядков на множествах значений признаков, обеспечивающих высокое качество классификации.
5. Создание стохастических композиций процедур корректного голосования над произведением частичных порядков.

Основные результаты диссертации, выносимые на защиту:

1. Разработана единая схема синтеза алгоритмов логической классификации с использованием терминологии Correct Voting Procedures (CVP).
2. Создана общая схема синтеза логических классификаторов в случае частично упорядоченных данных, которая может быть использована для описания классических логических алгоритмов классификации и предлагаемых в рамках данной работы моделей.
3. Предложена методика повышения качества классификации без потери корректности за счёт выбора частичного порядка на множествах значений признаков.
4. Разработан асимптотически оптимальный алгоритм дуализации над произведением цепей RUNC-M+. Дано теоретическое обоснование предлагаемому алгоритму на основе матричной формулировки задачи дуализации над произведением частичных порядков.
5. Созданы методики повышения качества классификации путём синтеза стохастических композиций логических классификаторов.
6. Экспериментально подтверждено, что предлагаемые в рамках данной работы процедуры результативно применимы для решения проблемы классификации по прецедентам.

Научная новизна. Автором в рамках данной работы впервые создана общая схема синтеза процедур CVP над произведением частичных порядков. Построены и исследованы практические классификаторы, базирующиеся на асимптотически оптимальных алгоритмах дуализации над произведением частичных порядков. Все полученные в рамках диссертационной работы результаты являются новыми.

Теоретическая и практическая значимость. Диссертационная работа содержит как теоретические, так и практические результаты. В работе рассмотрены методологические аспекты корректных алгоритмов логической классификации. Показано, что известные алгоритмы классификации, описанные в традициях разных научных направлений, могут быть синтезированы в рамках единой схемы.

В рамках направления CVP создана общая схема синтеза корректных логических алгоритмов классификации по прецедентам над произведением частичных порядков, согласно которой классические модели CVP – это классификаторы над произведением антицепей.

Дана матричная формулировка задачи дуализации над произведением частичных порядков, в рамках которой установлено, что в общем случае анализ прецедентной информации приводит к необходимости находить так называемые упорядоченные (тупиковые) покрытия целочисленной матрицы. Введённое понятие упорядоченного (тупикового) покрытия целочисленной матрицы – обобщение известного понятия (неприводимого) покрытия булевой матрицы. Доказана теорема о числе упорядоченных тупиковых покрытий целочисленной матрицы в типичном случае.

Исследована актуальная и ранее не изучавшаяся задача выбора на этапе предварительного анализа обучающей выборки «хороших» частичных порядков. Разработана «быстрая» процедура линейного упорядочения множеств допустимых значений признаков, эффективная по времени вычислений и позволяющая повысить качество классификации, но не гарантирующая корректность классификации. Поставлена задача выбора «корректных» частичных порядков на множествах допустимых значений признаков, и показано, что эта задача может быть решена на основе построения неприводимых покрытий специальной булевой матрицы.

На базе идей бэггинга и бустинга построены и экспериментально исследованы стохастические композиции логических классификаторов общего вида.

Степень достоверности. Достоверность полученных результатов обеспечивается доказательствами сформулированных утверждений и теорем, а также результатами экспериментов, проведённых автором.

Основные результаты работы докладывались на конференциях «Математические методы распознавания образов (ММРО-18)» (Таганрог, 2017), «Дискретные модели в теории управляющих систем» (Подмосковье, 2018), «Математические методы распознавания образов (ММРО-19)» (Москва, 2019), «Конференция по искусственному интеллекту (КИИ-2019)»

(Ульяновск, 2019), «Интеллектуализация обработки информации (ИОИ-13)» (Москва, 2020), «Информационные технологии и нанотехнологии (ИТНТ-2021)» (Самара, 2021), «Математические методы распознавания образов (ММРО-20)» (Москва, 2021), «Информационные технологии и нанотехнологии (ИТНТ-2023)» (Самара, 2023).

Список работ, опубликованных автором по теме диссертации.

1. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О дуализации над произведением частичных порядков // Машинное обучение и анализ данных 2017. Т. 3, Вып. 4. С. 239–249.
2. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. Дуализация над произведением цепей: асимптотические оценки числа решений // Доклады Академии наук, 2018. Т. 483. № 2. С. 130 – 133; Djukova E. V., Masliakov G. O., Prokofjev P. A. Dualization Problem over the Product of Chains: Asymptotic Estimates for the Number of Solution // Doklady Mathematics. 2018. Vol. 98. No. 3. P. 564–567.
3. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О поиске максимальных независимых элементов частичных порядков (случай цепей) // Прикладная математика и информатика. Москва: МАКС Пресс, 2018. Вып. 58. С. 12–20.
4. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. Задача монотонной дуализации и её обобщение — дуализация над произведением цепей // Труды 10-й Международной конференции «Дискретные модели в теории управляющих систем». Москва и Подмосковье 23-25 мая 2018 г. Труды / Отв. ред. В. Б. Алексеев, Д. С. Романов, Б. Р. Данилов. Москва: МАКС Пресс. 2018. С. 117 – 119.
5. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. Задача дуализации над произведением цепей: асимптотика типичного числа решений // Тезисы докладов 12-й Международной конференции «Интеллектуализация обработки информации (ИОИ-2018)» (Москва, Россия – Гаэта, Италия). М.: ТОРУС ПРЕСС. 2018. С. 12–159.
6. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О числе максимальных независимых элементов частичных порядков (случай цепей) // Информатика и её применения, 2019. Т. 13. Вып. 1. С. 25–32;

7. Djukova E. V., Masliakov G. O., Prokof'ev P. A. Finding Maximal Independent Elements of Products of Partial Orders (the Case of Chains) // Computational Mathematics and Modeling. 2019. Vol. 30. Iss. 1. P. 7–12.
8. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О логическом анализе данных с частичными порядками в задаче классификации по прецедентам // Ж. вычисл. матем. и матем. физ., 2019. Т. 59. № 9, С. 1605–1616; Djukova E. V., Masliakov G. O., Prokof'ev P. A. On the Logical Analysis of Partially Ordered Data in the Supervised Classification Problem // Computational Mathematics and Mathematical Physics. 2019. Vol. 59. Iss. 9. P. 1542–1552.
9. Djukova E. V., Masliakov G. O., Prokof'ev P. A. Logical Classification of Partially Ordered Data // 7th Russian Conference, RCAI 2019, Ulyanovsk, Russia. October 21–25, 2019. Proceedings. Editors: Kuznetsov, Sergei O., Panov, Aleksandr I. P. 115–126.
10. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. Классификация над произведением частичных порядков // Математические методы распознавания образов: 19-я Всероссийская конференция с международным участием, ММРО-2019. Москва. 26–29 декабря 2019 г. Тезисы докладов. С. 29–31.
11. Бакланова А. О., Дюкова Е. В., Масляков Г. О. Исследование зависимости качества классификации от выбора частичных порядков на множествах значений признаков // Интеллектуализация обработки информации. Тезисы докладов 13-й Международной конференции, г. Москва. 2020. С. 21 – 25.
12. Дюкова Е. В., Масляков Г. О., Прокофьев П. А. О выборе частичных порядков на множествах значений признаков в задаче классификации // Информатика и её применения, 2021. Т. 15. Вып. 4. С. 74–80.
13. Дюкова Е. В., Масляков Г. О. О корректной классификации над произведением частичных порядков // Сборник трудов по материалам VII Международной конференции и молодёжной школы (ИТНТ-2021). Т. 3. Самара, 2021. С. 34292; Djukova E. V., Masliakov G. O. Correct classification over a product of partial orders // IEEE Proceedings of the VII International Conference on Information Technology and Nanotechnology (ITNT-2021), Samara, Russia, 2021. P. 1–5.
14. Дюкова Е. В., Масляков Г. О. Корректная классификация над произведением частичных порядков // Математические методы распознавания образов: 20-я Всероссийская конференция с

международным участием, ММРО-2021. Москва. 7–10 декабря 2021 г. Тезисы докладов. С. 59–63.

15. *Djukova E. V., Masliakov G. O.* Correct classification over a product of partial orders // IEEE Proceedings of the IX International Conference on Information Technology and Nanotechnology (ITNT-2023), Samara, Russia, 2023. P. 1–5.

В изданиях ВАК опубликованы работы [2, 6, 8, 12]. Англоязычные версии работ [2, 8] опубликованы в журналах, индексируемых в Web Of Science Core Collection. В изданиях, индексируемых в Scopus, опубликованы работы [7, 9, 15] и англоязычная версия работы [13].

Содержание диссертации и основные положения, выносимые на защиту отражают **персональный вклад** автора в опубликованные работы. Все представленные в диссертации результаты получены лично автором под научным руководством д.ф.-м.н. Е. В. Дюковой.

Диссертационная работа Маслякова Г. О. «Корректная классификация над произведением частичных порядков» является законченной самостоятельной научно-квалификационной работой и удовлетворяет всем требованиям пунктов 9, 10 и 14 Положения о присуждении учёных степеней, утверждённого постановлением Правительства Российской Федерации от 24 сентября 2013 года №842, а также Паспорту научной специальности 1.2.3. – «Теоретическая информатика, кибернетика», в частности по следующим пунктам:

- пункт 3: теория сложности алгоритмов и вычислений
- пункт 17: анализ больших данных, обнаружение закономерностей в данных и их извлечение.

Диссертационная работа Маслякова Г. О. «Корректная классификация над произведением частичных порядков» рекомендуется к защите на соискание учёной степени кандидата физико-математических наук по специальности 1.2.3. – «Теоретическая информатика, кибернетика» (физико-математические науки).

Заключение принято на заседании семинара отделения 1 Федерального государственного учреждения «Федеральный исследовательский центр

«Информатика и управление» Российской академии наук». Присутствовало на заседании 14 человек. Результаты голосования «за» – 14 человек, «против» – нет, «воздержалось» – нет.

Протокол №1 от 12 сентября 2023 г.

Председатель семинара:

д.ф.-м.н.

зав. отдела №13 ФИЦ ИУ РАН

 В. И. Цурков

Секретарь:

к.ф.-м.н.

н.с. отдела №13 ФИЦ ИУ РАН

 Е. К. Егорова