

УДК 004.021

РАЗРАБОТКА АЛГОРИТМА АВТОМАТИЗИРОВАННОГО ИЗВЛЕЧЕНИЯ ШАБЛОНОВ ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ ИЗ НАБОРОВ БАНКОВСКИХ ТРАНЗАКЦИЙ

Татаров Д.А., Менщиков А.А. (Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – к.т.н., доцент Менщиков А.А.

(Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Данная работа посвящена разработке алгоритма автоматизированного извлечения шаблонов поведения пользователей из банковских транзакций, которые могут использоваться в системах обнаружения мошенничества в наборах банковских транзакций, а также в системах генерации данных.

Введение. В настоящее время при разработке методов, алгоритмов и систем детектирования аномалий в банковских транзакциях исследователям довольно сложно, а порой невозможно найти релевантные данные для обучения и тестирования разработанных методов. Например, самая популярная платформа для соревнований и исследований в области машинного обучения Kaggle, по запросу «транзакции по банковским картам», выдает всего 5 наборов данных, из которых только три релевантны запросу. Для решения проблемы малого количества релевантных наборов данных был предложен алгоритм генератора банковских транзакций на основании марковских моделей и сценариев поведения пользователей, однако для его корректной работы необходимо наличие большого числа шаблонов поведения пользователей.

Основная часть. Одним из возможных методов решения задачи детектирования мошеннических транзакций в банковских системах является поведенческий анализ. Для корректной работы данного метода необходимо наличие набора данных для тренировки алгоритма машинного обучения. Таким набором данных обычно выступает эталонное поведение пользователя или шаблон (сценарий) этого поведения. Также на основании сценариев поведения возможно проводить генерацию данных для различных целей, в том числе для обучения других моделей машинного обучения и систем детектирования аномалий.

Для выделения шаблонов из наборов данных, находящихся в открытом доступе, был предложен алгоритм, основанный на двухэтапном процессе: определение значимых признаков и последующее выделение признаков шаблона поведения, основанного на этих признаках. Также полученные шаблоны могут быть объединены в более крупные с использованием методов кластеризации.

Выводы. В результате работы был разработан алгоритм для извлечения шаблонов поведения пользователей из наборов банковских транзакций. Данный алгоритм будет применен для генерации шаблонов поведения пользователей в системах анализа аномалий банковских транзакций, а также в системах генерации данных, в частности, в системе генерации банковских транзакций.

Татаров Д.А. (автор)

Подпись

Менщиков А.А. (автор)

Подпись

Менщиков А.А. (научный руководитель)

Подпись