

УДК: 004.942

Название: Прогнозирование цепочек трат с использованием точечных случайных процессов.

Авторы:

Лысенко А.В., Университет ИТМО, г. Санкт-Петербург;

Шиков Е.Н., 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:

Наличие технических возможностей по агрегации больших массивов данных в корпоративных информационных системах приводит к возможности построения предсказательных моделей поведения клиентов по истории их активностей. Например, в банковской сфере можно выполнять персонализацию предложений на основе прогнозирования покупательских потребностей клиента (ожидаемого времени и категории покупки следующей транзакции клиента). Такие знания могут позволить банку предвосхитить желания клиента — предоставить информацию об актуальных скидках на определенную категорию товаров, предсказать денежный поток в той или иной торговой-сервисной точке и др. Также на основе прогнозов платежной активности производится разработка бонусных программ совместно с предприятиями сферы торговли и услуг. Поэтому разработка и исследование эффективности методов прогнозирования цепочек трат клиентов является актуальной темой для исследований.

Целью данного исследования является разработка генеративной модели трат клиента, позволяющей предсказывать факт наступления и время трат в различных категориях на основании транзакционной истории клиента.

Данные для исследования (транзакционная история 28 тысяч клиентов за период 2016-2018 года) были предоставлены банком-партнером. Данные были разделены на обучающую и тестовую выборку (15 и 2 месяцев, соответственно). Были отобраны клиенты по нижнему порогу частоты транзакций, а также выбраны 10 наиболее активно используемых категорий транзакций из доступных 87.

Для построения модели предсказания цепочек транзакций был использован аппарат точечных случайных процессов, а именно, неоднородный пуассоновский процесс и процесс Хоукса. В качестве исходных признаков на вход модели подавались временные отметки и категории транзакций клиентов.

Для оценки качества построенных моделей отдельно оценивались точности предсказания времени и категории генерируемых событий. В качестве метрик качества использовалось усредненное по всем клиентам значение абсолютной ошибки времени и доля верно предсказанных категорий событий от общего числа предсказанных событий. Данные метрики были рассчитаны для пуассоновского процесса и процесса Хоукса при вариации размерности данных и гиперпараметров моделей.

Результаты моделирования показывают, что процесс Хоукса больше подходит для предсказания трат, что объясняется наличием зависимости между покупками в разных категориях для данной модели. Метрики качества для модели на основе процесса Хоукса составили: для средней ошибки – 16 часов, для точности по 10 категориям – 26%.

Автор _____ / Лысенко А.В. /

Научный руководитель _____ / Боченина К.О. /

Директор мегафакультета ТИНТ _____ / Бухановский А.В. /