

Добавление внешней информации для улучшения векторных представлений в моделях последовательностей событий

М. Д. Ковалева, научный руководитель: А.А.Зайцев

Московский физико-технический институт

16 декабря 2023 г.

Постановка задачи

Проблема

Рассматривается задача создания векторных представлений для последовательностей событий банковских транзакций. В основном, для создания векторных представлений используются модели, обучаемые контрастивными методами. Но такие модели не учитывают внешнюю информацию и «зацикливаются» на конкретных пользователях.

Цель работы

Улучшение векторных представлений посредством добавления внешней информации об окружающих пользователях.

Задачи работы

- 1) изучение существующих моделей для создания векторных представлений для последовательностей транзакций
- 2) разработка методов добавления внешней информации
- 3) валидация разработанных методов

Особенности данных

- ▶ Данные банковских транзакций представляют из себя набор нерегулярных временных рядов
- ▶ Особенности по сравнению с обычными временными рядами:
 - ▶ **Нерегулярность:** временные промежутки между данными непостоянны
 - ▶ **Распределение по пользователям:** имеется не один временной ряд каких-то событий, а множество, по одному временному ряду для каждого пользователя
 - ▶ **Многомерность:** описание одной транзакции обычно состоит из нескольких признаков, которые могут включать как численные, так и категориальные переменные. В нашем случае: тип транзакции, объем транзакции
- ▶ Выборка из n последовательностей событий $D = \{\mathbf{S}^i\}_{i=1}^n$. Каждая последовательность представлена в виде набора пар $\mathbf{S}^i = \{(t_j^i, \mathbf{Z}_j^i)\}_{j=1}^{T_i}$, $t_j^i \in [0, T_i]$ — упорядоченные времена событий, а вектор $\mathbf{Z}_j^i \in \mathbb{R}^d$ — описание j -ого события i -ой последовательности.

Особенности задачи

- ▶ **Получение представлений:** Для того, чтобы получить представление события, мы используем отдельную нейронную сеть, кодировщик $e(\cdot)$.

$$e(\mathbf{S}) = \mathbf{H}, \text{ где } \mathbf{H} = \{(t_j, \mathbf{h}_j)\}_{j=1}^T, \mathbf{h}_j \in \mathbb{R}^d.$$

- ▶ **Применение векторных представлений:** Представления содержат в себе информацию о пользователе и его состоянии в определенный момент. Применяются для решения прикладных задач. Мы рассматриваем:
 - ▶ Глобальная валидация: Классификация пользователей (по признаку уйдет ли клиент из банка или отдаст ли клиент кредит)
 - ▶ Локальная валидация: уйдет ли пользователь в ближайший месяц
 - ▶ Предсказание типа следующей транзакции пользователя

Baseline решение

Базовое решение CoLES: контрастивная модель.

Состоит из

- ▶ TrxEncoder: энкодер транзакций, состоит из нескольких линейных слоев
- ▶ SeqEncoder: энкодер последовательности состоит из рекуррентной сети LSTM или GRU

Обучается self-supervised контрастивным методом:

- ▶ Использует классическую контрастивную функцию потерь
$$\mathcal{L}_{uv} = Y_{uv} \frac{1}{2} d(c_u, c_v)^2 + (1 - Y_{uv}) \frac{1}{2} \max\{0, \rho - d(c_u, c_v)\}^2$$
, где (u, v) - рассматриваемая пара объектов (в нашем случае это последовательности транзакций), $d(c_u, c_v)$ - расстояние между векторными представлениями рассматриваемых объектов, ρ - гиперпараметр, Y_{uv} равен 1, если пара положительная и 0 если пара отрицательная.
- ▶ Положительные пары: подпоследовательности одного пользователя
- ▶ Отрицательные пары: подпоследовательности разных пользователей

Предложенное решение

Одним из путей улучшения представлений CoLES может быть добавление информации о ситуации в мире. Получать ее мы будем агрегируя представления всех остальных пользователей в данный момент.

Используемые методы агрегирования:

- ▶ Mean
- ▶ Max
- ▶ Attention: $\mathbf{B}_t = X \text{softmax}(X^T \mathbf{h}_t)$
- ▶ Learnable attention:
 $\mathbf{B}_t = X \text{softmax}(X^T A \mathbf{h}_t)$

где $\mathbf{h}_t \in R^m$ - вектор представлений для рассматриваемого пользователя,
 $X \in R^{m \times n}$ - матрица, составленная из представлений всех n пользователей на данный момент времени,
 $A \in R^{m \times m}$ - обучаемая матрица

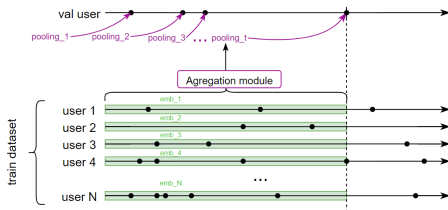


Рис.: Общий пайплайн получения глобальных представлений

Результаты

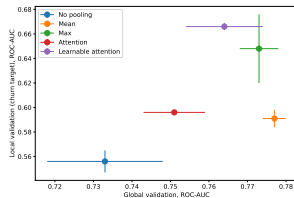
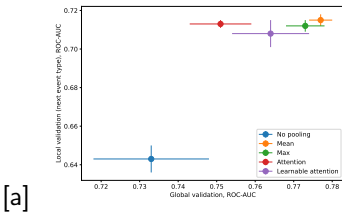


Рис.: Результаты для выборки Churn

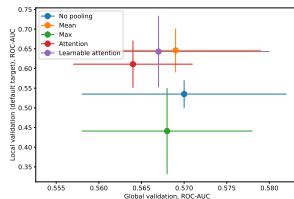
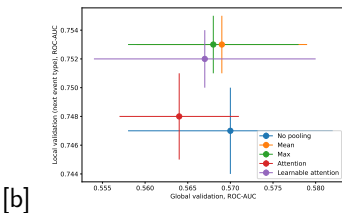


Рис.: Результаты для выборки Default

Рис.: Сравнение качества различных типов агрегации в модели глобальных представлений на локальной и глобальной валидации. В качестве локальной задачи рассматривается прогнозирование следующего типа события (слева) и прикладная задача классификации транзакций (справа)

Выводы

- ▶ Использование глобальных представлений улучшает метрики в большинстве случаев
- ▶ **Learnable attention** чаще всего оказывается в лидерах или наравне с ними, особенно в задачах локальной валидации. Подходы на основе механизма внимания помогают хорошо выделять локальные паттерны в последовательностях
- ▶ **Mean** и **Max** хорошо показывают себя в глобальных задачах. Они дают некоторое среднее представление обо всех пользователях, что может помогать в глобальных задачах классификации, но также они могут сглаживать локальные представления, что негативно сказывается на качестве локальных задач.

Список литературы

1. *Babaev D. et al. CoLES: Contrastive Learning for Event Sequences with Self-Supervision // Proceedings of the 2022 International Conference on Management of Data. — 2022.*
2. *Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality Reduction by Learning an Invariant Mapping. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06) 2 (2006), 1735–1742.*