

ОТЗЫВ

на автореферат диссертации Ахматшина Фарида Галиулловича
«Модели и алгоритмы автоматической группировки объектов для систем
анализа и хранения данных на основе методов семейства k -средних»,
представленной на соискание ученой степени кандидата технических наук по
специальности 2.3.1 – Системный анализ, управление и обработка
информации, статистика

Работа Ахматшина Фарида Галиулловича посвящена разработке новых алгоритмов автоматической группировки, используемых в системах анализа и хранения данных, повышению эффективности алгоритмов кластерного анализа при обработки больших данных.

Вследствие увеличения объемов обрабатываемых и анализируемых данных усложняется внутренняя структура систем анализа данных, включая подсистемы хранения данных, что требует применения алгоритмов машинного обучения, в том числе алгоритмов кластерного анализа, для повышения эффективности управления в системах хранения данных (дисковые массивы, векторные базы данных).

Для классификации объектов и разделения их на группы разработано множество методов и алгоритмов. Однако, универсальных алгоритмов кластерного анализа, обеспечивающих высокую эффективность разделения объектов при незначительных различиях в данных практически нет. Кроме того, увеличение объема обрабатываемых данных выявляет низкую вычислительную эффективность существующих алгоритмов. Повышению вычислительной производительности, качества результата автоматической группировки объектов для систем анализа и хранения данных посвящена данная работа.

В работе получены важные научные результаты:

1. Новый подход к нормализации данных для предобработки входных данных, используемых в системах анализа данных результатов неразрушающих испытаний образцов промышленной продукции.

2. Новый алгоритм кластеризации с использованием жадной эвристической процедуры выбора радиуса локальных концентраций по размеченным данным.

3. Новый алгоритм кластеризации для построения индекса векторной базы данных для приближенного поиска ближайших соседей, обеспечивающий компромисс между точностью и временем вычислений, существенно улучшает метрику полноты в задачах приближенного поиска ближайших соседей.

4. Новый алгоритм автоматической группировки повторяющихся фрагментов блоков данных на основе алгоритма k -средних совместно с локально-чувствительным хэшированием (LSH), обеспечивающий увеличение эффективности сжатия данных в системах хранения данных.

5. Новая процедура инициализации центров кластеров для алгоритмов автоматической группировки больших объемов данных, использующая вспомогательную структуру данных – массив слагаемых для вычисления суммы квадратов расстояния.

Результаты прикладных и теоретических задач получены с применением методов системного анализа, исследования операций, теории размещения, теории оптимизации, прошли широкую апробацию, опубликованы в ведущих рецензируемых журналах.

По тексту автореферата имеются следующие замечания:

1. Неясно, как осуществлялся выбор количества (30) запусков для каждого алгоритма кластеризации для их сравнительного анализа, и есть ли связь между количеством запусков и достигнутыми результатами.

2. В автореферате на стр. 14 автор утверждает, что имеется корреляция между значениями целевой функции и полнотой, однако на рис. 4 эта связь не очевидна.

Указанные замечания не снижают ценности рассматриваемой работы.

Представляемая на защиту работа соответствует требованиям, предъявляемым к кандидатским диссертациям, а ее автор Ахматшин Фарид Галиуллович заслуживает присуждения ученой степени кандидата технических наук по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика.

Декан факультета управления и бизнеса
ФГБОУ ВО «Ангарский государственный
технический университет»,
Доктор технических наук,
доцент

Истомин Андрей Леонидович

3 марта 2025 г.

Подпись Истомина А.Л. удостоверяю

Проректор по научной работе
ФГБОУ ВО «Ангарский государственный
технический университет»,
Доктор технических наук, доцент



А.В. Бальчугов

Адрес организации: 665830, Ангарск, квартал 85а, дом 5.
Тел. 8(3955) 67-89-15, postmaster@angtu.ru