

АНАЛИЗ РЕЗУЛЬТАТОВ ПРОХОЖДЕНИЯ ПРАКТИЧЕСКИХ ЗАДАНИЙ В МАССОВЫХ ОТКРЫТЫХ ОНЛАЙН
КУРСАХ (МООК) НА ПРИМЕРЕ LEARNSQL.RU

Артамонова В. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – старший преподаватель Говоров А.И.

(Университет ИТМО, Санкт-Петербург)

Анализ результатов прохождения практических заданий в МООК на примере данных, полученных с прошлогодней сессии LearnSQL.ru, позволяет построить признаковые описания объектов в полученных данных, выявить паттерны поведения пользователей, на основании которых строится алгоритм классификации представленных пользователями решений задач на честные и списанные ответы.

Введение. Благодаря бурному развитию информационно-коммуникационных технологий массовые открытые онлайн курсы получили широкое распространение как новый формат обучения, по сравнению с традиционными способами как средне-специальное и высшее образование. Данные, извлеченные из сервисов, осуществляющих подобный формат дистанционного получения знаний, представляют большой интерес и их анализ является актуальной проблемой, решение которой предлагают как отечественные, так и зарубежные авторы. Частой темой исследований представляется вопрос о честности пользователей во время прохождения курса, вероятность того, что отправленное решение к заданию может быть списано.

Основная часть. В качестве решения предлагается провести первичный анализ данных с целью выявления объектов, которые могут оказать влияние на выполнение практического задания: сложность самого задания, количество попыток и время выполнения задания, типичность поведения пользователя. Выбранным объектам необходимо дать признаковое описание и выделить целевую переменную.

В качестве целевой переменной выступает признак “вероятность списывания”, которой может принимать значения “есть/нет”, что позволяет классифицировать тип задачи как бинарную классификацию. Так как представленные данные изначально не имеют такого признака, то необходимо провести их разметку. Решению присваивается присутствие вероятности списывания, если время, потраченное на выполнение задания на определенный процент меньше, чем рассчитанное среднее время выполнения данного задания. Признак “среднее время выполнения задания” считается с помощью выведенных весов сложности каждого задания, выявленных на основании поведения пользователей, среднего количества попыток, которое пользователь использует для дачи верного решения, и времени выполнения заданий пользователями.

После того, как произведена разметка данных, предлагается использовать три алгоритма машинного обучения для решения задачи бинарной классификации: логистическая регрессия, деревья решений, случайный лес. Алгоритм, показавший наивысшую точность, может быть далее использован для анализа новых данных.

Выводы. Разработанный алгоритм следует применять на новых данных, извлекаемых из сервиса LearnSQL.ru, с целью классификации ответов пользователей на честные и нечестные. Испытание данной разработки необходимо провести на результатах прохождения практических заданий с новой сессии онлайн-курса. Также данный алгоритм можно применять к анализу платформ МООК, при условии совпадения архитектуры предоставляемых данных.

Артамонова В. (автор)

Говоров А.И. (научный руководитель)

