

УТВЕРЖДАЮ
Первый заместитель
генерального директора
АО «НИИАС»,
д.т.н., профессор
Е.Н. Розенберг
2025 г.



ОТЗЫВ ВЕДУЩЕЙ ОРГАНИЗАЦИИ

Акционерного общества «Научно-исследовательский и проектно-конструкторский институт информатизации, автоматизации и связи на железнодорожном транспорте»
на диссертационную работу **Рябцева Антона Борисовича**
«Развитие метода динамического формирования групп объектов по принципу идентичности для больших данных», представленную на соискание учёной степени кандидата технических наук по специальности
2.3.8 – Информатика и информационные процессы

Актуальность темы диссертации

Актуальность разработки методов эффективного формирования групп идентичных объектов в динамичных средах обусловлена возрастающей потребностью в автоматизации процессов анализа и обработки данных в условиях их высокой изменчивости. Рост объёмов информации, увеличение скорости её обновления и необходимость оперативного принятия решений в таких сферах, как транспортная логистика, бизнес-аналитика и управление ресурсами, требуют новых подходов к группировке данных, способных адаптироваться к изменениям в реальном времени. Особую значимость приобретает задача повышения структурной согласованности в условиях динамичных графов, где нарушение старых и появление новых связей требуют оперативного пересчёта кластерной структуры. Существующие решения, ориентированные на слабую динамику или требующие регулярного перестроения кластеров, не всегда обеспечивают необходимую эффективность. Кроме того, важной проблемой остаётся оптимизация расчёта признаков идентичности объектов, особенно при использовании аналитических SQL-запросов, которые зачастую становятся узким местом в системах обработки больших данных.

Работа Рябцева А.Б. посвящена разработке метода динамического формирования групп идентичных объектов с учётом ограничений реального времени, высокой изменчивости системы и необходимости поддержания

согласованности. Предлагаемый подход сочетает методы машинного обучения и динамической кластеризации, что позволяет существенно повысить эффективность обработки данных в прикладных задачах. В качестве результата данной работы предлагается техническое решение, являющееся ключевым компонентом системы формирования групп схожих товарных предложений в транспортной логистике и на товарных площадках.

Структура и содержание диссертации

Диссертация состоит из введения, пяти глав, заключения и списка литературы. **Во введении** обоснованы актуальность темы исследования, научная новизна и практическая значимость результатов, полученных в работе, представлены выносимые на защиту научные положения, приведён обзор актуальных исследований в области. **В первой главе** представлен анализ теоретических основ задачи динамического формирования групп объектов по принципу идентичности, с акцентом на понятиях интероперабельности, структурной согласованности и оценки качества группировки. Также проведена формализация задачи, обозначены основные требования к алгоритмам и обоснован выбор подхода, сочетающего бинарную модель оценки идентичности и графовый алгоритм кластеризации, в частности, алгоритм распространения меток (Label Propagation Algorithm, LPA). **Вторая глава** посвящена задаче оценки схожести объектов как ключевому этапу, предшествующему их группировке, формализована функция схожести и рассмотрены практические аспекты формирования признаков, а также возможные архитектурные ограничения, связанные с необходимостью выполнения аналитических SQL-запросов в некоторых задачах. **В третьей главе** представлен анализ традиционных и современных подходов к оптимизации выполнения аналитических SQL-запросов, включая методы на основе машинного обучения, и сделан вывод о пределах их применимости в условиях динамично изменяющихся данных. **В четвёртой главе** представлены развитие и экспериментальное обоснование выбора метода динамического формирования групп, включая формализацию структуры связей, сравнительный анализ алгоритмов кластеризации и описание оригинальной двухэтапной модификации алгоритма LPA. **В пятой главе** приведены методика и результаты экспериментального подтверждения применимости предложенного метода, включая сравнительный анализ качества кластеризации и примеры практического использования на реальных данных. **В заключении** представлены основные результаты работы.

Основные результаты диссертации и их научная новизна

1) В работе представлено развитие метода динамического формирования групп объектов по принципу идентичности для больших и изменчивых данных, обеспечивающего высокую структурную согласованность групп. В отличие от классических методов кластеризации, которые требуют заранее заданного числа кластеров

либо имеют ограничения по масштабируемости и чувствительности к параметрам, предложенный метод позволяет эффективно работать с данными с заранее неизвестной структурой и числом групп, адаптируясь к их постоянным изменениям в условиях реального времени.

2) В работе получены новые научные результаты, составляющие основу методики группировки объектов. Разработанные подходы позволяют решать ключевые задачи обработки больших данных: эффективное вычисление признаков идентичности, устойчивое графовое объединение семантически схожих объектов и оценку качества группировки в динамичных условиях.

3) В работе обоснованы выбор и модификация комбинированного подхода к группировке. В отличие от известных методов, в работе обоснована комбинация бинарной классификации для оценки парной схожести объектов и алгоритма распространения меток (Label Propagation Algorithm) для формирования групп. Данная комбинация является наиболее эффективной для задач динамического объединения семантически идентичных объектов.

4) В работе предложен модифицированный двухэтапный алгоритм кластеризации на основе LPA. Алгоритм является новым и позволяет повысить структурную согласованность получаемых групп за счёт калибровки пороговых параметров на основе допустимой доли ошибочных связей. В отличие от аналогов, алгоритм устойчив к неполной или зашумлённой информации благодаря отдельной обработке сильных и слабых связей, что обеспечивает высокую однородность после первого этапа и высокую полноту при сохранении однородности после второго этапа.

5) В работе предложена новая методика оценки качества группировки объектов для больших динамичных данных. Методика позволяет осуществлять мониторинг качества кластеризации в условиях отсутствия полной эталонной разметки. В методике определены: подход к приближённой оценке однородности на основе двухэтапного взятия выборки, показатель покрытия как автоматизированная альтернатива полноте, а также порядок экспертной верификации для интерпретации результатов.

Соответствие паспорту научной специальности

Область исследования и содержание диссертации соответствуют паспорту специальности 2.3.8 – Информатика и информационные процессы, в частности по пунктам:

Пункт 1: Разработка компьютерных методов и моделей описания, оценки и оптимизации информационных процессов и ресурсов, а также средств анализа и выявления закономерностей на основе обмена информацией

пользователями и возможностей используемого программно-аппаратного обеспечения.

Пункт 7: Разработка методов обработки, группировки и аннотирования информации, в том числе, извлечённой из сети интернет, для систем поддержки принятия решений, интеллектуального поиска, анализа.

Пункт 12: Разработка технологий извлечения и анализа информации в больших базах данных, в том числе, с использованием концепции многомерного представления (OLAP) и интеллектуального анализа данных (Data Mining) статического и в реальном масштабе времени, реализация моделей баз знаний.

Пункт 13: Разработка и применение методов распознавания образов, кластерного анализа, нейро-сетевых и нечётких технологий, решающих правил, мягких вычислений при анализе разнородной информации в базах данных.

Достоверность полученных результатов

Достоверность результатов работы подтверждена экспериментальной проверкой эффективности предложенных методов и алгоритмов на реальных данных предложений на товарных площадках; корректным сравнением с существующими подходами, продемонстрировавшим превосходство двухэтапной модификации LPA по балансу показателей однородности и полноты; успешным практическим внедрением результатов исследований в проектах отдела «Машинное обучение и матчинг» ООО «Озон Технологии»; публикациями основных положений диссертации в рецензируемых научных изданиях, входящих в список ВАК, и докладами на международных и российских конференциях.

Значимость полученных результатов

Результаты, полученные в диссертации Рябцева А.Б., имеют научную и практическую значимость.

Теоретическая значимость состоит в развитии метода динамического формирования групп объектов по принципу идентичности для больших и изменчивых данных. Разработанная методика количественной оценки качества группировки расширяет инструментарий для анализа качества кластеризации в динамичных системах.

Практическая значимость результатов исследования состоит в преимуществах, которые могут быть получены при их внедрении, а именно: повышение точности и полноты группировки семантически идентичных объектов, снижение уровня шума в данных, возможность адаптивной работы в условиях постоянного изменения информации и высокая масштабируемость для распределённых вычислительных систем.

Результаты данной работы использованы в рамках государственного задания номер 103-00001-25-02 в интересах развития транспорта РФ, а разработанные методы и алгоритмы внедрены в технологические процессы и используются в проектах ООО «Озон Технологии», что подтверждает их успешную апробацию на реальных данных и практическую эффективность.

Рекомендации по использованию результатов

Результаты работы могут быть использованы:

– В системах управления цепями поставок и логистическими платформами для автоматизированного формирования объединённых карточек грузов и транспортных средств, агрегации данных от разных поставщиков, а также для повышения качества данных путём дедубликации информации на цифровых торговых площадках.

– В задачах интеграции и консолидации данных из разнородных источников, в том числе при построении систем управления цифровыми двойниками, для идентификации и связывания записей, описывающих одну и ту же сущность.

– В системах анализа данных и бизнес-аналитики для транспортных компаний, позволяя повысить качество и согласованность операционных данных, что напрямую влияет на достоверность аналитических отчётов и эффективность управления перевозками, складскими запасами и оптимизацией маршрутов.

– В распределённых вычислительных системах для обработки больших объёмов динамически изменяющихся данных, где требуется устойчивая и масштабируемая кластеризация объектов в условиях реального времени.

– В исследовательской деятельности, связанной с моделированием транспортных потоков и логистических сетей, как инструментарий для анализа динамичных кластеризаций, а также как основа для дальнейшего развития методов оптимизации маршрутов и алгоритмов формирования групп перевозок в графовых моделях транспортной инфраструктуры.

Отдельные методы, представленные в работе, в частности, методика оценки качества группировки и модифицированный алгоритм LPA, могут иметь широкое применение в любых прикладных задачах, связанных с поиском семантически идентичных объектов и обеспечением структурной согласованности данных.

Замечания

1. Во введении при анализе проработанности темы раскрыты ограничения классических методов кластеризации, однако, по нашему мнению, необходимо представить более развернутое сопоставление современных графовых алгоритмов (Louvain, Infomap, Walktrap) с точки зрения их

применимости именно к задаче семантической идентичности, а не только к общим задачам поиска сообществ.

2. В главе 3, содержащей анализ методов оптимизации SQL-запросов, в сравнительной таблице результатов представлены данные о времени выполнения, но отсутствует сравнение подходов с точки зрения потребления вычислительных ресурсов.
3. В главе 4, представляющей модифицированный алгоритм LPA, не представлена детализация процедуры калибровки пороговых параметров на размеченных данных, включая критерии остановки для достижения заданных долей ошибок.

Заключительная оценка

Отмеченные недостатки не оказывают существенного влияния на общую положительную оценку диссертационной работы в целом. Основные положения диссертации изложены в 8 печатных работах, из них 3 в журналах из списка ВАК. Опубликованные работы достаточно полно отражают содержание диссертации. Кроме того, результаты данной работы докладывались на трёх конференциях, включая 2 международные. В диссертации приведены все необходимые ссылки на цитируемые источники в соответствии с п. 14 «Положения о присуждении ученых степеней».

Диссертационная работа Рябцева А.Б. «Развитие метода динамического формирования групп объектов по принципу идентичности для больших данных» является научно-квалификационной работой, в которой содержится решение научной задачи, имеющей значение для развития соответствующей области знаний. По научной новизне полученных научных результатов, теоретической и практической значимости, по критериям, установленным пунктом 9 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства Российской Федерации 24 сентября 2013 г. № 842 и приложений 2,3,4 Положения о совете по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, утвержденного приказом Министерства образования и науки РФ от 10.11.2017 г. №1093, предъявляемым к кандидатским диссертациям, ее автор, Рябцев Антон Борисович, заслуживает присуждения ему ученой степени кандидата технических наук по специальности 2.3.8 – Информатика и информационные процессы.

Диссертационная работа, автореферат и отзыв на работу рассмотрены на совместном заседании секций №7 «Управление активами, надежностью и рисками», №8 «Информационная и кибербезопасность» Научно-технического совета и Экспертного совета АО «НИИАС», протокол №
2 от 21.11. 2025 г.

Сведения о ведущей организации: Акционерное общество
«Научно-исследовательский и проектно-конструкторский
институт информатизации, автоматизации и связи на
железнодорожном транспорте»

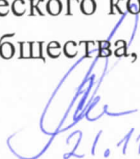
Сайт: www.niias.ru

Почтовый адрес: 109029, Москва, Нижегородская улица, 27, стр. 1

Рецензент:

Руководитель секции №8 НТС АО «НИИАС»,
главный эксперт Научно-технического комплекса
Технологий информационного общества,
д.т.н., профессор МГТУ
Дата:

Сабанов Алексей Геннадьевич



21.11.2025