

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

Хорошо известно, что способы обработки данных, такие как кластеризация, не всегда могут быть применены напрямую. Для унификации шкал исходные данные стандартизируют, используя затем евклидово расстояние или его квадрат. Обнаружение шума и выбросов в данных является одним из требований к методам нормализации. При этом специфические задачи могут требовать особых методов предобработки. Разнообразные методы кластерного анализа используют различные подходы к предобработке информации. Они комбинируют применение различных метрик, методов вычисления координат и расстояний для поиска оптимальных решений. Для большинства практических задач такой подход является оправданным.

Как следует из автореферата, Ахматшин Ф.Г. обосновывает разработку новых алгоритмов автоматической группировки. Он использует их в системах анализа и хранения данных, для повышения эффективности алгоритмов кластеризации при обработке больших данных в системах автоматической группировки объектов. В частности, он применяет их в составе векторной СУБД и подсистем компрессии данных в составе систем хранения данных.

Помимо теоретической значимости, состоящей в дополнении эффективных алгоритмов решения задач автоматической группировки, а также алгоритмов предобработки данных, для таких задач выполнена практическая реализация результатов. Программная разработка алгоритма автоматической группировки повторяющихся фрагментов блоков данных на основе алгоритма k -means, совместно с локально-чувствительным хэшированием (LSH) предназначена для использования в системах хранения данных. Она реализована в разработанной модели оптимального использования дискового пространства с учетом компрессии данных в рамках работ по договору на выполнение НИР.

К сожалению работа не лишена недостатков. К замечаниям можно отнести следующее:

1. Необходимо пояснение, что означает «уличение параметра тщательности сжатия» (стр. 17).
2. Нет пояснений, что обозначено A_{MB}^z и A_B^z (стр. 19).

3. Не совсем понятно, как определялся коэффициент пропорциональности вероятности выбора вектора вычисленному квадрату расстояния до всех выбранных центров.

4. Нет пояснений, каким образом проверялась статистическая значимость разницы (таблица 3) и между какими показателями она вычисляется.

5.

На основе автореферата можно сделать вывод, что работа соответствует всем требованиям, предъявляемым ВАК РФ к кандидатским диссертациям, а ее автор Ахматшин Фарид Галиуллович заслуживает присуждения ему ученой степени кандидата технических наук по специальности 2.3.1 – Системный анализ, управление и обработка информации, статистика.

Доктор физико-математических наук,
профессор, главный научный сотрудник
лаборатории «Математические модели
принятия решений» ФГБУН Институт
математики им. С.Л. Соболева СО РАН

Кочетов Юрий Андреев

Адрес организации:
просп. Академика Коптюга ,4.
г.Новосибирск, 630090, Россия
e-mail: jkochet@math.nsc.ru
Телефон: +7(383)3297583

