

ОТЗЫВ

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

В условиях экспоненциального роста объёмов данных актуальна разработка новых методов кластеризации данных и поиска ближайших соседей как ценного источника новых знаний о структуре хранимой информации, её сходстве и компактности. Ахмадиным Ф.Г. предложено новые решения задач анализа данных неразрушающих тестовых испытаний промышленной продукции; архивации и хранения данных; организации векторных баз данных. Разработаны новые эффективные алгоритмы автоматической группировки для систем анализа и хранения данных в составе векторной СУБД, а также подсистем компрессии данных.

Исследования, выполненные автором, позволили получить важные результаты, обладающие научной новизной:

1. Подход, комбинирующий нормализацию по допустимым значениям параметров оцениваемых характеристик продукции и оценки Джеймса-Штейна, позволяющий повысить эффективность предобработки входных данных, используемых в процессе анализа результатов неразрушающих испытаний образцов промышленной продукции.

2. Алгоритм кластеризации для системы анализа данных тестовых испытаний радиоизделий, включающий жадную эвристическую процедуру выбора радиуса локальных концентраций по размеченным данным.

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

4. Алгоритм автоматической группировки повторяющихся фрагментов блоков данных на основе алгоритма k -средних совместно с локально-чувствительным хэшированием, увеличивающий эффективность сжатия данных.

5. Процедура инициализации центров кластеров для алгоритмов автоматической группировки больших объёмов данных, способная быстро находить приемлемое начальное решение при большом объёме данных.

Полученные результаты представлены на международных конференциях и семинарах, опубликованы в профильных высокорейтинговых журналах, внедрены в работу ООО «Центр вычислительных технологий».

Замечания:

1. На страницах 11, 14, 20 показано, что для каждого алгоритма проведено 30 экспериментов. Следовало бы обосновать количество экспериментов с учётом различий алгоритмов. Неясно, насколько сильно изменяются значения целевой функции (Таблицы 3 и 4) при изменении числа запусков.

2. Ожидается, что разработанные алгоритмы позволяют улучшить процесс анализа результатов испытаний широкого спектра образцов промышленной продукции. Однако на рисунке 1 и в таблице 2 упомянута интегральная схема 140УД26А. Проводились ли испытания для других радиоизделий? Насколько сильно отличаются наборы данных и условия тестирования алгоритмов?

Замечания не влияют на положительную оценку диссертационного исследования. Работа Ахмадшина Ф.Г. соответствует требованиям пунктов 9-14 Положения о присуждении учёных степеней, утверждённого постановлением Правительства Российской Федерации от 24.09.2013 № 842 (ред. от 16.10.2024) «О порядке присуждения учёных степеней», предъявляемым к диссертациям на соискание учёных степени кандидата технических наук. Автор работы Ахмадшин Фарид Галиуллинович заслуживает присуждения учёной степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Старший научный сотрудник отдела Прикладной информатики ИВМ СО РАН
доктор технических наук по специальности 05.25.05 – Информационные системы и процессы

«__17__» марта 2025 г.

Ничепорчук Валерий Васильевич

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр «Красноярский научный центр
Сибирского отделения Российской академии наук» (ФИЦ КНЦ СО РАН)
обособленное подразделение – Институт вычислительного моделирования
Сибирского отделения Российской академии наук (ИВМ СО РАН)
660036, г. Красноярск, Академгородок, 50, стр. 44,
тел.: 8 (391) 243-27-56, e-mail: sek@icm.krasn.ru

Подпись Ничепорчука В. В. удостоверяю.

Учёный секретарь ИВМ СО РАН
канд. физ.мат. наук
«17» марта 2025 г.

А.В. Вяткин