

УДК: 004.942

Название: Разработка алгоритмов прогнозирования зарубежных поездок по геопространственным финансовым данным.

Авторы:

Ставинова Е.А., Университет ИТМО, г. Санкт-Петербург;

Боченина К.О., Университет ИТМО, г. Санкт-Петербург

Научный руководитель: Боченина К.О., Университет ИТМО, г. Санкт-Петербург

Тезис доклада:

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 модели, позволяющие прогнозировать локацию пользователя в следующем месяце на основе истории его транзакций. Для реализации поставленной цели использовался открытый набор данных о транзакциях по дебетовым картам одного из крупных российских банков (2794 клиентов, 718544 транзакции за период с января 2016 года по февраль 2018 года).

Все транзакции были преобразованы в цепочки меток локаций для каждого пользователя. Цепочки были разделены на тренировочную (15 месяцев) и тестовую (11 месяцев) выборку. Перед предсказанием было решено выявить различные типы путешественников. Для этого с помощью метода кластеризации k-means было выделено 3 кластера пользователей. Затем были оценены матрицы переходных вероятностей для каждого кластера. Для предсказания метки пользователя в следующем месяце использовались цепи Маркова на единственном признаке — локации.

Были получены следующие результаты процедуры кластеризации. Первая группа пользователей — это люди, которые проводят много времени не дома, в другом регионе России. Путешествуют за рубеж мало, поездки у представителей первой группы почти всегда длительные. Вторая группа представляет собой людей, которые редко отправляются в поездки, ездят за рубеж чаще, чем по стране. Путешествие представителей этой группы вероятнее всего занимает не более месяца. Третий кластер характеризуется длительными

поездками, преимущественно за рубеж. Также характерной чертой представителей третьего кластера является длительность поездки — с вероятностью более чем 0.5 она не ограничивается одним месяцем.

В работе была исследована точность предсказания при вариации доступной длины цепочки из тренировочного набора, по которой определялась принадлежность клиента к кластеру. Предсказание велось на один месяц вперед с использованием переходных вероятностей кластеров. Также было исследовано качество предсказаний на разные месяцы, таким образом, была изучена предсказательная способность по популяции в среднем. Для оценки качества предсказаний использовались метрики Precision и Recall для всех клиентов и по кластерам отдельно. Кроме того, были построены таблицы сопряженности для визуализации эффективности алгоритма в общем и отдельно по кластерам.

Автор _____ / Ставинова Е.А. /

Научный руководитель _____ / Боченина К.О. /