

Label Attention Network для последовательной классификации по нескольким меткам

Галина Боева

Московский физико-технический институт

Курс: Моя первая научная статья /M05-304

Эксперт: к.ф-м.н А. Зайцев

2024

Проблема

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
- 4) обоснование причинно-следственных связей с помощью построения графа на основе внимания

- ▶ Классификация с несколькими метками.
Thomas Hartvigsen, Cansu Sen, Xiangnan Kong, and Elke Rundensteiner. Recurrent halting chain for early multi-label classification. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1382–1392, 2020.
Wenyu Zhang, Devesh K Jha, Emil Laftchiev, and Daniel Nikovski. Multi-label prediction in time series data using deep neural networks. arXiv preprint arXiv:2001.10098, 2020.
- ▶ Основные подходы для задачи классификации с несколькими метками.
Xiao Shou, Tian Gao, Shankar Subramaniam, Debarun Bhattacharjya, and Kristin Bennett. Concurrent multi-label prediction in event streams. In AAAI Conference on Artificial Intelligence, 2023.
Fan Zhang, Shuai Wang, Yongjie Qin, and Hong Qu. Conv-based temporal sets prediction for next-basket recommendation. In 2023 International Conference on Frontiers of Robotics and Software Engineering (FRSE), pages 419–425. IEEE, 2023.

Постановка задачи

Набор меток - $\mathcal{L} = \{1, 2, \dots, L\}$. Рассмотрим набор из N последовательностей $D = (\mathbf{X}_n, \mathbf{Y}_n)_{n=1}^N$. Предполагается, что пары $(\mathbf{X}_n, \mathbf{Y}_n)$ являются случайными величинами, которые подчиняются неизвестному распределению $P(\mathbf{X}, \mathbf{Y})$. Пусть $T_n = |\mathbf{Y}_n|$ обозначает размер набора меток, связанного с \mathbf{X}_n .

В нашей постановке есть ограничение на размер прошлого, доступного модели. Обозначим τ означает количество событий, предшествующих рассматриваемому событию, тогда $D_t^u = (\mathbf{X}_j^u, \mathbf{Y}_j^u)_{j=t-\tau}^{t-1}$, следующих одному и тому же неизвестному распределению $P(\mathbf{X}, \mathbf{Y})$.

Формально $f(\cdot)$ имеет вид:

$$f : D_t^u \rightarrow (p_1, p_2, \dots, p_L),$$

где p_i — вероятность присутствия данной метки в наборе, для предсказания Y_t .

Цель изучения f - минимизировать ожидаемую потерю на подвыборке в соответствии с базовым распределением P :

$$\min_f \mathbb{E}_{(\mathbf{X}, \mathbf{Y}) \sim P} [\text{Loss}(\mathbf{Y}, f(\mathbf{X}))],$$

где Loss — это логарифмическая функция потерь.

Предложенный метод

Пусть $E = \{\mathbf{e}_0, \dots, \mathbf{e}_K\}$, $\mathbf{e}_i \in \mathbb{R}^d$ - последовательность входных представлений, где \mathbf{e}_0 соответствует эмбедингу \mathbf{z} , а все остальные \mathbf{e}_i соответствуют представлениям, фиксирующим историческую информацию с точки зрения метки. Вес внимания α_{ij} и обновленное представление \mathbf{e}'_i рассчитываются как:

$$\mathbf{e}'_i = \sum_{j=0}^K \alpha_{ij} (W^v \mathbf{e}_j); \alpha_{ij} = \text{softmax} \left(\frac{(W^q \mathbf{e}_i)^T (W^k \mathbf{e}_j)}{\sqrt{d}} \right)$$

где W^k - матрица ключевых весов, W^q - является весовой матрицей запроса, а W^v - весовой матрицей значения.

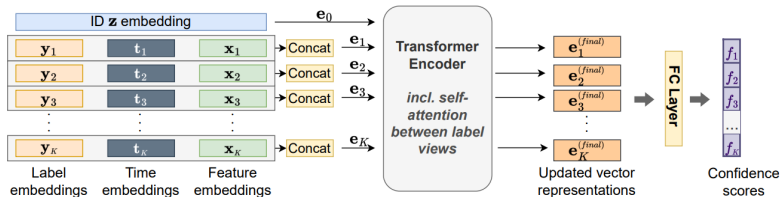


Рис.: Общий пайплайн получения глобальных представлений

Вычислительный эксперимент: Данные

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

Таблица: Характеристики наборов данных, используемых в задачах последовательной классификации с несколькими метками.

- ▶ Demand описывает исторический спрос на продукцию нескольких складов. Функция метки означает категорию продукта, а функция количества относится к соответствующему спросу.
- ▶ Sales — это исторические данные о продажах в разных магазинах. Метки относятся к категориям товаров, а сумма — это количество проданных товаров для определенной категории.

Вычислительный эксперимент: Основные результаты

Dataset	Model	Micro-AUC↑	Macro-AUC↑	Micro-F1↑	Macro-F1↑
Sales	LSTM	<u>0.8670</u>	<u>0.7346</u>	<u>0.5600</u>	<u>0.4389</u>
	TransformerBase	0.8604	0.7206	0.5348	0.3751
	CLASS2C2AE	0.8528	0.6881	0.5116	0.4217
	LANET (ours)	0.9069	0.7627	0.6235	0.4901
Demand	LSTM	0.8829	0.7633	<u>0.6746</u>	0.5929
	TransformerBase	0.8624	0.7240	0.6491	0.5678
	CLASS2C2AE	0.8342	0.7079	0.6738	0.5581
	LANET (ours)	<u>0.8806</u>	<u>0.7373</u>	0.7038	<u>0.5908</u>

Таблица: Сравнение нашего метода LANET с базовыми данными по двум различным наборам данных. Выделены лучшие значения, а значения вторые по рангу подчеркнуты.

Вычислительный эксперимент: Дополнительные исследования

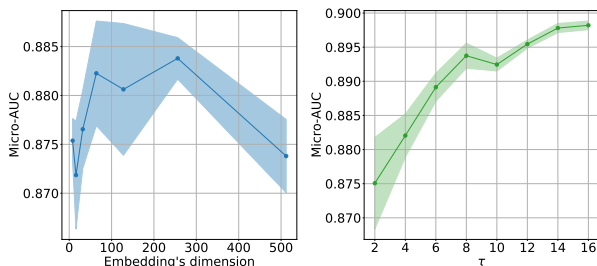


Рис.: а. Зависимость micro-AUC от параметра τ . б. Зависимость micro-AUC от размера векторных представлений.

	Micro-AUC \uparrow	Macro-AUC \uparrow	Micro-F1 \uparrow	Macro-F1 \uparrow
Label-attention	<u>0.881</u> \pm 0.007	<u>0.737</u> \pm 0.017	<u>0.704</u> \pm 0.018	0.591 \pm 0.003
Time-attention	0.837 \pm 0.004	0.682 \pm 0.007	0.674 \pm 0.009	0.588 \pm 0.002
Concat-attention	0.835 \pm 0.001	0.681 \pm 0.010	0.666 \pm 0.000	0.587 \pm 0.001
Gated-attention	0.829 \pm 0.030	0.668 \pm 0.048	0.672 \pm 0.010	0.578 \pm 0.011
Absence-indication	0.882 \pm 0.004	0.742 \pm 0.003	0.704 \pm 0.010	<u>0.589</u> \pm 0.004

Таблица: Сравнение показателей при расчете посещаемости между различными видами в LANET.

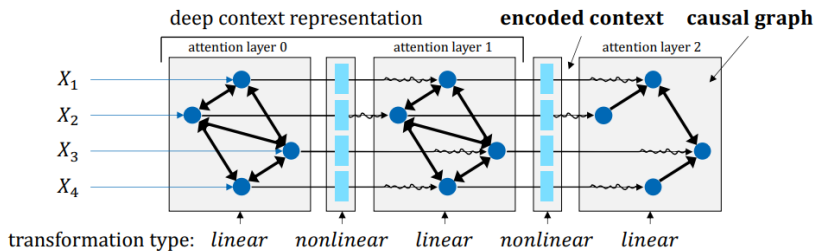


Рис.: CLEANN

- ▶ Проведены исследования по анализу различных наборов данных, используемых при сравнении реализованной модели LANET.
- ▶ Проведены ряд экспериментов для задачи классификации с несколькими метками на двух 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.

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)