

УДК 004.422.81

РАЗРАБОТКА ИНСТРУМЕНТА СРАВНЕНИЯ АЛГОРИТМОВ ВЕКТОРНОГО ПОИСКА

Бабаянц А.А. (Университет ИТМО), Томилов Н.А. (Университет ИТМО), Туров В.П. (Университет ИТМО)

Научный руководитель – кандидат технических наук, доцент Платонов А.В. (Университет ИТМО)

Аннотация. В докладе предлагается новый инструмент для сравнения алгоритмов поиска ближайших соседей. В отличие от существующих популярных решений, предложенный инструмент позволяет сравнивать алгоритмы не только с точки зрения качества и скорости поиска, но и с точки зрения используемых ресурсов, что позволит находить оптимальные параметры алгоритмов построения векторного индекса для практических задач.

Введение. Количество данных, генерируемых человечеством, стремительно растет. Генерируемые данные имеют не только текстовый формат, но содержат и другие модальности, такие как изображения и видео. Алгоритмы с применением глубокого обучения способны формировать векторные представления, позволяющие отображать семантическую составляющую мультимодальных документов, представленных в виде плотных векторов (dense embedding). Однако плотные векторные представления не позволяют построить инвертированный индекс, и потому не могут быть применены в классическом информационном поиске. В этом случае допустимо использовать алгоритмы поиска ближайших соседей, однако с увеличением объема информации эффективность поиска снижается, так как алгоритмы k-ближайших соседей имеют линейную сложность. Компромиссом является использование алгоритмов класса приближенного поиска ближайших соседей. Существует множество таких алгоритмов, каждый из которых имеет достоинства и недостатки, в том числе с точки зрения потребления вычислительных ресурсов, поэтому необходимо иметь инструмент, позволяющий выбирать алгоритм под текущую практическую задачу.

Основная часть. В докладе предлагается новый инструмент для проведения автоматизированного сравнения алгоритмов класса приближенного поиска ближайших соседей. Существующие программные решения сфокусированы на результативности алгоритмов и позволяют измерять пропускную способность (throughput) и полноту поиска (recall). Предлагаемый инструмент сравнения расширяет количество собираемых метрик и добавляет метрики использования вычислительных ресурсов системы. Такой подход позволяет оценить эффективность каждого алгоритма с точки зрения требований к вычислительным ресурсам.

Выводы. Реализованное программное обеспечение позволяет сравнивать алгоритмы приближенного поиска по большему количеству критериев, что позволяет точнее выбрать подходящий для каждой задачи алгоритм.

Туров В.П. (автор)

Подпись

Платонов А.В. (научный руководитель)

Подпись