

УДК 004.85, 004.056

МЕТОДЫ ПОВЫШЕНИЯ ПРОИЗВОДИТЕЛЬНОСТИ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ И ИХ ПРИМЕНЕНИЕ К ОБРАБОТКЕ БОЛЬШИХ ДАННЫХ

Роечко Д.В., Зыкин М.М. (Федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет ИТМО")
Научный руководитель – к.т.н., доцент (квалификационная категория "ординарный доцент") Меншиков А.А. (Федеральное государственное автономное образовательное учреждение высшего образования "Национальный исследовательский университет ИТМО")

В работе рассматриваются способы повышения производительности алгоритмов машинного обучения применительно к обработке больших данных. Проведен обзор методов снижения временных затрат на обучение алгоритмов, а также специализированных библиотек, адаптированных под работу с большими массивами информации, выбраны решения, применимые к задаче определения мошеннических транзакций.

Введение. Стремительный рост собираемых человечеством данных и вычислительных мощностей компьютеров дал новые возможности для развития и применения искусственного интеллекта. Однако прогресс в данной области практически напрямую связан с улучшением производительности оборудования, которое с течением времени может стать невозможным ввиду физических ограничений. Этот факт ставит под сомнение возможность создания более глубоких по своей структуре самообучающихся моделей, необходимых для решения усложняющихся задач. Одной из таких является распознавание нелегитимных банковских транзакций. Оно непосредственно связано с обработкой большого массива данных, кроме того, модель машинного обучения должна быть способна обрабатывать транзакции в режиме реального времени, что требует высокой скорости вычислений. Следовательно, использование методов оптимизации производительности является актуальным для данной задачи.

Основная часть. Предлагается применять техники повышения производительности алгоритмов машинного обучения, такие как векторизация вычислений, уменьшение размерности данных, пакетная нормализация и параллельные вычисления с использованием модели MapReduce. Последний из указанных методов положен в основу библиотек Hadoop Mahout и Spark MLlib, предназначенных для обработки больших данных. Следственно, необходимо провести обзор соответствующих методов и библиотек, а также оценить их применимость для реализации системы распознавания мошеннических транзакций в рамках проводимой научно-исследовательской работы.

Вывод. В результате проделанной работы был произведён обзор методов оптимизации алгоритмов машинного обучения, сформированы рекомендации для их дальнейшего применения к решению задачи идентификации мошеннических банковских транзакций.

Роечко Д.В. (автор)

Подпись

Зыкин М.М. (автор)

Подпись

Меншиков А.А. (научный руководитель)

Подпись