

ОТЗЫВ
официального оппонента д.ф.-м.н. С.Д. Двоенко
на диссертацию Драгунова Никиты Аркадьевича
«Поиск частых (нечастых) элементов декартова произведения конечных
частичных порядков и приложения»,
представленную на соискание ученой степени
кандидата физико-математических наук по специальности
1.2.3 –«Теоретическая информатика, кибернетика»

Актуальность темы диссертации. Тематика современного анализа данных охватывает весьма широкий диапазон. В современных условиях применяется термин «машинное обучение», который понимается широко и охватывает многие разделы различных дисциплин, в том числе и в области дискретной математики.

Методы дискретной математики в машинном обучении оказываются особенно актуальными при анализе малых выборок, а также, когда экспериментальные данные представляют собой информацию о сложноорганизованных структурах. Проблематика исследований заключается в том, чтобы показать, как и какие содержательные задачи анализа данных трансформируются в задачи из области дискретной математики. В общем случае, задачи анализа данных, решаемые методами дискретной математики, оказываются задачами переборного типа. При их решении особое значение приобретают формальные отношения на элементах дискретного множества и соответствующих подмножеств.

Актуальность работы определяется тем, что в ней рассмотрены задачи поиска частых и нечастых элементов в частично упорядоченных данных, к которым сводится ряд центральных задач логического анализа данных (анализа информации методами дискретной математики).

В прикладных областях данные обычно естественным образом упорядочены, но многие традиционные алгоритмы их обработки не учитывают отношения порядка. В связи с этим возникает необходимость разработки эффективных методов поиска частых и нечастых элементов для данных, представленных в виде декартова произведения конечных частично упорядоченных множеств. Известно, что такие методы были впервые применены при поиске ассоциативных правил и наиболее изучены в случае декартова произведения бинарных антицепей.

Анализ данных, представленных декартовым произведением частично упорядоченных множеств, сводится к одной из центральных труднорешаемых перечислительных проблем дискретной математики, которая хорошо известна как задача дуализации над произведением частичных порядков.

При построении эффективных перечислительных алгоритмов большое значение имеет время выполнения одного шага, которое оказывается полиномиальным лишь в некоторых частных случаях решения задачи дуализации. Известен инкрементальный подход, когда время шага алгоритма превышает полиномиальную оценку, но оказывается существенно меньше экспоненциальной. Этот подход, в основном, имеет теоретическое значение.

Другой подход, предложенный и развиваемый в исследованиях руководителя диссертанта д.ф.-м.н. Е.В. Дюковой, заключается в построении асимптотически оптимальных алгоритмов, которые в настоящее время лидируют по скорости обработки практических задач.

Диссертантом разработаны эффективные в «типичном случае» (для почти всех вариантов задачи) алгоритмы совместного перечисления максимальных частых и минимальных нечастых элементов декартова произведения частично упорядоченных множеств. Показано, что задача совместного перечисления указанных элементов частичных порядков может рассматриваться как известная в дискретной математике задача определения нулей (расшифровки) двузначной монотонной функции.

Решение задачи поиска частых и нечастых элементов рассматривается в данной работе как основа для решения актуальной задачи классификации по прецедентам. Исследуются подходы, ориентированные на синтез так называемых корректных алгоритмов, не ошибающихся на обучающей выборке. Исследования в этой области также ведутся д.ф.-м.н. Е.В. Дюковой в рамках научного направления, созданного в работах чл.-корр. РАН С.В. Яблонского и акад. РАН Ю.И. Журавлева.

Основные результаты и их научная новизна. В диссертации разработаны вычислительно-эффективные методы логического анализа информации, представленной в виде декартова произведения конечных частично упорядоченных множеств. Основные результаты заключаются в постановке и решении следующих задач.

1. Решена задача перечисления множеств максимальных частых и минимальных нечастых элементов декартова произведения частичных порядков на основе последовательно-совместного подхода к построению перечислительного алгоритма, теоретически доказана его корректность (утверждения 1–3) и экспериментально показана эффективность, когда мощности множеств частых и нечастых элементов примерно одинаковы. Предложенный подход является новым.

2. Показано, что к задаче перечисления множеств максимальных частых и минимальных нечастых элементов декартова произведения частичных порядков сводится задача расшифровки двузначной монотонной функции. Для ее решения построен асимптотически оптимальный алгоритм и доказана его корректность (утверждения 4–6). Экспериментально показано, что при задании такой функции на n -мерном k -значном кубе алгоритм работает быстрее известных при больших размерностях n . Разработанный алгоритм является новым.

3. Для решения задачи классификации по прецедентам разработаны быстрые алгоритмы логической классификации на основе корректных ЭК специального вида для декартова произведения антицепей (алгоритм REC) и для декартова произведения цепей (алгоритм REC+). Получены асимптотические оценки типичного числа и типичного ранга правильных ЭК для антицепей (теорема 1) и цепей (теоремы 2 и 3), позволяющие теоретически обоснованно выбирать параметры алгоритмов REC и REC+. Разработанные алгоритмы являются новыми.

Содержание работы. Диссертационная работа состоит из введения, трех глав, заключения и списка литературы из 75 наименований.

Введение. Обоснована актуальность темы диссертации, введены необходимые определения и обозначения, рассмотрены известные постановки задач перечисления частых (нечастых) элементов как задач монотонной дуализации и дуализации над декартовым произведением частичных порядков.

Рассмотрены два вида известных алгоритмов решения данных задач, а именно, инкрементальных алгоритмов квазиполиномиальной монотонной дуализации и алгоритмов

асимптотически оптимальной дуализации. Сделан вывод о практической неприменимости алгоритмов первого вида и актуальности разработки алгоритмов второго вида.

Сформулирована задача расшифровки монотонной двужанной функции, определенной на декартовом произведении частично упорядоченных множеств. Ранее указанная задача рассматривалась для декартова произведения цепей и антицепей. Сделан вывод о необходимости альтернативных, т.е. оптимальных не по Шеннону алгоритмов, решающих задачу расшифровки. Наряду с оптимальными по Шеннону (оптимальными в худшем случае) алгоритмами предложено строить асимптотически оптимальные алгоритмы, эффективные в среднем. Показана связь между поставленной задачей и поиском максимальных частых и минимальных нечастых элементов частичных порядков.

Рассмотрена задача классификации по прецедентам над произведением частичных порядков и предложено ее решение на основе перечисления частых элементов в данных.

Сформулирована основная цель исследования как разработка вычислительно эффективных методов логического анализа информации, представленной в виде декартова произведения частично упорядоченных конечных множеств, и применения этих методов в области дискретной математики и машинного обучения.

Глава 1. Исследована задача поиска максимальных частых и минимальных нечастых элементов в данных, представленных декартовым произведением конечных частично упорядоченных множеств. Выполнен обзор известных алгоритмов поиска частых элементов. Показана связь задачи перечисления максимальных частых и минимальных нечастых элементов с задачей дуализации над произведением частичных порядков.

Разработан новый последовательно-совместный алгоритм перечисления множеств максимальных частых и минимальных нечастых элементов, опирающийся на тривиальный последовательный и предложенный в работах Эльбассиони (2014 г.) совместный методы перечисления. Доказана корректность нового алгоритма. Для данных, представленных произведением конечных цепей, выполнено экспериментальное сравнение последовательного, совместного и последовательно-совместного методов.

Показано, что последовательно-совместный метод эффективен, когда мощности множеств частых и нечастых элементов сопоставимы. При существенном различии мощностей этих множеств наиболее эффективным оказывается последовательный метод.

Глава 2. Исследована задача расшифровки (вычисления нулей) двужанной монотонной функции, заданной оракулом и определенной на декартовом произведении конечных частично упорядоченных множеств, которая является одной из важных задач дискретной математики. Показана связь задачи расшифровки монотонной функции и задачи совместного перечисления максимальных частых и минимальных нечастых элементов декартова произведения частичных порядков.

Известные решения задачи расшифровки двужанной монотонной функции основаны на оптимальных по Шеннону алгоритмах, которые имеют минимальную сложность в «худшем случае». Такие алгоритмы существенно ограничены в практическом применении. Поэтому актуальна разработка альтернативных подходов, более эффективных для большинства практических задач.

Для решения задачи расшифровки построен и исследован алгоритм перечисления на основе т.н. последовательно-совместного подхода, разработанного в главе 1. Данный подход

основан на применении асимптотически оптимального алгоритма дуализации декартова произведения цепей RUNC-M+, построенного Дюковой Е.В., Масляковым Г.О. и Прокофьевым П.А. (2019 г.), и позволяет достичь высокой эффективности в «типичном случае».

Глава 3. Исследована задача корректной классификации над произведением частичных порядков. Разработан новый подход к логической классификации, основанный на поиске в описаниях прецедентов каждого класса часто встречающихся фрагментов специального вида, т.н. правильных представительных ЭК. Предложена двухэтапная процедура их построения.

Для декартова произведения антицепей разработан алгоритм REC перечисления правильных представительных ЭК. Алгоритм имеет полиномиальное время шага. Разработана его модификация REC+ для декартова произведения конечных цепей.

Для антицепей получена верхняя асимптотическая оценка числа правильных представительных ЭК. Показано, что ранг правильного представительного ЭК не превосходит некоторой параметрической величины, что позволяет теоретически обоснованно выбирать ранг искомых ЭК. Сравнение с известными алгоритмами случайного леса и логистической регрессии показало, что REC в среднем не уступает им по качеству, а в ряде задач превосходит их. Для цепей также получены аналогичные теоретические оценки типичного числа и ранга правильных ЭК.

Выполнено экспериментальное сравнение алгоритмов REC и REC+ на шести реальных задачах. Показано, что REC работает быстрее других хорошо известных логических классификаторов и на большинстве задач превосходит их по качеству классификации. Показано, что задание специальных линейных порядков на множествах значений признаков позволяет повысить качество классификации без привлечения дополнительной информации о выборке. Классификатор REC+ превосходит по качеству классификатор REC на пяти задачах из шести.

В диссертации сделан вывод о том, что были разработаны вычислительно-эффективные методы логического анализа информации, представленной в виде декартова произведения конечных частично упорядоченных множеств.

Достоверность и обоснованность результатов. Достоверность результатов диссертации подтверждена строгими математическими доказательствами (гл. 1: утверждения 1–3, гл. 2: утверждения 4–6, гл. 3: леммы 1–18, теоремы 1–4), экспериментальной проверкой разработанных алгоритмов на шести реальных задачах. Основные положения, выносимые на защиту, представлены 11 публикациями в рецензируемых научных изданиях, в том числе в пяти изданиях, рекомендованных ВАК. Основные результаты диссертации докладывались на шести конференциях в 2019–2025 гг.

Теоретическая и практическая значимость. Исследованные в диссертации методы нацелены на данные, представленные декартовым произведением конечных частичных порядков. Разработанные алгоритмы проверены на синтетических и реальных данных.

Новый алгоритм последовательно-совместного перечисления максимальных частых и минимальных нечастых элементов оказывается эффективнее известных алгоритмов последовательного и совместного перечисления, когда мощности множеств частых и нечастых элементов не сильно отличаются.

Алгоритм асимптотически оптимальной расшифровки двузначной монотонной функции над произведением цепей работает в среднем быстрее ориентированного на худший случай алгоритма В.Б. Алексеева (1976 г.) и требует меньшего числа обращений к оракулу.

Классификаторы REC и REC+ требуют меньше вычислительных ресурсов, чем классические модели логического распознавания, в ряде случаев превосходя их в качестве распознавания.

Получены асимптотические оценки типичного числа и ранга правильных представительных ЭК, которые показывают, что число правильных представительных ЭК существенно меньше, чем число тупиковых ЭК, поиск которых ведётся в логическом подходе к обработке информации. Таким образом, эффективность алгоритмов REC и REC+ подтверждается как экспериментально, так и теоретически. Полученные оценки позволяют обоснованно выбирать параметры алгоритмов (ранги искомых ЭК), не прибегая к перебору.

Замечания.

1. Стр. 51. По-видимому, определение s -частого элемента x противоречит определениям на стр. 3 и стр. 24.
2. Стр. 56. На рис. 6 и 7 значение s в шапке диаграммы не соответствует значению в подрисуночной подписи.
3. Стр. 81. (Табл. 4). Известно, что одним из признанных показателей качества классификации является мера F_1 , которая специально применяется для несбалансированных классов. Данная мера не является симметричной и имеет разные значения в зависимости от того, какой класс считается положительным прогнозом против остальных классов, как отрицательных прогнозов. По-видимому, применение данной меры позволило бы точнее оценить результаты испытаний алгоритмов в данной работе.

Принципиальных и иных замечаний по содержанию работы, стилю и изложению материала нет. Все указанные выше замечания не влияют на общую положительную оценку диссертационной работы.

Заключение по диссертации

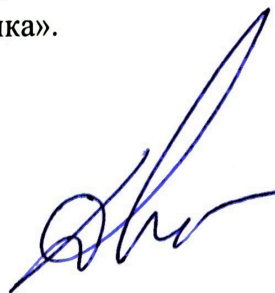
Диссертационная работа Н. А. Драгунова содержит решение актуальной задачи разработки вычислительно-эффективных методов логического анализа информации, представленной в виде декартова произведения конечных частично упорядоченных множеств, что соответствует паспорту специальности 1.2.3.

Основные результаты диссертации обладают научной новизной, имеют важное теоретическое и практическое значение. Они полностью представлены в 11 публикациях, в том числе в 5 статьях в изданиях из перечня ВАК и приравненных к нему международных наукометрических баз (Scopus, WoS), докладах на 6 международных и российских конференциях в 2019–2025 гг. Автореферат адекватно отражает содержание диссертации.

Представленная диссертация является научной квалификационной работой, основные результаты которой следует рассматривать как решение актуальных задач в области теоретической информатики и кибернетики с приложениями в дискретной математике и машинном обучении.

Диссертационная работа удовлетворяет всем требованиям и критериям Положения ВАК РФ «О порядке присуждения ученых степеней», утвержденного постановлением Правительства РФ от 24.09.2013 № 842, предъявляемым к диссертациям на соискание ученой степени кандидата физико-математических наук, а ее автор, Никита Аркадьевич Драгунов, заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 1.2.3 – «Теоретическая информатика, кибернетика».

Официальный оппонент,
профессор, д.ф.-м.н., профессор
ФГБОУ ВО «Тульский
государственный университет»

 С.Д. Двоенко

Адрес: 300012, Тула, пр. Ленина, 92, ТулГУ.

10. 09.2026

Тел.: 8 4872 257940

E-mail: info@tsu.tula.ru

*Тодманов
Игорь*

Двоенко



Драгунов

10.09.2026