
GraphSASRec: модель последовательных рекомендаций на основе самовнивания, дополненная графовыми представлениями.

A Preprint

Матвеев Артем
Московский государственный университет
имени М. В. Ломоносова
matfu21@ya.ru

Майсурадзе Арчил Ивериевич
Московский государственный университет
имени М. В. Ломоносова
artchil@mail.ru

2023

Abstract

Последовательные модели, решающие задачу предсказания следующего взаимодействия пользователя на основе кодирования его исторических событий, являются популярным решением для построения персонализированных рекомендательных систем как в индустрии, так и в академии. Преимуществами таких моделей являются: учет порядка, в котором следуют исторические события и оценка долгосрочных интересов пользователя. Однако подобные подходы недостаточно эксплуатируют полезный коллаборативный сигнал и плохо представляют объекты из длинного хвоста. Популярным решением этих проблем являются методы, основанные на применении графовых нейронных сетей к двудольному графу взаимодействий пользователей. В работе предлагается с другой стороны взглянуть на модель последовательных рекомендаций как на графовую сеть со стороны пользователя. Предлагается метод увеличения глубины этой сети, сохраняющий длину истории пользователя при минимальных накладных расходах. Предлагается способ корректировки смещения, вызванного семплированием из показательного распределения для задачи link-prediction. Наблюдаемые результаты демонстрируют улучшение с точки зрения метрик ранжирования и разнообразия выдачи.

Keywords Информационный поиск · Последовательные рекомендации · Графовые нейронные сети

1 Введение

Последовательные рекомендательные системы - это класс рекомендательных систем, которые принимают во внимание порядок взаимодействий пользователя и пытаются предсказать следующее его взаимодействие. Учет порядка является важной составляющей во многих рекомендательных сценариях: если пользователь только что купил мобильный телефон, то следующая покупка с большой долей вероятности будет аксессуаром к нему. Ранее такая задача решалась с помощью подходов, основанных на Марковских цепях ? или рекуррентных нейронных сетях ????. Но после появления архитектуры трансформер ? и его успеха в задачах распознавания естественного языка, наиболее успешными стали модели, основанные на обработке последовательностей действий пользователя с помощью трансформера ???.

Другим популярным подходом в рекомендательных системах являются графовые нейронные сети ??????. В этом подходе информация, которая есть в рекомендательной системе, рассматривается в виде графов. Большинство данных в любой рекомендательной системе по существу имеют графовую структуру. Например, данные взаимодействий в рекомендательном сервисе могут быть представлены в виде двудольного графа, вершины одной доли которого - пользователи, другой объекты (например,

товары), а наблюдаемый взаимодействия - ребра. Пусть нам дан такой граф. Ключевая идея графовых нейронных сетей заключается в итеративной агрегации признаков представлений соседей в графе и объединение этой агрегированной информации с представлением вершины, для которой рассматривается соседство ?. Такая операция называется распространением сообщений ?. Формально ее можно записать в следующем виде:

$$\text{Aggregation} : n_v^{(l)} = \text{Aggregator}_l(\{h_u^l, \forall u \in \mathcal{N}_v\}),$$

$$\text{Update} : h_v^{(l+1)} = \text{Updater}_l(h_v^{(l)}, n_v^{(l)}),$$

где $h_u^{(l)}$ определяется как представление вершины u после l -ого слоя графовой сети. Aggregator_l и Updater_l представляют собой обучаемые функции агрегации соседей и обновления представления вершины на l -ом слое. В качестве функции агрегации могут выступать как простые варианты по типу max-pooling, mean-pooling ?, так и более сложные, например: importance-pooling ?, агрегация на основе контекста ?, механизм внимания ?, агрегация на базе трансформера ?. В качестве функции обновления могут выступать как простые архитектуры на базе нескольких полносвязных слоев ??, так и более сложные на базе трансформера ?.

Успех графовых подходов в рекомендательных системах можно объяснить тремя причинами ?. Во-первых, выражая все данные в виде вершин и ребер графа, графовые нейронные сети предоставляют общий способ использовать все имеющиеся данные ?, тогда как традиционные рекомендательные системы чаще всего фокусируются на одном или небольшом количестве источников данных. Во-вторых, подобные модели могут явно утилизировать связи высокого порядка (товар X купил пользователь U , которому понравился товар Y , который в свою очередь похож на товар Z). В классических моделях этот учет происходит только неявно. Причем многие работы показывают, что от увеличения глубины графовой сети (то есть явного учета взаимодействий более высокого порядка) наблюдается рост целевых метрик ????. В-третьих, целевой сигнал в рекомендательных системах очень разреженный (например, покупка). Графовые подходы позволяют использовать методы, основанные на обучении с частичным привлечением учителя ?, что приводит к улучшению качества моделей.

При внимательном взгляде на модели последовательных рекомендаций, можно увидеть в них графовую нейронную сеть со стороны пользователя. Как было упомянуто выше, графовые нейронные сети выучивают лучшее семантическое пространство, если начинать использовать в модели связи все большего порядка. Однако, на практике, длина последовательности пользователя варьируется от сотен ??, до тысяч ?? исторических событий. В этом случае, при построении соседств следующих уровней, возникает проблема экспоненциального роста их размера ?, что приводит к невозможности применения графовых методов в классическом виде.

Основной вклад заключается в следующем:

- Предлагается с другой стороны взглянуть на задачу последовательных рекомендаций. Найти в ней сходства с подходами, связанными с графовыми нейронными сетями, и использовать методы из этой области для улучшения моделей последовательных рекомендаций.
- Предлагается метод увеличения глубины связей, утилизируемых в моделях последовательных рекомендаций, на основе предобученных графовых представлений. При этом сохраняющей длину истории пользователя при минимальных накладных расходах.
- Представляется способ корректировки, получающий асимптотически несмещенную оценку градиента для функции потерь в задаче link-prediction. Приводятся результаты, демонстрирующие улучшения с точки зрения метрик полноты по сравнению с решением, не использующим эту корректировку.
- Представляется модифицированная архитектура модели последовательных рекомендаций, учитывающая связи более высоких порядков. Демонстрируются результаты, показывающие, что такой подход приводит к росту метрик ранжирования.

2 Постановка задачи

2.1 Последовательные рекомендации

В последовательных рекомендациях рассматривается задача предсказания следующего положительного взаимодействия пользователя по последовательности его исторических действий $S^u = (S_1^u, S_2^u, \dots, S_{|S^u|}^u)$.

Во время обучения, на момент времени t , модель предсказывает следующий объект интереса пользователя на основе его взаимодействий, произошедших раньше момента t . На вход модели поступает последовательность $(S_1^u, S_2^u, \dots, S_{|S^u|-1}^u)$. Ожидаемый выход модели - следующее положительное взаимодействие $S_{|S^u|}^u$. В данной работе рассматривается модель SASRec (?), в основе которой лежит декодер блок трансформера ?, на выходе которого тоже последовательность. Поэтому задачу можно переформулировать в эквивалентном виде, как задачу предсказания сдвинутой версии последовательности $(S_2^u, S_3^u, \dots, S_{|S^u|}^u)$.

2.2 Предсказание ребра

Дан граф $G = (V, E, X)$, где V - множество вершин, E - множество ребер, $X \in \mathbb{R}^{|V| \times d}$ - d -размерные векторные представления входных вершин. Задача предсказания ребра формулируется как задача определения существования (или появления в будущем, если граф рассматривается как динамический ?) ребра e_{ij} между вершинами i и j , где $i, j \in V$, и $e_{ij} \notin E$.

3 Сопутствующие работы

3.1 Последовательные рекомендации в индустрии

В реальных рекомендательных системах процесс получения кандидатов для пользователя разбивается на две части: матчинг и ранжирование. На стадии матчинга применяются относительно легкие модели ??, которые из миллионов-миллиардов кандидатов отбирают сотни-тысячи. На стадии ранжирования применяются тяжелые модели ???, уточняющие прогнозы моделей с предыдущей стадии.

3.2 Трансдуктивные графовые модели

3.3 Индуктивные графовые модели

4 Корректировка смещения, вызванного семплированием

5 Метод

6 Эксперименты

7 Заключение

Список литературы

- T. M. Ahmed El-Kishky. Twihin: Embedding the twitter heterogeneous information network for personalized recommendation. 2022.
- C. M. Aleksandr Petrov. gsasrec: Reducing overconfidence in sequential recommendation trained with negative sampling. 2023.
- L. B. B. Hidasi, A. Karatzoglou and D. Tikk. Session-based recommendations with recurrent neural networks. 2016.
- A. B. A. J. S. C. Wu, A. Ahmed and H. Jing. Recurrent recommender networks. 2017.
- A. P. Carl Yang. Multisage: Empowering gcn with contextualized multi-embeddings on web-scale multipartite networks. 2020.
- Y. Z. Chen Gao. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. 2023.
- A. Damianou. Podcast recommendations and search query using gnns at spotify | graph learning workshop 2022. 2022.
- J. L. Fei Sun. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. 2019.
- B. Hidasi and A. Karatzoglou. Recurrent neural networks with top-k gains for session-based recommendations. 2017.

- Y. K. Jon Eskreis-Winkler. Xwalk: Random walk based candidate retrieval for product search. 2023.
- A. F. Kirill Khrylchenko. Personalized transformer-based ranking for e-commerce at yandex. 2023.
- G. G. Liudmila Prokhorenkova. Catboost: unbiased boosting with categorical features. 2017.
- C. H. Mengru Chen. Heterogeneous graph contrastive learning for recommendation. 2023.
- A. Z. Nikil Pancha. Pinnerformer: Sequence modeling for user representation at pinterest. 2022.
- Y. B. Petar Velickov. Graph attention networks. 2018.
- T. W. Qiang Cui. Herograph: A heterogeneous graph framework for multi-target cross-domain recommendation. 2020.
- Z. H. Qimai Li. Deeper insights into graph convolutional networks for semi-supervised learning. 2018.
- R. H. Rex Ying. Graph convolutional neural networks for web-scale recommender systems. 2018.
- R. S. Ruoxi Wang. Dcn v2: Improved deep cross network and practical lessons for web-scale learning to rank systems. 2020.
- C. F. S. Rendle and L. Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. 2010.
- N. P. Saket Gurukar. Multibisage: A web-scale recommendation system using multiple bipartite graphs at pinterest. 2022.
- F. S. Shiwen Wu. Graph neural networks in recommender systems: A survey. 2022.
- S. Vaswani. Attention is all you need. 2017.
- J. M. Wang-Cheng Kang. Self-attentive sequential recommendation. 2018.
- J. L. William L. Hamilton. Inductive representation learning on large graphs. 2018.
- K. D. Xiangnan He. Lightgcn: Simplifying and powering graph convolution network for recommendation. 2020.
- J. Y. Xinyang Yi. Sampling-bias-corrected neural modeling for large corpus item recommendations. 2019.
- P. E. Xue Xia. Transact: Transformer-based realtime user action model for recommendation at pinterest. 2023.
- C. H. Xuheng Cai. Lightgcl: Simple yet effective graph contrastive learning for recommendations. 2023.
- H. C. Yue Xu. Single-layer graph convolutional networks for recommendation. 2020.
- X. C. Yunfei Chu. Dynamic sequential graph learning for click-through rate prediction. 2021.
- Q. S. Zhiqiang Wang. Masknet: Introducing feature-wise multiplication to ctr ranking models by instance-guided mask. 2021.