

# 基于融合自注意力的图神经协同过滤模型研究<sup>①</sup>

宋永磊<sup>1</sup>, 曹树国<sup>2</sup>

1. 重庆第二师范学院 数学与信息工程学院, 重庆 400065; 2. 重庆第二师范学院 教务处, 重庆 400065

**摘要:** 针对协同过滤算法中协同过滤信号交互编码数据量过大, 难以有效提取所需的协作信号问题, 本文采用深度学习方法, 提出融合自注意力的图神经协同过滤模型(UAGNCF). 首先将自注意力机制运用到用户与项目之间的消息构建中, 然后将处理后的消息运用高阶连通性嵌入到图神经网络层, 最终将获得的消息聚合后再输出. 在 3 个数据集上进行实验分析, 实验结果表明与现有模型进行对比, 其召回率优于其他模型 5%~9%, 表明本文构建的 UAGNCF 模型合理、有效, 能够取得更好的推荐效果.

**关键词:** 协同过滤; 注意力机制; 高阶连通性; 图神经网络; 消息聚合

**中图分类号:** TP391

**文献标志码:** A

**文章编号:** 1000-5471(2021)07-0130-05

推荐系统一般是通过对用户和项目的参数进行矢量化表示, 并基于模型参数重建用户与商品的交互数据. 协同过滤(CF)模型<sup>[1-2]</sup>由两部分组成, 嵌入部分和交互建模. ①嵌入部分, 将用户和商品转换为向量化表示; ②交互建模, 根据嵌入的内容来重建历史的交互行为. 文献[1]中通过分析用户-商品项目交互矩阵来发现相似的商品项目, 并将这些相似商品项目进行聚合. 文献[2]与文献[1]相似, 区别在于其目标是寻找用户的相似项. 例如, 矩阵分解(MF)<sup>[3-5]</sup>直接将用户/商品 ID(Identity Document)嵌入为向量, 再与交互图进行内积. 在文献[4-5]中, 通过改进隐式反馈的矩阵分解来提升推荐效果. 尽管向量内积可以体现出交互用户和商品的喜好度, 但其线性不足以揭示用户和商品之间的复杂和非线性关系. 因此, 目前一些文献利用深度学习来提高交互关系, 以便得到用户和商品之间的非线性特征交互.

推荐系统的研究除了因式分解模型外, 相关文献提出了一些基于图的推荐模型, 在图上随机游走, 进行推荐<sup>[6-8]</sup>. 模型将用户-项目矩阵视为交互图, 并且在图上随机游走从而将间接偏好进行推荐. 文献[6]提出了 RP3( $\beta$ ), 这是一种通过对图节点 3 次随机游走, 重新对项目进行排名从而进行推荐的方法. 文献[7]扩展了评分算法使推荐更准确、更快速. 文献[9]中提出了基于交互图 G 随机游走参数的各种评分算法.

上述推荐方法对于一些特定用户、项目以及冷启动处理并不令人满意. 问题解决的关键在于嵌入部分的编码和协作交互信号编码, 嵌入编码决定了相关特征的完整性, 协作信号编码揭示了用户、项目之间的行为相似性. 具体而言, 文献[10]只考虑了对用户、项目的特征编码, 再定义模型训练目标函数, 在此方法中忽略了用户和项目之间的相互作用. 因此, 当嵌入不足以捕获协同信号时, 该方法具有交互作用, 可以弥补嵌入不足的缺陷<sup>[11]</sup>. 在文献[12]中, 使用高阶连接编码来解决忽略用户和项目之间相互作用的缺陷, 但仍然忽略了用户和商品的个性化.

通过以上研究发现, 将用户与商品的交互直接集成到编码部分, 会浪费很多不必要的资源. 在实际应用中, 交互行为规模可能会达到百万甚至更大, 从而难以提取所需的协作信号. 而且, 部分模型考虑了项目隐式关联信息却忽略了用户、项目的个性化特征, 导致最终在推荐时效果不理想. 因此, 本文提出融合用户自注

① 收稿日期: 2020-06-30

基金项目: 重庆市自然科学基金项目(cstc2018jcyjAX0810); 西南大学网络与继续教育 2019 年度教学研究项目(SWU1908003); 重庆第二师范学院校级教育教学改革项目(JG201819).

作者简介: 宋永磊, 硕士, 讲师, 主要从事推荐系统及大数据挖掘研究.

意力, 使用交互图结构中协作编码的自然方法, 结合用户特征进行高阶连接来解决此问题.

## 1 本文提出的模型

### 1.1 相关定义

在常见的推荐系统中, 设置  $M$  个用户  $U = \{u_1, u_2, \dots, u_M\}$  和  $N$  个项目  $V = \{v_1, v_2, \dots, v_N\}$ . 用户与项目之间的交互矩阵  $\mathbf{Y} \in R^{M \times N}$

$$\mathbf{R}_{ij} = \begin{cases} 1, & \text{用户 } i \text{ 对商品 } j \text{ 点击} \\ 0, & \text{其他} \end{cases} \quad (1)$$

在交互矩阵  $\mathbf{R}_{ij}$  中, 用户  $i$  对商品  $j$  点击, 用 1 表示; 其他的用 0 来表示, 这其中包括没有点击和未观察到的项目. 然后, 根据一些最新的推荐模型<sup>[14,20]</sup>, 设定用户、项目的嵌入向量分别为  $x_u \in \mathbf{R}^d$ ,  $x_v \in \mathbf{R}^d$ , 这里  $d$  表示嵌入向量维度.  $x_u, x_v$  可以看作是为参数矩阵构建的嵌入查找表.

在这里建立的嵌入表用作用户和商品嵌入的初始状态, 并以端到端的方式进行优化. 比如 MF<sup>[10]</sup> 和 NCF<sup>[11]</sup>, 它们直接将 ID 嵌入到交互层中, 然后用来预测分数. 本文由于细化嵌入步骤将协作信号显式注入到嵌入表中, 提高了嵌入表的表现形式, 从而优化了用户商品交互图的传播嵌入层.

### 1.2 UAGNCF 模型

本文提出了融合自注意力的图神经协同过滤模型. 针对协同过滤的两个关键因素: 数据嵌入和交互建模, UAGNCF 设计了两个部分. ① 对用户、商品项目进行自注意力编码学习个性化特征, ② 使用高阶传播学习用户与商品项目的交互.

### 1.3 自注意机制

目前一些主流的推荐模型, 大多数都在关注用户行为, 忽略了用户、项目个性化特征的潜在联系.

### 1.4 图卷积网络

卷积神经网络(CNN)在计算机视觉、自然语言处理以及其他一些应用中取得了巨大的成功. CNN 主要用于具有二维网格结构的数据, 如图像. 目前的一些研究使用 CNN<sup>[13-17]</sup> 来构建更通用的图结构数据.

### 1.5 高阶嵌入传播层

本节讨论基于 GCN<sup>[16]</sup> 的消息传递体系结构, 捕获沿图结构的协同信号并优化用户和项目的嵌入. 例如, 用户点击了某些商品项目, 这些商品项目可能具有某些商品共同特性. 因此, 可用其共同特性来计算商品项目之间的协作相似性. 用户在对项目进行选择时有很多的关联性, 由于这些隐式关联性无法直接提取, 从而引进了阶连通性<sup>[18]</sup>. 将用户与商品项目的交互信息形成用户商品项目嵌入表示, 然后将其进行聚合. 为此, 本文提出了多种图信息传播规则, 用于与商品项目形成的交互元组  $(u, v)$ , 表示用户  $U$  与项目  $V$  之间的信息构建.

### 1.6 UAGNCF 算法

UAGNCF 模型从用户和商品项目的特征  $X_u, X_v$  开始, 对于每个用户、商品项目的个性化赋予不同的权重信息. 然后, 将个性化特征输入注意力网络, 经过一层全连接层, 将有类似爱好的用户和商品项目聚集起来. 再输入高阶嵌入层, 与提取个性化的特征进行结合. 最后再使用聚合函数进行聚合, 得到嵌入表示.

## 2 实 验

融合用户自注意力的图神经协同过滤模型的基准预测(Baseline), 将 3 个公开数据集与最新的一些研究成果进行实验对比.

### 2.1 数据集

为了评估融合用户自注意力的图神经协同过滤模型的有效性, 在 3 个数据集上进行实验: Gowalla, Yelp2018 和 MovieLens, 数据集是公开可用的, 并且在域、大小和稀疏性方面有所不同.

Gowalla: 这是从 Gowalla 获得的数据集<sup>[19]</sup>, 用户通过签到共享他们的位置.

Yelp2018: 该数据集选自 2018 年的 Yelp 挑战赛<sup>[12]</sup>. 其中, 当地企业如餐馆和酒吧被看成商品.

MovieLens: MovieLens 是由明尼苏达州大学的 GroupLens 项目组开发的, 本文选用的 MovieLens 1M, 包含 6 040 名用户对 3 952 部电影的评价, 其中每个用户至少对 20 个电影进行评分, 约有 100 万条评分数据, 并且提供了用户职业、电影类别、电影时长等辅助信息<sup>[20]</sup>. 该数据集也是电影推荐中广泛使用的标准数据集.

为了确保数据集的质量, 使用 10-core 方法<sup>[21]</sup>, 即保留具有至少 10 种互动的用户和项目. 对于每个数据集, 随机选择每个用户 80% 的历史互动来构成训练集, 并将其余的视为测试集. 从训练集中随机选择 10% 的交互作为验证集调整模型超参数. 对于每个观察到的用户项目交互, 将其视为积极实例, 然后执行否定抽样策略, 将其与用户之前未交互过的一个消极项目配对.

## 2.2 实验设置

评估指标: 对于测试集中的每个用户, 本文将用户与未交互的所有项目视为消极项目. 然后, 除了训练集中使用的积极项之外, 将用户对所有项目的偏好得分进行输出. 为了评估前 K 个推荐和偏好排名的有效性, 采用了两种广泛使用的评估指标<sup>[7,19]</sup>: 召回率 recall @ K 和归一化折扣累积增益 ndcg @ K. 默认情况下, 设置 K=20. 本文实验测试中数据采用用户的平均指标.

实验环境: Ubuntu16.04 的操作系统, Intel Xeon(R) Gold 6126 处理器, RAM 大小为 256 G, 使用 NVIDIA RTX2080Ti 显卡.

参数设置: 本文使用 TensorFlow 实现融合自注意的图神经推荐模型. 模型 Embedding 的大小设定为 64, batch-size 的大小设定为 1 024. 本文使用 Xavier 初始化权重<sup>[6]</sup>、初始化模型参数, RMSprop 优化器对模型进行优化, 并且执行提前停止策略: 如果在验证数据上的 Recall@20 在连续 50 个 epoch 没有增加, 则停止.

## 2.3 实验分析

### 2.3.1 对比分析

本文在 Gowalla, Yelp2018, MovieLens 这 3 个数据集上进行实验, 使用 ndcg 和 recall 作为评定指标 (ndcg 表示查询结果的好坏, 值越大表示越好; recall 表示召回率, 值越大表示越好). 为了验证 UAGNCF 模型的有效性, 将 UAGNCF 与以下模型进行对比实验分析:

MF<sup>[10]</sup>: 通过贝叶斯个性排名(BPR)损失优化的矩阵分解, 该损失将用户项目直接交互, 仅用作交互函数的目标值.

NeuMF<sup>[11]</sup>: 该方法是较新的神经 CF 模型, 该模型使用元素级上方的多个隐藏层以及用户和项目嵌入来捕获其非线性特征交互, 采用两层平原架构, 其中每个隐藏层的尺寸保持不变.

NGCF<sup>[12]</sup>: 将用户-项目交互整合到嵌入过程中, 通过在神经图上传播来挖掘用户与项目的偏好, 从而进行推荐.

将 Gowalla, Yelp2018 和 MovieLens 数据集上 UAGNCF 模型和其他模型 Baseline 之间性能进行比较. UAGNCF 模型与 MF, NeuMF 模型对比, 因 MF, NeuMF 模型内部网络结构不足以捕获用户与商品项目之间的复杂关系, 进而限制了性能. NeuMF 模型在 3 个数据集上均优于 MF 模型, 说明了用户和项目嵌入之间非线性特征交互的重要性. NeuMF 与 NGCF 对比发现 NeuMF 的性能比 NGCF 劣势, 其对隐藏商品项目属性的捕获要弱一些, 从而导致性能降低. UAGNCF 模型与 NGCF 对比, NGCF 使用了高阶连通性, 捕获了商品项目的一些隐藏属性, 因 NGCF 对用户和商品项目的个性化考虑不全面, 导致其性能略低于 UAGNCF. 本文提出的融合自注意力图神经推荐(UAGNCF)模型, 在 3 个数据集上面进行实验, 效果明显优于其他模型.

### 2.3.2 聚合函数分析

为研究嵌入传播层(图卷积层)如何影响性能, 考虑了不同的聚合方法, 并有以下发现:

UAGNCF-gcn: 仅考虑线性变换, 这使得消息传播取决于前面的注意力, 体现了本文嵌入传播层的合理性和有效性.

UAGNCF-bi: 将  $x_u$  和  $x_v$  进行一次交互, 并将其个性化权重加重, 从而更好地进行推荐. 该方法在性能上优于 UAGNCF-gcn, 因其更好地捕获了  $x_u$  与  $x_v$  的亲合力.

UAGNCF-sum; 将 UAGNCF-gcn 与 UAGNCF-bi 连接在一起, 考虑线性变换时增加了用户与商品项目交互的亲合力, 提升了模型的性能, 从而证明了聚合函数的重要性.

### 2.3.3 自注意力分析

为研究自注意力机制如何影响性能, 将不同数据集和不同的聚合方法与 NGCF 模型进行对比, 得出融合自注意力机制的 UAGNCF 模型比 NGCF 的推荐效果更好. 因 UAGNCF 模型考虑了个性化特征, 对用户和商品项目进行注意力编码, 使用户与商品项目能更好地进行交互, 从而得到更好的推荐效果, 结果证明了 UAGNCF 模型引入自注意力机制的合理性与有效性.

### 2.3.4 性能分析

本文对 UAGNCF 与 NGCF 的召回率进行了性能测试. 数据显示 UAGNCF 比 NGCF 具有更好的收敛性. 因为本文在优化小批量图的交互时, 进一步提取了间接连接的用户和商品项目, 结果证明 UAGNCF 比 NGCF 在嵌入空间中执行嵌入传播更具优越性.

## 3 结 语

通过实验分析发现, 在推荐系统中不仅要关注用户-项目交互图, 还要关注用户、项目的一些个性化特征. 本文使用自注意力机制自动学习交叉特征权重, 将以前忽略的特征提取出来加以利用; 然后将学习到的特征再加入到高阶传播层, 学习用户、项目的隐式特征. 通过此模型表明, 推荐系统中用户、商品项目、用户-商品项目交互行为都很重要, 将这 3 者的特征因素综合考虑更加全面. 本文提出的融合自注意力图神经协同过滤 (UAGNCF) 模型增加了用户自注意力机制, 更好地利用用户和商品特性, 将高阶连通性与自注意力机制使用多聚合方式进行聚合, 更好地契合了两者的特征, 从而提高了模型的预测能力和泛化能力.

### 参考文献:

- [1] WEI S Y, YE N, ZHANG S, et al. Item-Based Collaborative Filtering Recommendation Algorithm Combining Item Category with Interestingness Measure [C]//2012 International Conference on Computer Science and Service System. Nanjing: IEEE, 2012.
- [2] ZHANG H W, SHEN F M, LIU W, et al. Discrete Collaborative Filtering [C]//Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. New York: ACM, 2016.
- [3] KOREN Y, BELL R, VOLINSKY C. Matrix Factorization Techniques for Recommender Systems [J]. Computer, 2009, 42(8): 30-37.
- [4] HE X N, ZHANG H W, KAN M Y, et al. Fast Matrix Factorization for Online Recommendation with Implicit Feedback [C]//Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2016.
- [5] KOREN Y. Factorization Meets the Neighborhood: a Multifaceted Collaborative Filtering Model [C]//Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD. New York: ACM, 2008.
- [6] CHRISTOFFEL F, PAUDEL B, NEWELL C, et al. Blockbusters and Wallflowers: Accurate, Diverse, and Scalable Recommendations with Random Walks [C]//Proceedings of the 9th ACM Conference on Recommender Systems. New York: ACM, 2015.
- [7] COOPER C, LEE S H, RADZIK T, et al. Random Walks in Recommender Systems: Exact Computation and Simulations [C]//Proceedings of the 23rd International Conference on World Wide Web. New York: ACM, 2014.
- [8] FOUSS F, PIROTTE A, RENDERS J M, et al. Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation [J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(3): 355-369.
- [9] FOUSS F, PIROTTE A, SAERENS M. A Novel Way of Computing Similarities between Nodes of a Graph, with Application to Collaborative Recommendation [C]//The 2005 IEEE/WIC/ACM International Conference on Web Intelligence. Compiegne: IEEE, 2005.
- [10] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian Personalized Ranking from Implicit Feedback [C]//Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. Montréal: AUAI Press, 2009.

- [11] HE X, LIAO L, ZHANG H, et al. Neural Collaborative Filtering [C]//Proceedings of the 26th International Conference on World Wide Web. Perth; International World Wide Web Conferences Steering Committee, 2017.
- [12] WANG X, HE X, WANG M, et al. Neural Graph Collaborative Filtering [C]//In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19), Paris; ACM, 2019.
- [13] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral Networks and Locally Connected Networks on Graphs [C]//International Conference on Learning Representations (ICLR2014). Banff; CBLS, 2014.
- [14] HAMILTON W, YING Z, LESKOVEC J. Inductive Representation Learning on Large Graphs [C]. California; Advances in Neural Information Processing Systems, 2017.
- [15] HENAFF M, BRUNA J, LeCun Y. Deep Convolutional Networks on Graph-Structured Data [C]. Montréal; Conference on Neural Information Processing Systems, 2015.
- [16] KIPF T N, WELING M. Semi-Supervised Classification With Graph Convolutional Networks [C] Toulon; International Conference on Learning Representations, 2017.
- [17] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks [C]. Vancouver; International Conference on Learning Representations, 2018.
- [18] YANG J H, CHEN C M, WANG C J, et al. HOP-Rec: High-Order Proximity for Implicit Recommendation [C]//Proceedings of the 12th ACM Conference on Recommender Systems. Vancouver; ACM, 2018.
- [19] LIANG D, CHARLIN L, MCINERNEY J, et al. Modeling User Exposure in Recommendation [C]//Proceedings of the 25th International Conference on World Wide Web. Montreal; International World Wide Web Conferences Steering Committee, 2016.
- [20] HARPER F M, KONSTAN J A. The MovieLens Datasets: History and Context [C]. New York; Ksii Transactions on Internet and Information Systems, 2016.
- [21] HE R, MCAULEY J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends With One-Class Collaborative Filtering [C]//Proceedings of the 25th International Conference on World Wide Web. Montreal; International World Wide Web Conferences Steering Committee, 2016.

## Graph Neural Collaborative Filtering Model Research Based on Self-Attention

SONG Yong-lei<sup>1</sup>, CAO Shu-guo<sup>2</sup>

1. College of Mathematics and Information Engineering, Chongqing University of Education, Chongqing 400065, China;

2. Office of Academic Affairs, Chongqing University of Education, Chongqing 400065, China

**Abstract:** Aiming at the problem that the amount of interactive coding data of the collaborative filtering signal in the collaborative filtering algorithm is too large, it is difficult to effectively obtain the required collaborative signal. The deep learning method is used here to propose a graph neural collaborative filtering model fused with self-attention (UAGNCF). First, apply the free attention mechanism to the message integration between users and projects, and then apply high-level connectivity to the embedded graph neural network layer for the processed messages, and finally aggregate the obtained messages before outputting. Experimental analysis on three data sets. The experimental results show that, compared with the existing model, its recall rate replaces other models by 5%—9%, indicating that the constructed UAGNCF model is reasonable and effective, and can obtain better recommendations effect.

**Key words:** collaborative filtering; attention mechanism; high-order proximity; graph neural network; message aggregation

责任编辑 夏 娟