

Marked Temporal Point Process with Neural ODE.

Насыров Р.Р., Зайцев А.А.

МФТИ

17 мая 2025 г.

Содержание

1 Введение

- Актуальность
- Новизна

2 Постановка задачи

- Neural ODE
- Temporal Point Process
- DM TPP

3 Результаты

Введение

Актуальность

- Обработка последовательностей событий (TPP) важно в медицине, финансах
- События характеризуются доп. данными (тип, координаты), которые нужно учитывать
- События могут возникать в произвольный момент времени

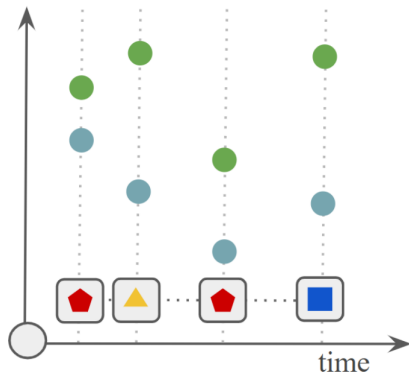


Рис.: Пример TPP с 1 категориальным и 2 непрерывными признаками.

Цель исследования: предложить метод эмбединга нерегулярных маркированных TPP (irregularly-sampled MTPP), пригодный для решения сопутствующих задач

Задача: бинарная классификация MTPP клиентов

Постановка задачи

Алгоритм Neural ODE

- Эволюция $h(t)$ описывается ODE, в котором векторное поле f задается нейросетью:

$$\frac{dh(t)}{dt} = f_{\theta}(h(t), t)$$

- Генеративная модель:

$$\begin{cases} h_{t_0} \sim p(h_{t_0}) \\ h_{t_i} = \text{ODESolve}(h_{t_0}, f_{\theta}, t_i) \\ x_{t_i} \sim p_{\phi}(x|h_{t_i}) \end{cases}$$

Где t_i — времена, в которых нужно вычислить временной ряд, x — значение, соответствующее h , p_{ϕ} — параметр декодера.

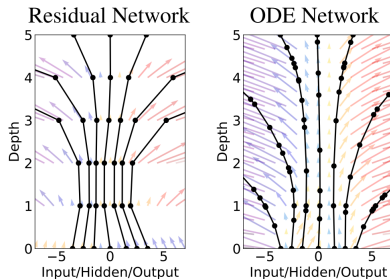


Рис.: Отличие residual- (слева) и ODE- (справа) нейросетей

Постановка задачи

Temporal Point Process

- 1 TRP — стох. процесс, реализация которого — набор $[t_0, \dots, t_N]$
- 2 Обозначим набор событий до t как $\mathcal{H}_t = [t_0, \dots, t_i]$, а pdf следующего события как $f(t|\mathcal{H}_{t_i}) = f^*(t)$, cdf как $F^*(t) = F(t|\mathcal{H}_{t_i})$
- 3 Вместо pdf f^* используется функция *интенсивности*
$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}$$
- 4 в Marked TRP у каждого события появляется дискретная метка $k \in [1, \dots, K]$

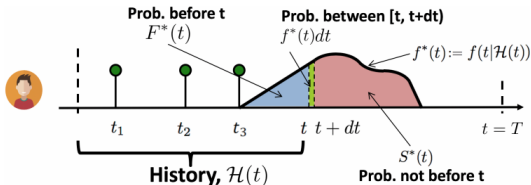


Рис.: Temporal Point Process

Постановка задачи

Marked Temporal Point Process

- 1 МТПР — стох. процесс, реализация которого — набор $\mathcal{S} = [(t_0, k_0), \dots, (t_N, k_N)]$ времен и типов событий.
- 2 Функция интенсивности: $\lambda^*(t, k) = \lambda(t, k | \mathcal{H}_t)$ факторизуется на вероятность события произойти $\lambda_g^*(t)$ и вероятность типа произошедшего события $f^*(k|t)$

$$\lambda^*(t, k) = \lambda_g^*(t) \cdot f^*(k|t)$$

- 3 Правдоподобие:

$$L(\mathcal{S}) = \left[\prod_{i=1}^N \lambda_g^*(t_i) \right] \cdot \left[\prod_{i=1}^N f^*(k_i|t_i) \right] \cdot \exp\left(-\int_0^{t_N} \lambda_g^*(u) du\right)$$

Постановка задачи

Decoupled MTPP with Neural ODE

- 1 Событие (e_i, k_i) , $k_i \in [1, \dots, K]$ имеет эмбединг $W_e(k_i)$
- 2 Эмбединг каждого события: $h(t_i, e_i) = W_e(k_i)$ изменяется под действием Neural ODE независимо от остальных:

$$dh(t; e_i) = \gamma_\theta(h(t; e_i), t, k_i)dt$$

- 3 Затем эмбединги в финальный момент времени t используются для вычисления $\lambda_g^*(t)$ и $f^*(k|t)$ с помощью агрегирующих функций:

$$\lambda_g^*(t) = A(g_\mu(h(t; e_n)), g_\mu(h(t; e_n)), \dots, g_\mu(h(t; e_n)))$$

$$f^*(k|t) = A_k(g_f(h(t; e_1)), g_f(h(t; e_2)), \dots, g_f(h(t; e_n)))$$

Где функции $g_\mu : \mathbb{R}^d \rightarrow \mathbb{R}$ и $g_f : \mathbb{R}^d \rightarrow \mathbb{R}$ моделируют индивидуальный вклад каждого события в интенсивность и вероятность.

Постановка задачи

Схема Decoupled MTPP with Neural ODE

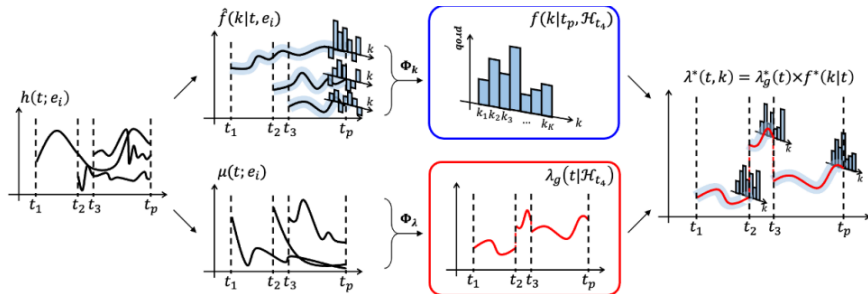


Рис.: DMTPP-ODE

Постановка задачи

- Решается задача классификации.
- Обозначим $\mathcal{S} \in \mathbb{S}$ — МТПР:

$$\mathcal{S} = \{e_i\}_{i=0}^N = \{(t_i, k_i)\}_{i=0}^N; \quad 0 \leq t_i < t_{i+1} \leq 1; \quad k_i \in [1, \dots, K]$$

Где K — количество типов событий, t_i время события. \mathbb{S} множество всех МТПР.

- Модель классификации параметризуется весами нейросети, $h_\theta : \mathbb{S} \rightarrow [0, 1]$ и минимизирует комбинацию supervised Cross-Entropy loss L_{CE} и unsupervised лосса предсказания МТПР L_{MTPP}

$$L_{CE} = \sum_i BCE(h_\theta(\mathcal{S}_i), y_i)$$

$$\ln(L) = \sum_{i=1}^N \ln(\lambda_g^*(t_i)) - \int_0^{t_N} \lambda_g^*(u) du + \sum_{i=1}^N \ln(f^*(k_i | t_i))$$

$$\theta^* = \arg \min_{\theta} \sum_{L(\mathcal{S})} L_{CE}(L(\mathcal{S})) + \ln(L(\mathcal{S}))$$

Результаты и выводы

Агрегация	Accuracy	ROC AUC
Mean	69.2% \pm 0.9%	71.4% \pm 0.8%
LSTM	74.0% \pm 2.0%	77.5% \pm 2.6%
Transformer	70.3% \pm 3.0%	71.6% \pm 5.0%

Таблица: Результаты классификации на датасете оттока клиентов Churn. Отображены mean и 3 std для трех типов агрегаций с усреднением по 5 запускам.

Выводы

- 1 Агрегация по умолчанию (среднее значение) показала худшие результаты.
- 2 LSTM показала лучшие результаты как по точности, так и по ROC AUC.
- 3 Модель трансформатора, похоже, была недостаточно обучена, так как у нее самое большое расхождение.

Следующие шаги

- Использовать Neural CDE для аккумулялирования эффекта от событий одного типа
- Оценивать неопределенность типа события

Литература

- 1 Ricky TQ Chen et al. “Neural ordinary differential equations”. In: Advances in neural information processing systems 31 (2018).
- 2 Nan Du et al. “Recurrent marked temporal point processes: Embedding event history to vector”. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016, pp. 1555–1564.
- 3 Patrick Kidger et al. “Neural controlled differential equations for irregular time series”. In: Advances in neural information processing systems 33 (2020), pp. 6696–6707.
- 4 Hongyuan Mei and Jason M Eisner. “The neural hawkes process: A neurally self-modulating multivariate point process”. In: Advances in neural information processing systems 30 (2017).
- 5 Jakob Gulddahl Rasmussen. “Lecture notes: Temporal point processes and the conditional intensity function”. In: arXiv preprint arXiv:1806.00221 (2018).