

УДК 004.942

**ИССЛЕДОВАНИЕ ИНДИВИДУАЛЬНОЙ ПРЕДСКАЗУЕМОСТИ ФАКТА  
ЗАРУБЕЖНОЙ ПОЕЗДКИ В СЛЕДУЮЩЕМ МЕСЯЦЕ ДЛЯ КЛИЕНТОВ БАНКА**

**Ставинова Е.А.** (Университет ИТМО), **Боченина К.О.** (Университет ИТМО)

**Научный руководитель – к.т.н., доцент Боченина К.О.**

(Университет ИТМО)

В данном докладе будет рассмотрена задача предсказания факта совершения зарубежной поездки клиентом банка на основании его транзакций. Кроме того, с помощью существующих метрик предсказуемости будет оценена индивидуальная предсказуемость данного факта для каждого клиента из имеющегося набора данных.

Данные о транзакциях клиентов предприятий финансовой сферы, фиксирующие время, сумму и место совершения трат, позволяют исследовать различные аспекты платежного поведения клиентов, включая геопространственное поведение. Так, широкое распространение получили алгоритмы выявления подозрительных транзакций и случаев мошенничества с банковскими картами по геолокациям трат. Также наличие данных о пространственно-временных паттернах трат клиента позволяет разрабатывать широкий спектр прикладных рекомендательных сервисов.

В данном исследовании была поставлена следующая цель: разработать способ оценки предсказуемости факта совершения зарубежных поездок на основании транзакционных данных. Для реализации поставленной цели использовался открытый набор данных о транзакциях по дебетовым картам одного из крупных российских банков (2756 клиентов, 699523 транзакции за период с января 2016 года по февраль 2018 года).

Все транзакции были преобразованы в цепочки меток локаций для каждого пользователя, таким образом, каждый пользователь получил представление в виде дискретного временного ряда. На основании анализа литературных источников были выявлены наиболее подходящие для решения поставленной задачи метрики предсказуемости для временных рядов. Первая метрика оценивает предсказуемость как меру того, насколько хорошо текущее значение временного ряда объясняется его предыдущими значениями.

В зависимости от значений первой метрики предсказуемости были выделены пять групп клиентов, для которых были исследованы метрики качества предсказания (Precision и Recall). Далее были разработаны алгоритмы, которые относят клиента к одной из выделенных групп предсказуемости на основании истории его транзакций. Кроме того, было проведено изучение и сравнение точности данных алгоритмов.

Что касается второй метрики, то она сосредоточена на анализе зависимости распределения меток (классов) от признаков. С помощью данной метрики предсказуемости было исследовано качество описания метки сгенерированными из транзакционной истории признаками.

Оцененная индивидуальная предсказуемость клиента может внести существенный вклад в улучшение точности прогнозирования зарубежной поездки в следующем месяце. По имеющимся транзакционным данным клиента возможно определить класс предсказуемости, к которому он относится, и сделать прогноз на следующий месяц, выбрав классификатор, наиболее подходящий для того или иного класса предсказуемости. Процесс выбора оптимального классификатора для каждого из классов предсказуемости еще нуждается в дальнейшем анализе. Кроме того, с помощью анализа качества описания меток имеющимися признаками можно подобрать оптимальный набор признаков, который приведет к улучшению точности работы модели в дальнейшем.

Ставинова Е.А. (автор)

Подпись

Боченина К.О. (научный руководитель)

Подпись