

# Добавление внешней информации для улучшения векторных представлений в моделях последовательностей событий

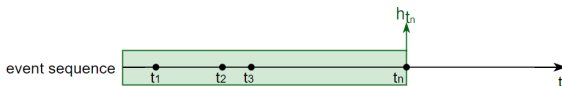
Мария Дмитриевна Ковалева  
Научный руководитель: к.ф.-м.н. А. А. Зайцев

Кафедра интеллектуальных систем ФПМИ МФТИ  
Специализация: Интеллектуальный анализ данных  
Направление: 03.04.01 Прикладные математика и физика  
и  
Data Science, Skoltech

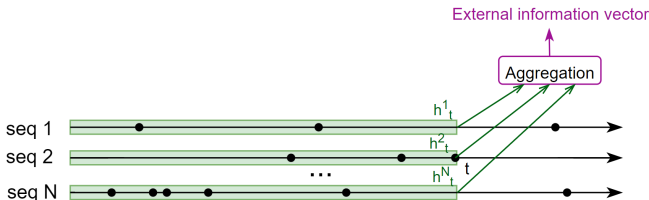
2024

# Введение

**Основная задача:** постоение векторных представлений  $\mathbf{h}_t$  для последовательностей событий



**Недостатки:** существующие решения не учитывают внешнюю информацию



**Идея:** внешняя информация содержится в самих последовательностях и может быть представлена в виде их агрегации

## Цель работы:

Улучшить векторные представления для моделей последовательностей событий, используя агрегации для учета внешней информации

## Задачи:

1. Разработка методов агрегации для учета внешней информации
2. Валидация разработанных методов на данных банковских транзакций

# Постановка задачи и базовые методы

$D = \{S^i\}_{i=1}^n$  — выборка последовательностей событий

$S^i = \left\{ \left( t_j^i, \mathbf{z}_j^i \right) \right\}_{j=0}^{T^i}$  — последовательность событий

$t_j^i \in [0, T^i]$  — время события;  $\mathbf{z}_j^i \in \mathbb{R}^d$  — описание события

**Построение векторных представлений:**  $e(S^i) = H^i$

- ▶  $e$  - энкодер: обычно полносвязная сети для кодирования описания событий + рекуррентная сеть
- ▶  $H^i = \left\{ \left( t_j^i, \mathbf{h}_j^i \right) \right\}$  - векторные представления

**Контрастивное обучение для энкодера:**

$$L_{km} = I_{k=m} d(\mathbf{h}^k, \mathbf{h}^l)^2 + \frac{1}{2} (1 - I_{k=m}) \max \{0, \rho - d(\mathbf{h}^k, \mathbf{h}^l)\}^2,$$

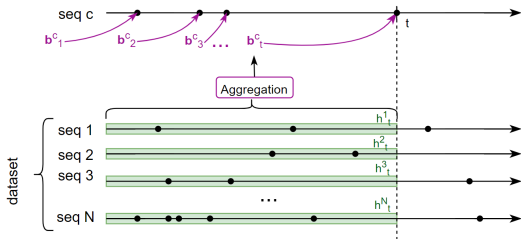
где  $d$  - расстояние между векторными представлениями,

$\rho$  - гиперпараметр

**Авторегрессионное обучение для энкодера:** Функция потерь состоит из кроссэнтропии для категориальных признаков и MSE для непрерывных признаков.

# Получение представлений внешней информации и таксономия методов агрегации

## Построение вектора внешней информации



## Таксономия методов агрегации

При построении методов агрегации можно использовать близость текущей последовательности и последовательностей из обучающего набора.

### Близость последовательностей:

- ▶ по векторным представлениям
- ▶ по времени

Общая формула построения агрегации:  $\mathbf{b}^c_\tau = f(H, f_e(H, \mathbf{h}^c_{t < \tau}), f_t(\tau, T))$

Здесь  $H = [\mathbf{h}^1_{t < \tau}, \dots, \mathbf{h}^n_{t < \tau}]$  — матрица с векторными представлениями всех последовательностей в данный момент времени в столбцах,

$T = [t^1_{t < \tau}, \dots, t^n_{t < \tau}]$  — вектор составленный из времен последних на данный момент событий для всех последовательностей;

$f_e$  и  $f_t$  — функции учитывающие близость по векторам и по времени,  $f$  — функция агрегации, обычно взвешенная сумма векторов из  $H$ .

# Предложенные методы агрегации

## Классические:

- ▶ **Mean:**  $\mathbf{b}_\tau^c = \frac{1}{n} \sum_{i=1}^n \mathbf{h}_{t < \tau}^i$
- ▶ **Max:**  $\mathbf{b}_\tau^c = \max(H)$

## Вдохновленные процессом Хоукса\*:

- ▶ **Exp Hawkes:**  
 $\mathbf{b}_\tau^c = H \exp(-(\tau \mathbf{1} - T))$
- ▶ **Exp learnable Hawkes:**  $\mathbf{b}_\tau^c = \phi_{NN}(\text{concat}(H, \mathbf{h}_\tau^c)) \exp(-(\tau \mathbf{1} - T))$
- ▶ **Attention Hawkes:**  $\mathbf{b}_\tau^c = H(\text{softmax}(H^T \mathbf{h}_\tau^c) \odot \exp(-(\tau \mathbf{1} - T)))$

## Основанные на внимании:

- ▶ **Attention:**  
 $\mathbf{b}_\tau^c = H \text{softmax}(H^T \mathbf{h}_\tau^c)$
- ▶ **Learnable attention:**  
 $\mathbf{b}_\tau^c = H \text{softmax}(H^T A \mathbf{h}_\tau^c)$
- ▶ **Symmetrical attention:**  
 $\mathbf{b}_\tau^c = H \text{softmax}(H^T S^T S \mathbf{h}_\tau^c)$
- ▶ **Kernel attention:**  
 $\mathbf{b}_\tau^c = H \text{softmax}(\phi(H^T) \phi(\mathbf{h}_\tau^c))$

Здесь  $A$  и  $S$  — матрицы с обучаемыми параметрами,  $\phi$  и  $\phi_{NN}$  — обучаемые преобразования (двухслойный перцептрон).

---

\* Laub P. J., Taimre T., Pollett P. K. Hawkes processes. arXiv preprint arXiv:1507.02822. — 2015.

# Валидация предложенных методов

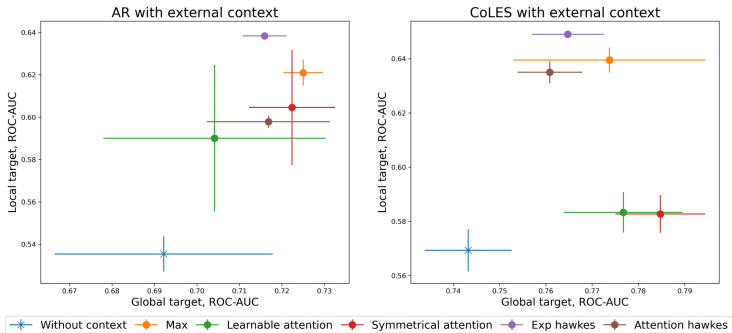
## Данные банковских транзакций:

- ▶ Последовательность  $S^i$ : последовательность транзакций одного пользователя банка
- ▶ Событие  $Z_j^i$ : транзакция (merchant category code + amount)
- ▶ Таргет: ушел ли клиент из банка

## Валидация:

1. **Глобальная:** исследует последовательность как единое целое
  - ▶ Задача бинарной классификации
  - ▶ Для предсказания используется модель бустинга
2. **Локальная:** исследует локальные свойства последовательностей используя плавающее окно
  - a Предсказание типа следующей транзакции
  - b Локальная бинарная классификация
  - ▶ Для предсказания используется MLP голова

# Результаты



Лучшие модели находятся правее и выше.

- ▶ Добавление **внешней информации** улучшает метрики
- ▶ Метод **Exp Hawkes** лучший по локальным свойствам
- ▶ **Классические** и **основанные на внимании** методы лучшие по глобальным свойствам



# Выносятся на защиту

1. Предложено использование агрегации векторных представлений для учета внешней информации в последовательностях событий
2. Предложена таксономия методов агрегации для учета внешней информации и реализованы конкретные методы
3. Показано что добавление внешней информации повышает качество векторных представлений при использовании их в разлных прикладных задачах с реальными данными.

## Публикации

1. Bazarova, A.\*, Kovaleva, M.\*, Kuleshov, I.\*, Romanenkova, E.\*, Stepikin, A.\*, Yugay, A.\*, Mollaev, D., Kireev, I., Savchenko, A., and Zaytsev, A. Universal representations for financial transactional data: embracing local, global, and external contexts. arXiv preprint arXiv:2404.02047 (2024)

---

\*Равный вклад.

## Дополнительные слайды. Результаты.

	Global target			
	Contrastive learning		Autoregressive learning	
	ROC-AUC	PR-AUC	ROC-AUC	PR-AUC
Without contex	0.743 $\pm$ 0.009	0.792 $\pm$ 0.014	0.692 $\pm$ 0.025	0.734 $\pm$ 0.032
Mean	0.773 $\pm$ 0.004	0.828 $\pm$ 0.003	0.722 $\pm$ 0.007	0.776 $\pm$ 0.005
Max	0.774 $\pm$ 0.021	0.818 $\pm$ 0.032	<b>0.725 <math>\pm</math> 0.005</b>	<b>0.777 <math>\pm</math> 0.002</b>
Attention	0.760 $\pm$ 0.014	0.808 $\pm$ 0.017	0.696 $\pm$ 0.014	0.744 $\pm$ 0.017
Learn. attention	0.777 $\pm$ 0.013	0.830 $\pm$ 0.013	0.704 $\pm$ 0.026	0.751 $\pm$ 0.020
Sym. attention	<u>0.785 <math>\pm</math> 0.010</u>	<u>0.835 <math>\pm</math> 0.005</u>	0.722 $\pm$ 0.010	0.769 $\pm$ 0.004
Kernel attention	0.775 $\pm$ 0.003	0.824 $\pm$ 0.002	<u>0.709 <math>\pm</math> 0.019</u>	<u>0.760 <math>\pm</math> 0.003</u>
Exp Hawkes	<u>0.765 <math>\pm</math> 0.008</u>	0.814 $\pm$ 0.009	0.716 $\pm$ 0.005	<u>0.767 <math>\pm</math> 0.013</u>
Exp learn. Hawkes	0.764 $\pm$ 0.008	0.812 $\pm$ 0.008	0.714 $\pm$ 0.025	0.758 $\pm$ 0.020
Attention Hawkes	0.761 $\pm$ 0.007	0.796 $\pm$ 0.009	<u>0.717 <math>\pm</math> 0.014</u>	0.751 $\pm$ 0.023
	Local target			
Without contex	0.569 $\pm$ 0.008	0.321 $\pm$ 0.003	0.535 $\pm$ 0.008	0.299 $\pm$ 0.011
Mean	0.592 $\pm$ 0.005	0.342 $\pm$ 0.005	0.543 $\pm$ 0.006	0.312 $\pm$ 0.006
Max	0.640 $\pm$ 0.005	<b>0.400 <math>\pm</math> 0.006</b>	0.621 $\pm$ 0.006	0.256 $\pm$ 0.008
Attention	0.600 $\pm$ 0.009	0.348 $\pm$ 0.010	0.534 $\pm$ 0.016	0.301 $\pm$ 0.007
Learn. attention	0.583 $\pm$ 0.007	0.330 $\pm$ 0.008	0.590 $\pm$ 0.035	<u>0.338 <math>\pm</math> 0.025</u>
Sym. attention	0.583 $\pm$ 0.007	0.329 $\pm$ 0.007	<u>0.605 <math>\pm</math> 0.027</u>	<u>0.350 <math>\pm</math> 0.020</u>
Kernel attention	0.582 $\pm$ 0.007	0.329 $\pm$ 0.007	<u>0.572 <math>\pm</math> 0.021</u>	0.330 $\pm$ 0.023
Exp Hawkes	<b>0.649 <math>\pm</math> 0.000</b>	0.366 $\pm$ 0.003	<b>0.638 <math>\pm</math> 0.001</b>	<b>0.351 <math>\pm</math> 0.001</b>
Exp learn. Hawkes	0.581 $\pm$ 0.012	0.322 $\pm$ 0.013	0.539 $\pm$ 0.034	0.293 $\pm$ 0.025
Attention Hawkes	<u>0.635 <math>\pm</math> 0.004</u>	<u>0.359 <math>\pm</math> 0.005</u>	0.598 $\pm$ 0.003	0.331 $\pm$ 0.001

Результаты валидации для различных методов. Лучшие значения выделены **жирным**, вторые значения подчеркнуты, третьи подчеркнуты дважды.