

ОТЗЫВ

Официального оппонента к.т.н. П.Е. Голосова
на диссертационную работу **Рябцева Антона Борисовича** на тему
**«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 неразрывно связана с проблемами масштабируемости и производительности алгоритмов обработки данных в распределённых средах. Исследование Рябцева А.Б. затрагивает критически важный аспект — построение эффективного алгоритмического аппарата для работы с большими и изменчивыми данными.

Научная новизна работы заключается в обосновании и реализации метода динамической кластеризации на основе алгоритма распространения меток (LPA) для распределённых вычислительных систем, что обеспечивает его масштабируемость при обработке больших объёмов данных. Оригинальной является разработанная двухэтапная модификация алгоритма LPA, обладающая свойством инкрементальности и устойчивостью к неполной информации, что позволяет применять его в условиях постоянного обновления данных без полного пересчёта кластерной структуры. Работа содержит новую методику оценки качества группировки, адаптированную для мониторинга в распределённых системах обработки данных и не требующую трудоёмкой полной верификации результатов.

Достоверность результатов подтверждена не только высокими метриками качества кластеризации, но и успешной интеграцией метода в промышленную инфраструктуру, что служит доказательством его жизнеспособности и производительности на реальных нагрузках. Кроме того, результаты данной работы использованы в рамках государственного задания номер 103-00001-25-02.

Теоретическая значимость работы заключается в развитии теоретических основ построения масштабируемых систем обработки больших данных. Теоретически значимыми результатами являются итоги проведённого исследования возможностей оптимизации планов выполнения аналитических SQL-запросов, которое вносит вклад в теорию оптимизации запросов, выявляя практические ограничения методов машинного обучения в динамичных средах. Разработанная двухэтапная модификация алгоритма LPA вносит вклад в теорию обработки потоковых данных, обеспечивая устойчивость кластеризации

в условиях постоянного обновления информации. Разработанная методика оценки качества группировки объектов вносит вклад в теорию анализа динамических графов и потоковых данных, предоставляя формальный аппарат для контроля качества кластеризации в реальном времени.

Практическая значимость. Внедрение результатов работы в ООО «Озон Технологии» демонстрирует, что предложенное решение работает в высоконагруженной среде и решает конкретные бизнес-задачи. Метод представляет значительный интерес для построения масштабируемых систем обработки данных в различных предметных областях.

Работа состоит из введения, пяти глав, заключения и списка литературы.

Во введении обоснованы актуальность темы исследования, научная новизна и практическая значимость результатов, полученных в работе. Введение включает выносимые на защиту научные положения, обзор актуальных исследований по теме работы.

Первая глава формирует теоретический фундамент исследования, вводя и анализируя ключевые концепции интероперабельности и структурной согласованности, а также проводя обзор метрик для оценки качества группировки, что завершается строгой формальной постановкой задачи.

Вторая глава сосредоточена на проблеме оценки парной схожести, формализуя соответствующую функцию и выявляя основное узкое место системы — низкую производительность аналитических запросов, с помощью которых в некоторых задачах формируются признаки объекта.

Третья глава даёт критическую оценку перспективам применения машинного обучения для оптимизации этих запросов, констатируя, что существующие методы требуют значительных ресурсов для поддержания актуальности моделей и плохо адаптируются к изменениям в данных.

В четвёртой главе представлен оригинальный двухэтапный алгоритм, который последовательно строит кластеры на основе связей высокой и средней надёжности, что позволяет повысить полноту группировки без потери однородности.

Пятая глава содержит экспериментальное обоснование метода, включающее сравнительное тестирование на реальных данных, которое подтверждает его эффективность, а также разработку практических методик для оценки качества группировки в промышленной эксплуатации.

В заключении содержится краткое изложение основных результатов работы.

Замечания:

1. Архитектура предложенного метода в целом описана, но отсутствует детализированная схема рабочего процесса в распределённой среде.

В Главе 4 упомянута реализация на MapReduce, но для полноты картины необходимо описать, как именно организованы этапы: вычисление признаков, построение графа, двухэтапная кластеризация LPA

- и обновление групп при поступлении новых данных.
2. Целесообразно было бы дополнить положения, выносимые на защиту, количественным сравнением методов кластеризации не только по V-мере, но и по времени выполнения и масштабируемости на больших графах, чтобы подчеркнуть комплексные преимущества предложенного подхода.
 3. При описании адаптации метода для распределённых систем следовало кратко обозначить ключевые архитектурные проблемы, которые пришлось решить для обеспечения масштабируемости.

Общее заключение: Диссертационная работа Рябцева А.Б. является законченным научным исследованием, имеет научную ценность, возможность практического применения. Работа является актуальной и содержит новые научные результаты, полученные автором лично, и свидетельствующие о личном вкладе в науку. Работа обладает внутренним единством, содержание исследования соответствует поставленным в работе целям и задачам.

Диссертационная работа Рябцева А.Б. «Развитие метода динамического формирования групп объектов по принципу идентичности для больших данных» соответствует пп. 9-14 «Положения о присуждении учёных степеней», утверждённого постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, предъявляемым к диссертациям на соискание учёной степени кандидата технических наук по специальности 2.3.8 «Информатика и информационные процессы», а её автор, Рябцев А.Б., заслуживает присуждения ему учёной степени кандидата технических наук по данной специальности.

Официальный оппонент

кандидат технических наук,

директор Института общественных наук Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации

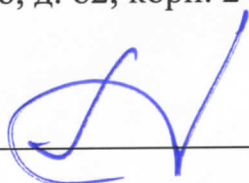
Адрес: 119571, г. Москва, проспект Вернадского, д. 82, корп. 2

05.11.2025 г.

ЗАВЕРЯЮ

УЧЕНЫМ СЕКРЕТАРЬ



 / П.Е. Голосов