



УТВЕРЖДАЮ

Директор ФИЦ ИУ РАН

академик РАН

И.А. Соколов

«17» ОКТАБРЯ 2018 г.

ЗАКЛЮЧЕНИЕ

Федерального государственного учреждения

«Федеральный исследовательский центр «Информатика и управление»

Российской академии наук (ФИЦ ИУ РАН)»

Диссертационная работа Виноградова Дмитрия Вячеславовича «Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток», представленная на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 - «Теоретические основы информатики», выполнена в отделе «Интеллектуальный анализ данных и поддержка научных исследований» Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН).

Соискатель – Виноградов Дмитрий Вячеславович — в 1986 году окончил механико-математический факультет Московского государственного университета им. М.В. Ломоносова по специальности «математика». Диссертацию на соискание ученой степени кандидата физико-математических наук «Метаматематические исследования правдоподобных рассуждений типа ДСМ» по специальности 05.13.17 «Теоретические основы информатики» (диплом КТ № 025616) защитил в 2000 году в диссертационном совете, созданном во Всероссийском институте научной и технической информации Российской академии наук (ВИНИТИ РАН). С 2016 года является старшим научным сотрудником ФИЦ ИУ РАН в отделе «Интеллектуальный анализ данных и поддержка научных исследований».

Научный консультант – доктор технических наук, заслуженный деятель науки РФ, профессор Финн Виктор Константинович. Работает в отделе «Интеллектуальный анализ данных и поддержка научных исследований» ФИЦ ИУ РАН с 2016 года, главный научный сотрудник.

По итогам обсуждения принято следующее заключение:

Диссертация Виноградова Дмитрия Вячеславовича «Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток» представляет собой цельное и законченное научное исследование, посвященное математическому исследованию вероятностных алгоритмов машинного обучения с использованием идей прикладной теории решеток.

Тема диссертации является **актуальной** для развития методов машинного обучения, предназначенных для интеллектуального анализа данных большого объема. Подобные массивы данных естественно возникают в современной науке и в практических областях человеческой деятельности при использовании современных средств сбора, хранения и обработки данных.

Цели и задачи диссертации достигнуты:

1. Предложить общий вероятностный подход к интеллектуальному анализу данных на основе прикладной теории решеток.
2. Разработать вероятностные алгоритмы машинного обучения, с использованием понятий и результатов анализа формальных понятий (АФП) – современного раздела теории решеток.
3. Исследовать математические свойства предложенных автором вероятностных алгоритмов.

Основные результаты диссертации четко сформулированы:

1. Доказательство корректности алгоритма представления объектов с помощью битовых строк и оценка вычислительной сложности этого алгоритма.
2. Оценки вероятностей переобучения для логико-комбинаторного метода вычисления всех сходств обучающих примеров.
3. Оценка эффективности ленивой схемы вероятностных вычислений для поиска сходств.
4. Оценки среднего времени вычислений сходства и скорости перемешивания случайных блужданий при его поиске (для случая Булевой алгебры).
5. Оценка в метрике тотальной вариации различия между распределения вероятностей выдачи результатов спаривающими цепями Маркова, неостановленной и остановленной по сумме длин предварительно вычисленных траекторий.
6. Оценка достаточного числа сходств для правильного доопределения по аналогии с заданной надежностью всех примеров, содержащих причины с вероятностью, не ниже заданной (аналог результатов Вапника-Червоненкиса).

Научная новизна работы представляется несомненной:

1. Вероятностный подход к машинному обучению на основе прикладной теории решеток ранее не исследовался. Предыдущие подходы (ДСМ-метод и анализ формальных понятий) были детерминированными.

2. Открытие феномена «переобучения» для логико-комбинаторного метода вычисления всех сходств является новым.

Личное участие автора в получении научных результатов несомненно:

Все результаты диссертации получены автором самостоятельно и опубликованы с достаточной полнотой в ведущих рецензируемых отечественных и зарубежных журналах. Совместные работы автора представляют собой лишь приложения изложенных в диссертации идей соискателя в различных прикладных задачах.

Достоверность результатов и выводов:

1. Математическим фундаментом для проведенных исследований являются как классические методы дискретной математики, теории вероятностей и случайных процессов, так и современные методы прикладной теории решеток.
2. Все математические утверждения приводятся с подробными и строгими доказательствами.
3. Выводы диссертации подтверждаются как строгими математическими результатами, так и компьютерными экспериментами над массивами из репозитория данных для тестирования алгоритмов машинного обучения (UCI MLR).

несомненна.

Практическая ценность экспериментально обоснована:

1. Предложенные в работе методы и алгоритмы можно использовать для интеллектуального анализа данных в больших массивах, а также данных, подвергающихся искажению с помощью некоторого вероятностного механизма.

2. Проверка теоретических идей в программной ВКФ-системе, примененной к массивам данных из репозитория UCI MLR, продемонстрировала превосходство предложенного подхода над некоторыми комбинаторными алгоритмами анализа данных, а также возможность их применения к данным большого объема.

Апробация результатов диссертации включает в себя:

1. Результаты работы неоднократно докладывались на научных семинарах ФИЦ ИУ РАН, ВИНТИ РАН, МФТИ, РГГУ, а также на 10 международных и всероссийских конференциях.
2. По материалам диссертационной работы автором были опубликованы более 20 научных работ, из которых требованиям ВАК удовлетворяют 17.

3. Часть результатов диссертации вошли в отчеты по проектам РФФИ и программ Президиума РАН.
4. Основные результаты диссертационного исследования составляют часть учебных курсов «Теория сходства в интеллектуальных системах» и «Интеллектуальный анализ данных и машинное обучение», которые Д.В. Виноградов читает студентам отделения интеллектуальных систем в гуманитарной сфере РГГУ.

Содержание диссертации соответствует паспорту специальности 05.13.17 - «Теоретические основы информатики»:

1. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных (п. 5)
2. Моделирование формирования эмпирического знания (п. 7)
3. Разработка методов обеспечения высоконадежной обработки информации (п. 11)

Согласно формуле специальности «Теоретические основы информатики» к ней относятся, в числе прочего, «... исследования методов преобразования информации в данные и знания; создание и исследование ... методов машинного обучения и обнаружения новых знаний». Таким образом, исследование вероятностно-комбинаторного метода обучения, основанного на бинарной операции сходства, соответствует данной специальности.

Публикации автора:

По материалам диссертационной работы автором были опубликованы более 20 научных работ, из которых требованиям ВАК удовлетворяют 17. Наиболее значимые работы:

1. Galitsky, B.A., Kuznetsov, S.O. and Vinogradov, D.V. Applying hybrid reasoning to mine for associative features in biological data // Journal of Biomedical Information, Vol. 40, - Issue 3. - 2007. - p. 203-220
2. Панкратова Е.С., Виноградов Д.В. Формальное описание настройки интеллектуальных ДСМ-систем на область клинической и лабораторной диагностики // Научная и техническая информация, Сер. 2. - 2011. - № 9. - С. 1-5
3. Виноградов Д.В. Вероятностное порождения гипотез в ДСМ-методе с помощью простейших цепей Маркова // Научная и техническая информация, Сер. 2. - 2012. - № 9. - С. 20-27
4. Виноградов Д.В. Автоматическое порождение гипотез в ДСМ-методе с помощью цепей Маркова // В кн.: Труды 13 национальной конференции по искусственному интеллекту КИИ-2012, Т. 2. - 2012. - С. 121-127
5. Vinogradov, D.V. A Markov chain approach to random generation of formal concepts // Proceedings of the Workshop "Formal Concept Analysis Meets Information Retrieval (FCAIR 2013)": CEUR Workshop Proceedings, Vol. 977. - 2013. - p. 127-133

6. Виноградов Д.В. Качество вероятностно порожденных ДСМ-гипотез // В кн.: Материалы 6 всероссийской мультikonференции по проблемам управления (МКПУ-2013) Т. 1. - 2013. - С. 14-16
7. Виноградов Д.В. ВКФ-метод порождения гипотез: программная реализация // В кн.: Труды 14 национальной конференции по искусственному интеллекту КИИ-2014, Т. 2. - 2014. - С. 252—258
8. Vinogradov, D.V. VKF-method of hypotheses generation // Communications in Computer and Information Science, Vol. 436. - 2014. - p. 237-248
9. Виноградов Д.В. Вероятностно-комбинаторный подход к автоматическому порождению гипотез // В кн.: Гуманитарные чтения РГГУ—2014. - М.: РГГУ. - 2015. - С. 771-775
10. Виноградов Д.В. Вероятность порождения случайного ДСМ-сходства при наличии контр-примеров // Научная и техническая информация, Сер. 2. - 2015. - № 3. - С. 1-5
11. Виноградов Д.В. Сильная концентрация времени работы алгоритма поиска сходств. В кн.: Материалы 8 всероссийской мультikonференции по проблемам управления (МКПУ-2015) Т. 1. - 2015. - С. 42-45
12. Виноградов Д.В. Предельная вероятность порождения случайного сходства при наличии контр-примеров // Научная и техническая информация, Сер. 2. - 2017. - № 2. - С. 17-19
13. Виноградов Д.В. Эффективность ленивых вычислений для поиска сходств в ВКФ-системе // Научная и техническая информация, Сер. 2. - 2017. - № 4. - С. 19-23
14. Виноградов Д.В. Анализ результатов применения ВКФ-системы: успехи и открытая проблема // Научная и техническая информация, Сер. 2. - 2017. - № 5. - С. 1-4
15. Виноградов Д.В. ВКФ-метод интеллектуального анализа данных: обзор результатов и открытых проблем // Искусственный интеллект и принятие решений. - 2017. - № 2. - С. 9-16
16. Vinogradov, D.V. Accidental formal concepts in the presence of counterexamples // Proceedings of International Workshop on Formal Concept Analysis for Knowledge Discovery (FCA4KD 2017): CEUR Workshop Proceedings, Vol. 1921. - 2017. - p. 104-112
17. Виноградов Д.В. Надежность предсказания по аналогии // Научная и техническая информация, Сер. 2. - 2017. - № 7. - С. 11-15
18. Виноградов Д.В. О надежном предсказании ВКФ-гипотезами. В кн.: Материалы 10 всероссийской мультikonференции по проблемам управления (МКПУ-2017) Т. 1. - 2017. - С. 48-50
19. Виноградов Д.В. Скорость сходимости к пределу вероятности порождения случайного сходства при наличии контр-примеров // Научная и техническая информация, Сер. 2. - 2018. - № 2. - С. 21-24

20. Виноградов Д.В. Учет предварительных оценок скорости порождения сходств спаривающей цепью Маркова // Информатика и ее применения. - 2018. - № 1. - С. 50-55
21. Виноградов Д.В. О представлении объектов битовыми строками для ВКФ-метода // Научная и техническая информация, Сер. 2. - 2018. - № 5. - С. 1-4
22. Vinogradov, D.V. Machine learning based on similarity operation // Communications in Computer and Information Science, Vol. 934. - 2018. - p. 46-59

ВЫВОД:

Диссертационная работа Виноградова Дмитрия Вячеславовича «Вероятностно-комбинаторный формальный метод обучения, основанный на операции сходства» представляет собой цельное и законченное научное исследование, имеющее высокую научную и практическую значимость. Все положения, представленные к защите, являются новыми и научно обоснованными. Результаты диссертации были получены автором самостоятельно и опубликованы с достаточной полнотой в ведущих рецензируемых отечественных и зарубежных журналах, которые соответствуют требованиям ВАК. Совокупность теоретических результатов, полученных автором в диссертационном исследовании, можно квалифицировать как открытие нового научного направления в области теоретической информатики — вероятностных методов обучения, основанных на теории решеток. Таким образом, диссертационная работа Д.В. Виноградова «Вероятностно-комбинаторный формальный метод обучения, основанный на теории решеток» рекомендуется к защите на соискание ученой степени доктора физико-математических наук по специальности 05.13.17 — «Теоретические основы информатики».

Заключение принято на заседании научного семинара отдела № 16 Федерального исследовательского центра «Информатика и управление» Российской академии наук, на котором присутствовало 8 человек.

Результаты голосования: «за» — 8 чел., «против» — нет, «воздержались» — нет.

Протокол № 2 от 16 сентября 2018 года

Председатель _____ д.ф.-м.н., с.н.с. Забейайло М.И.

Секретарь _____ к.т.н. Фабрикантова Е.Ф.