



УТВЕРЖДАЮ

Проректор Национального
исследовательского университета
«Высшая школа экономики»

к.э.н., доцент Рощин Сергей Юрьевич

« 10 » июля 2017 г.

ЗАКЛЮЧЕНИЕ
федерального государственного автономного
образовательного учреждения высшего образования
«Национальный исследовательский университет
«Высшая школа экономики»

Диссертация Кашницкого Юрия Савельевича на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой» выполнена в департаменте анализа данных и искусственного интеллекта федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики».

В период подготовки диссертации соискатель ученой степени Кашницкий Юрий Савельевич работал в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики» в Международной научно-учебной лаборатории интеллектуальных систем и структурного анализа с 01.11.2013 г. по настоящее время в должности стажера-исследователя.

В 2013 г. окончил федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (государственный университет)» по специальности «прикладная математика и информатика».

В 2016 г. окончил очную аспирантуру федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики» по научной специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ».

В октябре 2015 года Кашницкий Ю.С. был прикреплен в качестве экстерна для сдачи кандидатского экзамена по специальности 05.13.17 «Теоретические основы информатики». В мае 2017 года был прикреплен в качестве экстерна для сдачи кандидатских экзаменов по предметам «Иностранный язык» и «История и философия науки».

Справка об обучении с результатами сдачи кандидатских экзаменов выдана в 2017 г. в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики».

Научный руководитель – д.ф.-м.н., профессор Кузнецов Сергей Олегович – работает в федеральном государственном автономном образовательном учреждении высшего образования «Национальный исследовательский университет «Высшая школа экономики» в должности руководителя департамента анализа данных и искусственного интеллекта.

По итогам обсуждения принято следующее заключение.

Актуальность темы

В последнее время в машинном обучении все большую популярность приобретают методы анализа сложно структурированных данных. Такие задачи встречаются в химической информатике, анализе текстов и изображений, когда надо работать с данными, представленными последовательностями или графами. В задачах классификации подобных

данных лучше всего себя проявляет метод опорных векторов с ядрами. Однако этот метод неинтерпретируемый, по выводу алгоритма практически невозможно понять, почему тот или иной объект так или иначе классифицирован.

Для решения задач классификации сложно структурированных данных требуются интерпретируемые алгоритмы, основанные на правилах, с которыми далее могут работать эксперты.

Обоснованность научных положений

Достоверность научных результатов и обоснованность научных положений и рекомендаций, содержащихся в диссертации, основаны на использовании трудов ведущих отечественных и зарубежных исследователей в области классификации сложно структурированных данных. Разработка математического аппарата, лежащего в основе предлагаемых алгоритмов, опиралась на труды международного сообщества по анализу формальных понятий и получила его одобрение.

Результаты диссертационного исследования докладывались и обсуждались на российских и международных научно-практических конференциях и получили одобрение научной общественности.

Личное участие соискателя ученой степени в получении результатов, изложенных в диссертации

Личный вклад автора состоит в едином представлении методов классификации сложно структурированных данных с помощью узорных структур и их проекций, в предложении специального вида проекций узорных структур для данных с количественными признаками, обобщающего подход к обучению на основе деревьев решений, а также в создании программного комплекса, реализующего предложенные алгоритмы классификации данных, представленных последовательностями и графами.

Степень достоверности результатов проведенных соискателем ученой степени исследований

Полученные результаты согласуются с опубликованными в литературе данными других исследователей и не противоречат данным, представленным в независимых источниках по данной тематике. В работе использованы современные библиотеки, реализованные на языках программирования Java и Python: SPMF, Gaston и Scikit-learn.

Научная новизна работы

Наиболее существенные научные результаты, полученные лично автором, заключаются в том, что предложен новый эффективный подход к классификации данных со сложной структурой на основе узорных структур, а также предложен специальный вид проекций узорных структур для данных с количественными признаками, обобщающий подход к обучению на основе деревьев решений. Кроме того, создан комплекс программ для классификации данных со сложной структурой на основе решеток замкнутых описаний. Соответствующие алгоритмы были апробированы на многих наборах данных с категориальными и количественными признаками, а также на данных по токсичности химических веществ со сложной структурой в виде молекулярных графов.

Теоретическая значимость

Теоретическая значимость работы состоит в едином представлении методов классификации сложно структурированных данных с помощью узорных структур и их проекций. Методы классификации числовых данных, в том числе деревьев решений, представлены с помощью проекций интервальных узорных структур.

Выработан подход к классификации на основе правил, гарантирующий нахождение правил с лучшим выбранным критерием, чем правила, полученные с помощью деревьев решений, и имеющий вычислительную сложность меньшую, чем у алгоритма случайного

леса. Для этого введены и исследованы дискретизирующие проекции для интервальных узорных структур.

Практическая значимость результатов проведенных соискателем ученой степени исследований

Разработанные соискателем алгоритмы анализа данных в виде графов и последовательностей могут быть использованы на практике для получения наборов классифицирующих правил, которые могут быть интерпретированы человеком.

Результаты диссертационной работы также использовались при проведении семинарских занятий по курсу «Машинное обучение и анализ данных» для студентов бакалавриата факультета компьютерных наук Национального исследовательского университета «Высшая школа экономики», а также в программах дополнительного профессионального образования Национального исследовательского университета «Высшая школа экономики».

Апробация работы

1. Третья международная конференция по анализу изображений, сетей и текстов АИСТ 2014. Екатеринбург, Россия, апрель 2014 г. Доклад на тему: «Visual analytics in FCA-based triclustering».
2. Семинар «What can FCA do for Artificial Intelligence?» при Европейской конференции по искусственному интеллекту ECAI. Прага, Чехия, июль 2014 г. Доклад на тему: «Can FCA-based recommender system suggest a proper classifier?».
3. Ph.D.-семинар при Европейской конференции по машинному обучению и теоретическим основам и практике обнаружения знаний в базах данных ECML/PKDD. Нанси, Франция, июль 2014. Доклад на тему: «Recommender-based Multiple Classifier System».
4. Семинар «What can FCA do for Artificial Intelligence?» при международной объединенной конференции по искусственному интеллекту IJCAI. Буэнос-Айрес, Аргентина, июль 2015 г. Доклад на тему: «Lazy Associative Graph Classification».
5. Семинар «What can FCA do for Artificial Intelligence?» при международной объединенной конференции по искусственному интеллекту IJCAI. Буэнос-Айрес, Аргентина, июль 2015 г. Доклад на тему: «Lazy Classification with Interval Pattern Structures: Application to Credit Scoring».
6. Семинар «What can FCA do for Artificial Intelligence?» при Европейской конференции по искусственному интеллекту ECAI. Гаага, Нидерланды, сентябрь 2016. Доклад на тему: «Interval Pattern Concept Lattice as a Classifier Ensemble».
7. Пятая международная конференция по анализу изображений, сетей и текстов АИСТ 2016, г. Екатеринбург, Россия, апрель 2016 г. Доклад на тему: «Lazy Learning of Succinct Classification Rules for Complex Structure Data».
8. Пятнадцатая национальная конференция по искусственному интеллекту с международным участием КИИ-2016. Смоленск, Россия, октябрь 2016 г. Доклад на тему: «Методы поиска точных и интерпретируемых классифицирующих правил для данных со сложной структурой».
9. 13-ая международная конференция по решеткам понятий и их приложениям (The 13th International Conference on Concept Lattices and Their Applications). Москва, Россия, июль 2016. Доклад на тему: «Global Optimization in Learning with Important Data: an FCA-Based Approach».
10. Семинар Департамента анализа данных и искусственного интеллекта НИУ ВШЭ, Москва, Россия, октябрь 2015 г. Доклад на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой».
11. Семинар Департамента анализа данных и искусственного интеллекта НИУ ВШЭ, Москва, Россия, июнь 2016 г. Доклад на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой».

12. Семинар Департамента анализа данных и искусственного интеллекта НИУ ВШЭ, Москва, Россия, март 2016 г. Доклад на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой».
13. Конференция «Технологии Больших Данных» (ТБД-2016). Москва, Россия, июнь 2016. Доклад на тему: «Локальная интерпретируемость алгоритмов машинного обучения».
14. Семинар отдела Интеллектуальных систем ВЦ РАН им. А.А. Дородницына. Москва, 20.10.2016 г. Доклад на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой».

Полнота изложения материалов диссертации в публикациях

Основные результаты диссертационного исследования опубликованы в 11 работах общим объемом 5,55 п.л.; личный вклад автора составляет 4,35 п.л.

Работы, опубликованные автором в ведущих рецензируемых научных журналах и изданиях, рекомендованных ВАК:

1. Кашницкий Ю.С. Визуальная аналитика в задаче трикластеризации многомерных данных // Труды МФТИ. Труды Московского физико-технического института (государственного университета). 2014. Т. 6. № 3. С. 43-56 – 0,5 п.л.
2. Кашницкий Ю.С. Неточные множества для оптимизации SQL-запросов / Slezak D., Кузнецов С.О. // Открытые системы. СУБД. 2014. № 10. С. 44-45 – 0,15 п.л. (личный вклад автора – 0,1 п.л.)

Работы, опубликованные автором в рецензируемых научных изданиях, входящих в международные реферативные базы данных и системы цитирования :

3. Kashnitsky Y. Global Optimization in Learning with Important Data: an FCA-Based Approach // Proceedings of the 13th International Conference on Concept Lattices and Their Applications. 2016. Vol. 1624, ch. 19. P. 189-201 (Scopus) – 0,5 п.л. (личный вклад автора – 0,4 п.л.)
4. Kashnitsky Y. Interval Pattern Concept Lattice as a Classifier Ensemble / Kuznetsov S.O. // Proceedings of the 5th International Workshop «What can FCA do for Artificial Intelligence?». 2016. P. 105-112 (Scopus) – 0,45 п.л. (личный вклад автора – 0,4 п.л.)
5. Kashnitsky Y. Lazy Associative Graph Classification / Kuznetsov S.O. // Proceedings of the 4th International Workshop «What can FCA do for Artificial Intelligence?». 2015. P. 63-74 (Scopus) – 0,55 п.л. (личный вклад автора – 0,45 п.л.)
6. Kashnitsky Y. Lazy Classification with Interval Pattern Structures: Application to Credit Scoring / Masyutin A., Kuznetsov S.O. // Proceedings of the 4th International Workshop «What can FCA do for Artificial Intelligence?». 2015. P. 43-54 (Scopus) – 0,55 п.л. (личный вклад автора – 0,2 п.л.)
7. Kashnitsky Y. Can FCA-based Recommender System Suggest a Proper Classifier? / Ignatov D.I. // Proceedings of the 3rd International Workshop «What can FCA do for Artificial Intelligence?». 2014. Vol. 1257, ch. 3. P. 17-26 (Scopus) – 0,4 п.л. (личный вклад автора – 0,35 п.л.)
8. Kashnitsky Y. Visual analytics in FCA-based clustering // Supplementary Proceedings of the 3rd International Conference on Analysis of Images, Social Networks and Texts. 2014. Vol. 1197, ch. 12. P. 69-80 (Scopus) – 0,45 п.л.

Работы, опубликованные автором в других изданиях:

9. Kashnitsky Y. Lazy Learning of Succinct Classification Rules for Complex Structure Data // Supplementary Proceedings of the 5th International Conference on Analysis of Images, Social Networks and Texts. 2016. Vol. 1710, ch. 8. P. 73-84 – 0,5 п.л.
10. Кашницкий Ю.С. Ансамблевый метод машинного обучения, основанный на рекомендации классификаторов / Игнатов Д.И. // Интеллектуальные системы. Теория и приложения. 2015. Т. 19. № 4. С. 37-55 – 0,6 п.л. (личный вклад автора – 0,5 п.л.)
11. Kashnitsky Y. Recommender-based Multiple Classifier System // The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in

Databases, PhD Session Proceedings. 2014. P. 21-30 – 0,5 п.л. (личный вклад автора – 0,5 п.л.)

Публикации полностью соответствуют теме диссертационного исследования и раскрывают её основные положения.

Ценность научных работ соискателя ученой степени

Ценность научных публикаций заключается в получении результатов, обладающих признаками научной новизны: в предложении единого подхода к классификации сложно структурированных данных, в разработке для них композиционных методов классификации и в теоретическом обобщении этих методов классификации с помощью аппарата анализа формальных понятий. Полученные в исследовании выводы дополняют научные представления о принципах и подходах к интерпретируемой классификации сложно структурированных данных.

Диссертация Кашницкого Юрия Савельевича на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой» – это законченная научно-квалификационная работа, которая соответствует требованиям пунктов 9, 10, 14 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, а также Паспорту научной специальности 05.13.17 Теоретические основы информатики (технические науки), в том числе по следующему пункту:

п. 5 – Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений (разработан метод классификации данных со сложной структурой на основе решающих правил, предложенные алгоритмы могут работать в том числе и с последовательностями и графами в приложениях по анализу текста, поведенческих паттернов и химических свойств веществ).

Диссертация Кашницкого Юрия Савельевича на тему: «Методы замкнутых описаний в задаче классификации данных со сложной структурой» рекомендуется к защите на соискание ученой степени кандидата технических наук по специальности 05.13.17 «Теоретические основы информатики» (технические науки).

Заключение принято на заседании департамента анализа данных и искусственного интеллекта федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики» 6 июня 2017 г., протокол № 2.3.6-01/06062017-05.

Присутствовало на заседании 5 человек. Результаты голосования: «за» – 4 человека, «против» – 0 человек, «воздержалось» – 1 человек.

Старший научный сотрудник
департамента анализа данных
и искусственного интеллекта,
к.т.н., доцент



Незнанов Алексей Андреевич

Подпись заверяю
СПЕЦИАЛИСТ ПО КАДРАМ 1 КАТ
КОНЕВСКАЯ Л.В. 
«03» июня 2017 г.

