

Выявление взаимосвязи между метками с помощью алгоритма, основанного на собственном внимании для задачи классификации с несколькими метками

Г. Л. Боева, научный руководитель: А.А.Зайцев

Московский физико-технический институт

16 декабря 2023 г.

Постановка задачи

Проблема

Большая часть доступной пользовательской информации может быть представлена в виде последовательности событий с отметками времени. Современные подходы фокусируются на архитектуре преобразования последовательных данных, вводящей self-attention к элементам в последовательности. В этом случае мы учитываем временные взаимодействия событий, но теряем информацию о взаимозависимостях меток.

Цель работы

Создание подхода, основанного на механизме собственного внимания над метками, предшествующими прогнозируемому шагу.

Задачи работы

- 1) изучение существующих моделей, работающих в области предсказаний множества меток
- 2) разработка метода на основе внимания для предсказания множества меток
- 3) валидация разработанных методов

Постановка задачи

Рассмотрим классификацию с несколькими метками для последовательности $S = (X_i, Y_i)_{i=1}^{t-1}$. Y_i - набор меток, X_i - набор признаков, специфичных для каждой временной метки от 1 до $t - 1$.

Множество $Y_i \subseteq Y$, где $Y = 1, 2, \dots, K$ - множество всех возможных меток. Размер набора X_i равен размеру набора Y_i . Также введем дополнительный вектор признаков z , описывающий рассматриваемую последовательность S в целом, например, идентификатор пользователя.

Цель: спрогнозировать Y_t для следующей временной метки.

Функция $f(X_{t-\tau}, \dots, X_{t-1}, Y_{t-\tau}, \dots, Y_{t-1}, z) \in [0, 1]^K$, которая принимает историческую информацию о событиях в качестве входных данных и выводит вектор вероятностей присутствия для каждой из K меток.

$S^t = \{(X_j, Y_j)\}_{j=t-\tau}^{t-1}$, где τ означает количество событий, предшествующих временной метки t .

$g(f(\cdot))$ - модель принятия решений по меткам, которая преобразует доверительные баллы в метки, сравнивая оценку для k -й метки с выбранным пороговым значением β_k :

$$g(f_k(\cdot)) = \begin{cases} 1, & f_k(\cdot) > \beta_k \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Данные

ID	Date	Labels	Features
1	12-08-2016	[1, 18, 89]	[2.1, 0.4, 0.7]
	21-08-2016	[3, 8]	[0.3, 1.5]
	23-08-2016	[1, 18]	[1.7, 0.5]
0.	28-08-2016	[?, ?, ...]	

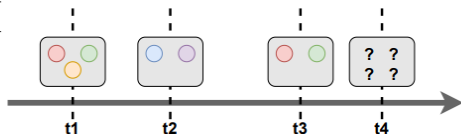


Рис.: Образец набора данных событий с метками времени, который можно использовать в задаче классификации по нескольким меткам и ее визуализация.

Dataset	# events	Median set size	Max set size	# unique labels	Diff
Sales	47 217	16	48	84	0.0632
Demand	5 912	13	24	33	0.0957
Liquor	291 029	14	66	107	0.0413
Transactions	784 520	3	23	77	0.1079
Orders	226 522	2	13	61	0.0518

Рис.: Характеристики наборов данных, используемых в задачах последовательной классификации по нескольким меткам.

Предложенный метод

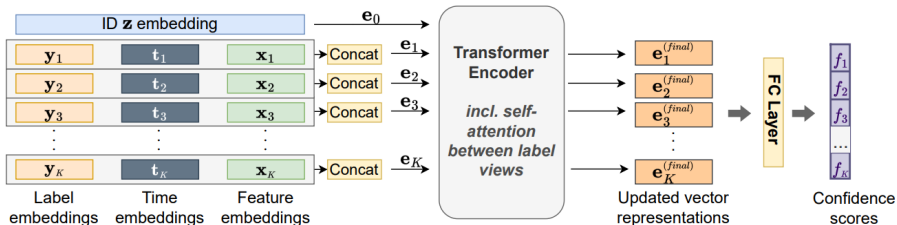


Рис.: Общий пайплайн получения глобальных представлений

Основные результаты

Model	Macro-AUC	Macro-F1
LSTM	0.752	0.222
Transformer	0.750	0.228
LANET	0.954	0.451

Таблица: Сравнение показателей из базовых показателей и LANET в наборе данных электронной коммерции.

	LSTM	TransformerBase	CLASS2C2AE	LANET (ours)
Micro-AUC	<u>2</u>	3.2	3.6	1.2
Macro-AUC	<u>2</u>	3.2	3.6	1.2
Micro-F1	<u>1.8</u>	3.4	3.4	1.6
Macro-F1	1.8	3.2	2.8	<u>2.2</u>

Рис.: Средний ранг для различных показателей, усредненный по 5 рассмотренным наборам данных. Лучшие значения выделены жирным шрифтом, а вторые лучшие значения подчеркнуты.

Дополнительные исследования

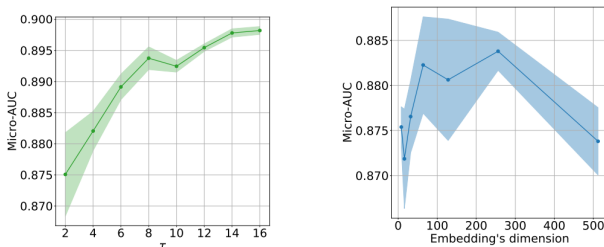


Рис.: а. Зависимость micro-AUC от параметра τ . б. Зависимость micro-AUC от размера векторных представлений.

τ - означает количество событий, предшествующих рассматриваемому событию с временной меткой t , которой присваивается целевая метка Y_t .

Inputs	Micro-AUC \uparrow	Macro-AUC \uparrow	Micro-F1 \uparrow	Macro-F1 \uparrow
All	0.881 \pm 0.007	0.737 \pm 0.017	0.704 \pm 0.018	0.591 \pm 0.003
No amount	0.825 \pm 0.051	0.705 \pm 0.039	0.698 \pm 0.010	0.574 \pm 0.014
No time	0.869 \pm 0.007	0.721 \pm 0.020	0.698 \pm 0.004	<u>0.590</u> \pm 0.003
No ID	<u>0.880</u> \pm 0.003	<u>0.732</u> \pm 0.006	<u>0.703</u> \pm 0.022	0.588 \pm 0.004

Рис.: Как различные типы представлений в LANET влияют на качество.

Выводы

- ▶ Проведены исследования по анализу различных наборов данных, используемых при сравнении реализованной модели LANET.
- ▶ Проведены ряд экспериментов для задачи классификации с несколькими метками на пяти различных выборках и сравнение с базовыми подходами в данной области.
- ▶ Проведена оценка метрики в зависимости от гиперпараметра, отвечающего за информацию о предыдущих временных метках, размера входных представлений или влияния признаков на конечный результат.

Список литературы

1. Kovtun E., Boeva G. *Label Attention Network for sequential multi-label classification: you were looking at a wrong self-attention* // arXiv — 2023.
2. Zhuzhel, V., Grabar, V., Boeva, G., Zabolotnyi, A., Stepikin, A., Zholobov, V., Ivanova, M., Orlov, M., Kireev, I., Burnaev, E., Rivera-Castro, R., Zaytsev, A.: *Continuous-time convolutions model of event sequences* (2023)

Следующий этап работы

Проблема

В рассмотренной литературе не рассматриваются подходы, которые улавливают связь меток в задаче Next Basket Recommendation.

Цель работы

Создание подхода, основанного на механизме собственного внимания над метками, который был бы применен в задачах рекомендаций набора меток для пользователя.

Задачи работы

- 1) изучение существующих моделей, работающих в области предсказаний множества меток
- 2) разработка метода на основе внимания для предсказания множества меток
- 3) валидация разработанных методов

Следующий этап работы

Постановка задачи

Корзина \mathbf{b} — набор элементов, т. е. $\mathbf{b} = \{i_1, i_2, \dots, i_j, \dots, i_{|\mathbf{b}|}\}$, где $i_j \in \mathcal{I}$, и где \mathcal{I} обозначает вселенную всех предметов. Для данного пользователя у нас есть доступ к последовательности n исторических корзин (по возрастанию в хронологическом порядке, так что более свежие элементы находятся в конце) обозначается как $\mathcal{H} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_i, \dots, \mathbf{b}_n]$, где $b_i \subset \mathcal{I}$. Цель исследования создать модель, которая учитывает исторические корзины \mathcal{H} в качестве входных данных и прогнозирует следующую корзину \mathbf{b}_{n+1} в качестве рекомендации.

Список литературы

1. Hui Fang, Guibing Guo, Danning Zhang, and Yiheng Shu. *Deep learningbased sequential recommender systems: Concepts, algorithms, and evaluations*. In *Web Engineering: 19th International Conference, ICWE 2019, Daejeon, South Korea, June 11–14, 2019, Proceedings 19*, pages 574–577. Springer, 2019.
2. Ming Li, Sami Jullien, Mozhdeh Ariannezhad, and Maarten de Rijke. *A next basket recommendation reality check*. *ACM Transactions on Information Systems*, 41(4):1–29, 2023. .