

ОТЗЫВ

на автореферат диссертации Казаковцева Владимира Львовича
«Алгоритмы ускоренного поиска в векторных базах данных»,
представленной на соискание ученой степени кандидата технических наук по
специальности 2.3.1 – Системный анализ, управление и обработка
информации, статистика

Приближенный поиск ближайших соседей (Approximate Nearest Neighbor Search, ANN) представляет собой эффективный метод, позволяющий находить достаточно близкие к запросу векторы в высокоразмерных пространствах, когда точный поиск становится вычислительно нецелесообразным из-за огромных объемов данных и высокой размерности. В отличие от точных алгоритмов, ANN сознательно отказывается от стопроцентной точности в пользу существенного выигрыша в скорости и масштабируемости. Традиционные структуры данных и методы, использующие геометрические принципы (например, неравенство треугольника), в многомерных пространствах быстро теряют эффективность - это хорошо известное проявление «проклятия размерности», которое делает точный поиск ближайших соседей неприменимым на практике.

Автореферат указывает на то, что в диссертационной работе Казаковцева В. Л. разрабатываются новые алгоритмы приближенного поиска ближайших соседей, ориентированные на векторные базы данных. Актуальность исследования обоснована объективным и непрерывным ростом как объема, так и размерности обрабатываемых данных в современных информационных системах, требующих новых эффективных алгоритмов обработки данных, которые могут быть встроены в уже существующие информационные системы.

Теоретическая и практическая значимость полученных результатов не вызывает сомнений. Предложенные алгоритмы и операторы вносят заметный вклад в развитие методов индексирования и поиска в векторных базах данных, а также в теорию мультимодальной кластеризации и эволюционных алгоритмов. Практическая ценность подтверждается возможностью их внедрения в промышленные системы управления векторными базами данных, которые способны эффективно работать с сотнями миллионов и миллиардами объектов. Дополнительным подтверждением востребованности темы служит поддержка исследований грантами, выполнение госзаданий и наличие хозяйственных договоров с ведущими технологическими компаниями. Опубликованные статьи, участие в конференциях и регистрация программ для ЭВМ также служат подтверждением актуальности настоящей работы.

В работе можно выделить следующие замечания:

1. Возможно, следовало также провести исследование зависимости границ классов сложности в зависимости от объема данных и размерности набора данных эмбедингов.

2. Нет явного указания на границы применимости разработанного классификатора сложности (начиная с каких размерностей для определения сложности запроса достаточно использовать только число результирующих кластеров).

3. Нет явного указания на границы применимости разработанной упрощенной жадной агломеративной процедуры.

Несмотря на перечисленные замечания, работа содержит новые решения, которые могут представлять интерес для практического применения в системах векторного поиска, а также могут использоваться для развития алгоритмов векторного поиска, основанных на принципах, отличных от IVF. Работа соответствует требованиям, предъявляемым к кандидатским диссертациям по специальности 2.3.1.

Доктор технических наук, профессор
кафедры прикладной математики
ФГБОУ ВО «Кемеровский
государственный университет»

Крутиков Владимир Николаевич

Адрес организации:
ул. Красная, д.6,
г. Кемерово, 650000, Россия
e-mail: krutikovvn@gmail.com
Телефон: +7 (3842) 58-12-26

