

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА 24.2.403.01
НА БАЗЕ ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО
ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ НАУКИ
И ТЕХНОЛОГИЙ ИМЕНИ АКАДЕМИКА М.Ф. РЕШЕТНЕВА»
МИНИСТЕРСТВА НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ ПО ДИССЕРТАЦИИ НА СОИСКАНИЕ УЧЕНОЙ СТЕПЕНИ
КАНДИДАТА НАУК

аттестационное дело № _____

решение диссертационного совета от 28.03.2025 г. №6

о присуждении Ахматшину Фариду Галиуллиновичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

Диссертация «Модели и алгоритмы автоматической группировки объектов для систем анализа и хранения данных на основе методов семейства k -средних» по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика принята к защите 17.01.2025 протокол №4 диссертационным советом 24.2.403.01 на базе Федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» Министерства науки и высшего образования Российской Федерации (660037, г. Красноярск, просп. им. газ. «Красноярский рабочий», 31, приказ Минобрнауки от 06.06.2017 № 544/нк).

Соискатель Ахматшин Фарид Галиуллинович 1980 года рождения в 2003 году окончил специалитет Лесосибирского филиал СибГТУ, в 2024 году окончил очную аспирантуру Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева. Работает старшим преподавателем кафедры информационных и технических систем в филиале СибГУ им. М.Ф.Решетнева в г.Лесосибирске, Министерство науки и высшего

образования Российской Федерации, а также по совместительству работает инженером в научно-исследовательской лаборатории института информатики и телекоммуникаций СибГУ им. М.Ф.Решетнева.

Диссертация выполнена на кафедре системного анализа и исследования операций Федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева» Министерства науки и высшего образования Российской Федерации.

Научный руководитель – доктор технических наук, профессор Казаковцев Лев Александрович, Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, профессор кафедры системного анализа и исследования операций.

Официальные оппоненты:

Кравец Олег Яковлевич, доктор технических наук, профессор, ФГБОУ ВО «Воронежский государственный технический университет», профессор кафедры «Автоматизированные и вычислительные системы».

Муравьев Сергей Борисович, кандидат технических наук, ФГАОУ ВО «Национальный исследовательский университет информационных технологий, механики и оптики», доцент Института прикладных компьютерных наук.

Ведущая организация Федеральное государственное бюджетное образовательное учреждение высшего образования «Кемеровский государственный университет» в своем положительном отзыве, подписанном Каган Еленой Сергеевной, кандидатом технических наук, доцентом, заведующим кафедрой прикладной математики указала, что диссертационная работа Ахматшина Ф.Г. является завершенным научно-исследовательским трудом на актуальную тему, выполнена самостоятельно и на высоком научном уровне. Полученные автором результаты достоверны, выводы и заключения являются обоснованными. Особую ценность работе придает её направленность на решение практических задач систем анализа и хранения

данных, на повышение эффективности алгоритмов кластеризации при обработке больших данных в системах автоматической группировки объектов, в том числе в составе векторной СУБД и подсистем компрессии данных в составе систем хранения данных. Представленная диссертационная работа отвечает требованиям п.9 «Положения о порядке присуждения ученых степеней» Постановления правительства Российской Федерации от 24.09.2013 г. № 842, предъявляемым к кандидатским диссертациям, а её автор Ахматшин Фарид Галиуллович заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика.

Соискатель имеет по теме диссертации 15 научных работ, из них 5 статей, опубликованных в ведущих рецензируемых научных журналах и изданиях, рекомендованных ВАК, и 5 работ в изданиях, индексируемых в международных системах цитирования Scopus и Web of Science, в других изданиях 4, наличие одного свидетельства о государственной регистрации программы для ЭВМ. Научные работы посвящены вопросам нормализации данных, разработке алгоритма кластеризации, разработке процедуры инициализации центров кластеров для алгоритмов кластеризации, сохраняющей вычислительную эффективность при увеличении объема обрабатываемых данных.

Наиболее значимые из них:

1. Сокращение оценки усадки Джеймса-Штейна для решения задачи автоматической группировки однородных производственных партий / Ф. Г. Ахматшин, И. А. Петрова, Л. А. Казаковцев, И. Н. Кравченко // Проблемы машиностроения и надежности машин. – 2024. – № 3. – С. 65-75. – DOI 10.31857/S0235711924030093. (переводная версия: Akhmatshin F. G., Reducing the James–Stein Shrinkage Estimator for Automatically Grouping Heterogeneous Production Batches / F. G. Akhmatshin, I. A. Petrova, L. A. Kazakovtsev, I. N. Kravchenko // Journal of Machinery Manufacture and Reliability. – 2024. – Vol. 53, No. 3. – P. 254-262. – DOI 10.1134/S1052618824700043).

2. Ахматшин Ф. Г. О нормализации данных в задаче автоматической группировки промышленной продукции по однородным производственным партиям / Ф. Г. Ахматшин, И. Р. Насыров, В. Л. Казаковцев, Л. А. Казаковцев // Системы управления и информационные технологии. – 2020. – № 2(80). – С. 86-89.

3. Ахматшин Ф. Г. Подбор свободного параметра алгоритма FOREL-2 в задаче автоматической группировки промышленной продукции по однородным производственным партиям / Ф. Г. Ахматшин // Системы управления и информационные технологии. – 2021. – № 4(86). – С. 28-31. – DOI 10.36622/VSTU.2021.86.4.006.

4. Ахматшин Ф. Г. Алгоритм FOREL-2 с жадной эвристикой выбора радиуса поиска локальных сгущений / Ф. Г. Ахматшин, Л. А. Казаковцев // Системы управления и информационные технологии. – 2022. – № 3(89). – С. 39-42. – DOI 10.36622/VSTU.2022.89.3.009.

5. Ахматшин Ф. Г. О методе инициализации для алгоритмов кластеризации / Ф. Г. Ахматшин // Системы управления и информационные технологии. – 2024. – № 1(95). – С. 4-10.

6. Ахматшин Ф. Г. О сжатии блоков данных с использованием алгоритма кластеризации k-средних / Ф. Г. Ахматшин // Системы управления и информационные технологии. – 2024. – № 3(97). – С. 68-72.

7. Ahmatshin F. G., Kazakovtsev L. A. Mini-batch K-means++ clustering initialization. //XXIII International Conference Mathematical Optimization Theory and Operations Research MOTOR-2024 Omsk Russia, June 30 - July 06, 2024. – pp. 293-307.

8. Ahmatshin F. G. Impact of data normalization methods and clustering model in the problem of automatic grouping of industrial products / F. G. Ahmatshin, L. A. Kazakovtsev // Journal of Physics: Conference Series, Krasnoyarsk, Russian Federation, 25 сентября – 04 октября 2020 года. Vol. 1679. – Krasnoyarsk, Russian Federation: Institute of Physics and IOP Publishing Limited, 2020. – P. 32085. – DOI 10.1088/1742-6596/1679/3/032085.

9. Ahmatshin F. G. Greedy Heuristics for the Choice of the Radius of Local Concentrations in Forel-2 Algorithm / F. G. Ahmatshin, L. A. Kazakovtsev // Hybrid methods of modeling and optimization in complex systems : Proceedings of the International Workshop —Hybrid methods of modeling and optimization in complex systems (HMMOCS 2022), Krasnoyarsk, 22–24 November 2022. – London, United Kingdom: European Proceedings, 2023. – P. 366-371. – DOI 10.15405/epct.23021.45.

На диссертацию и автореферат поступили отзывы от:

1. доктора технических наук, Дулесова Александра Сергеевича, профессора кафедры «Цифровые технологии и дизайн» Хакасского государственного университета им. Н.Ф. Катанова;

2. доктора технических наук, Бронова Сергея Александровича, профессора кафедры «Информационные технологии и математическое обеспечение информационных систем» ФГБОУ ВО «Красноярский государственный аграрный университет»;

3. доктора физико-математических наук, Кочетова Юрия Андреевича, профессора, главного научного сотрудника лаборатории «Математические модели принятия решений» ФГБУН Институт математики им. С.Л. Соболева СО РАН;

4. доктора технических наук, Истомина Андрея Леонидовича, декана факультета управления и бизнеса ФГБОУ ВО «Ангарский государственный технический университет»;

5. кандидата технических наук, Сташкова Дмитрия Викторовича, ведущего специалиста по АСУ ТП горнорудной Промышленности АО «СИНЕТИК»;

6. кандидата физико-математических наук, Красикова Виталия Александровича, старшего научного сотрудника НОЦ ФНС России и МГТУ им. Н.Э. Баумана;

7. доктора технических наук, Ничепорчука Валерия Васильевича, старшего научного сотрудника отдела Прикладной информатики ИВМ СО РАН.

Все отзывы положительные. В замечаниях отмечены: отсутствие некоторых обозначений в автореферате, отсутствие пояснений о применяемом методе определения числа кластеров, отсутствие определения параметра тщательности сжатия, отсутствие пояснений по поводу выбора количества запусков алгоритмов в вычислительных экспериментах, отсутствие формального определения семейства алгоритмов k-средних. Замечания не носят критический характер и не касаются научной новизны и практической значимости диссертационной работы.

Выбор официальных оппонентов и ведущей организации обосновывается тем, что они являются специалистами в области разработки методов и алгоритмов для решения оптимизационных задач, задач автоматической группировки объектов, имеют публикации в данной области, а также в других областях системного анализа, управления и обработки информации.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований:

– *предложен* новый подход к нормализации данных для входных данных, используемых в системах анализа данных результатов неразрушающих испытаний образцов промышленной продукции, комбинирующий нормализацию по допустимым значениям оцениваемых характеристик продукции и оценку Джеймса–Штейна, который обеспечивает повышение точности решения задачи кластеризации на 10% по индексу Рэнда на примере задачи автоматической группировки электрорадиоизделий;

– *разработан* новый алгоритм кластеризации для системы анализа данных электрорадиоизделий на основе данных тестовых испытаний с использованием жадной эвристической процедуры выбора радиуса локальных концентраций по размеченным данным, который обеспечивает повышение

точности (по индексу Рэнда) и скорости получения результатов автоматической группировки по сравнению с алгоритмом k -средних;

– *разработан* новый алгоритм кластеризации для построения индекса векторной базы данных для приближенного поиска ближайших соседей, который обеспечивает компромисс между точностью и временем вычислений, существенно улучшает метрику полноты в задачах приближенного поиска ближайших соседей;

– *разработан* новый алгоритм автоматической группировки повторяющихся фрагментов блоков данных на основе алгоритма k -средних совместно с локально-чувствительным хэшированием (LSH), который обеспечивает увеличение эффективности сжатия данных в системах хранения данных;

– *предложена* новая процедура инициализации центров кластеров для алгоритмов автоматической группировки больших объемов данных, использует вспомогательную структуру данных – массив слагаемых для вычисления суммы квадратов расстояния, который снижает вычислительные затраты в сравнении с алгоритмом k -means++ без снижения качества получаемого начального решения.

Теоретическая значимость исследования обосновывается тем, что:

– *применительно к проблематике диссертации* известный инструментарий методов автоматической группировки объектов дополнен эффективными алгоритмами решения задач автоматической группировки, а также алгоритмом предобработки данных для таких задач.

Значение полученных соискателем результатов исследования для практики подтверждается тем, что

- разработанные алгоритмы позволяют повысить вычислительную эффективность систем автоматической группировки объектов, а также точность их результатов, оцененную по индексу Рэнда.

– *разработаны* алгоритмы для автоматической группировки повторяющихся фрагментов блоков данных на основе алгоритма k -средних

совместно с локально-чувствительным хэшированием, которые **использованы** в составе системы хранения данных ООО «Центр вычислительных технологий»);

- разработанные алгоритмы кластерного анализа использовались в составе адаптивных алгоритмов машинного обучения при выполнении проекта Минобрнауки РФ FEFE-2023-0004 «Адаптивные методы синтеза и управления процессами проектирования компонентов сложных систем», а также при выполнении договора на выполнение научно-исследовательских работ по совершенствованию векторных СУБД.

Рекомендации об использовании результатов диссертационного исследования:

Результаты диссертационной работы Ахматшина Ф.Г. могут быть использованы в составе систем хранения данных, в частности – в системах управления векторными базами данных и в подсистемах компрессии данных распределенных дисковых хранилищ, а также в системах анализа данных тестовых испытаний промышленной продукции.

Оценка достоверности результатов исследования выявила:

– теоретические выкладки основываются на корректном применении математического аппарата, корректном применении и совершенствовании современных методов оптимизации и машинного обучения;

– **установлено**, что достоверность результатов диссертационного исследования подтверждается результатами проверки работоспособности и сравнительной эффективности разработанных алгоритмов в ходе проведения вычислительных экспериментов в достаточном количестве;

– **показано**, что результаты диссертационного исследования успешно применены в составе системы управления дисковыми хранилищами в ООО «Центр вычислительных технологий», о чем получен соответствующий акт.

Личный вклад соискателя состоит в проведении всех этапов исследования самостоятельно, непосредственном участии в апробации результатов, подготовке публикаций, разработке и верификации алгоритмов

решения поставленных задач. Научные положения, выносимые на защиту, основные выводы, результаты экспериментов принадлежат автору.

Диссертационный совет пришел к выводу о том, что диссертация Ахматшина Фариды Галиулловича «Модели и алгоритмы автоматической группировки объектов для систем анализа и хранения данных на основе методов семейства k -средних» является завершенной научно-квалификационной работой, в которой изложены новые научно обоснованные решения задач кластерного анализа, дополняющие множество эффективных алгоритмов решения задач автоматической группировки, а также алгоритмов предобработки данных для таких задач. Диссертация соответствует критериям п.9, установленным Положением о присуждении ученых степеней, и паспорту специальности «Системный анализ, управление и обработка информации, статистика».

На заседании 28 марта 2025 года диссертационный совет принял решение присудить Ахматшину Ф.Г. ученую степень кандидата технических наук по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика.

При проведении открытого голосования диссертационный совет в количестве 17 человек (7 человека дистанционно), из них 8 докторов наук по специальности 2.3.1, участвовавших в заседании, из 21 человека, входящих в состав совета, проголосовали: за – 16, против – 1, воздержавшихся – нет.

Председатель

Диссертационного совета

Ученый секретарь

Диссертационного совета

28.03.2025



Ковалев

Игорь Владимирович

Панфилов

Илья Александрович