

ОТЗЫВ

на автореферат диссертации Ахматшина Фариды 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

Диссертационное исследование Ф.Г. Ахматшина посвящено разработке новых алгоритмов автоматической группировки для использования в системах анализа и хранения больших данных с целью увеличения эффективности управления ими. Диссертант провел анализ существующих методов автоматической группировки данных, способов их предобработки, алгоритмов инициализации начальных центров кластеров, выявил их низкую вычислительную эффективность при увеличении объема обрабатываемых данных. Эти аспекты делают тему диссертационного исследования Ф.Г. Ахматшина весьма актуальной.

В диссертационном исследовании автор предложил новый подход к нормализации данных для их предобработки, позволяющий повысить точность кластеризации, что подтверждается ее результатами, приведенными в табличной и графической форме. Кроме того, разработаны новые алгоритмы кластеризации, демонстрирующие большую эффективность по сравнению с классической реализацией алгоритма k -средних. Тем не менее, отмечается, что текущая реализация алгоритма кластеризации для построения индекса для векторной базы данных при увеличении объема хранимых данных требует значительных вычислительных ресурсов по сравнению с алгоритмом Ллойда. Автор планирует дальнейшие исследования в направлении повышения эффективности используемой в алгоритме жадной эвристической процедуры. Также автор предлагает новый способ сжатия данных и подтверждает приведенными показателями времени сжатия и размера архива повышение производительности этого процесса. Необходимо отметить также разработанную автором новую процедуру инициализации центров кластеров в алгоритмах кластеризации, применяемую к большим данным. Ее использование позволит уменьшить вычислительную сложность агломеративного алгоритма и улучшить его производительность, что подтверждается приведенными результатами вычислительных экспериментов.

Между тем, диссертанту можно сделать следующие замечания, касающиеся содержательной стороны его работы.

1. Из текста автореферата не ясно, что понимается под однородными партиями изделий, и какие диапазоны значений характеристик влияют на группировку продуктов.

2. Не совсем ясно, как применение предложенного подхода к нормализации для предобработки входных данных позволяет выявлять наиболее значимые контролируемые параметры продукции.

3. Нет пояснений, почему, несмотря на большее значение целевой функции, в алгоритме LEES приоритет отдан показателю точности.

4. На рисунке 2 нет значений, соответствующих методу MBKM++.

Кроме того, необходимо отметить наличие пунктуационных и синтаксических ошибок, путаницу с обозначениями n , N , k . Номера рисунков идут не в возрастающем, а в убывающем порядке.

Указанные замечания не снижают научно-практическую ценность диссертационного исследования. Судя по автореферату, у данного диссертационного исследования присутствуют все необходимые признаки актуальности, достоверности, научной новизны, теоретической и практической значимости полученных лично соискателем научных результатов. Диссертация представляет собой завершенное научное исследование, соответствует требованиям, установленным Положением о присуждении ученых степеней, 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.

Красиков Виталий Александрович, к.ф.-м.н.,
старший научный сотрудник
НОЦ ФНС России и МГТУ им. Н. Э. Баумана,
105005, Москва, Госпитальный переулок, 4-6 стр. 3, этаж 2//
4 марта 2025г.