

一种双向注意力的知识图谱神经网络推荐

(河南理工大学 计算机学院, 河南 焦作 454000)

摘要: 在推荐系统中, 基于知识图谱的神经网络推荐以图形作为输入, 可以很好地将节点信息和拓扑结构相结合进行预测。然而现有方法中, 很少考虑图结构中存在的对称关系以及信息聚合时梯度消失的问题, 本文提出一种双向注意力机制的知识图谱神经网络推荐算法, 首先将图神经网络与对称注意力机制相结合, 然后采用双向翻译模型对知识图谱中用户-项目信息进行特征的嵌入表示, 使得注意力机制在决策权重时考虑的关系更全面。其次, 在对节点和邻居信息训练过程中, 为避免过拟合问题引入了多通道激活函数。最后, 在两个真实数据集上与经典算法进行对比, 验证了本文所提出算法的有效性。

关键字: 双向嵌入; 注意力机制; 知识图谱; 推荐

中图分类号: 文献标志码:

A neural network recommender algorithm based on idirectional

Abstract: Comparing with the traditional neural network, the neural network based on knowledge graph (KG) takes the graph as the input in the recommender system, which can well combine the node information and topology for prediction. However, the existing methods rarely consider the symmetry relationship in KG and the disappearance of the gradient when the information is aggregated. A Neural Network Recommender Algorithm Based on Bidirectional Graph Attention (BAGNR) is proposed. Firstly, the graph neural network with the symmetrical attention mechanism and uses the bidirectional translation model are combined to embed the KG information, so that the KG relations are considered comprehensively. Then, to avoid too many parameters which will lead to the increase of calculation and the problem of overfitting in the process of training, a dynamic activation function is introduced. Empirical results on two benchmark datasets demonstrate our model outperforms state-of-the-art methods.

Keywords: Bidirectional embedding; Attention mechanism; Knowledge graph; Recommender

0 引言

作为缓解信息过载问题的有效工具, 推荐系统能够弥补搜索引擎不能实现真正个性化服务的短板, 已广泛应用于电子商务、社交媒体、广告系统等各种在线系统。推荐系统能够在用户没有准确的需求时, 从用户的历史交互信息捕捉用户兴趣偏好, 提供个性化推荐服务。协同过滤是实现个性化推荐的经典算法^[1], 它包含两个关键部分: 一是将用户和项目的交互信息转换为矢量表示; 二是基于转化为矢量的嵌入信息构建历史交互序列。在众多协同过滤

收稿日期: 2020-02-22 修回日期: (不用填写)

基金项目: 国家自然科学基金资助项目(012345678); 国家重大科技专项项目(123456789)

第一作者简介:

通讯作者简介:

算法中, 矩阵分解^[2,3]是基于协同过滤的主流方法之一, 它将用户和项目投影到一个共享的潜在空间中, 利用潜在的特征向量来表示用户或者项目, 采用内积的形式进行建模^[4]。与此相关的许多工作都致力于增强矩阵分解, 例如基于邻居的模型与项目内容的主题相结合^[5]以实现特征的通用建模。尽管矩阵分解对于协同过滤有效, 但矩阵分解性能会受到交互函数线性内积简单选择的影响。为了解决矩阵分解内积简单选择的影响, He 等人^[6]提出了神经协同过滤模型, 使用深层神经网络从数据中学习交互函数。基于翻译的协同过滤模型使用欧几里德距离度量作为交互函数^[2]。尽管这些方法是有效的, 但在面对冷启动问题仍具有欠缺, 并不能产生足以令人满意的协同过滤嵌入。现有的方法只利用描述性特征来构建嵌入表示, 而不考虑用户与用户之间潜在的关系, 将每个交互看作独立建模, 不足以从中提取具体的协同信号。

为了捕获协同知识图中的高阶隐式关系。有学者利用神经注意机制^[7,8]来获取传播过程中每个邻居的权重, 通过级联传播的注意力权重可以揭示高阶连通性的重要性。He 等人^[9]以端到端的方式显式地建模了知识图谱中关系的高阶连通性, 但该方法嵌入实体的所有近邻实体的语义信息, 导致计算量过于庞大。除此之外, 由于在注意力传播层是以尾实体到头实体的形式传播信息, 由于信息的传播是双向的, 并没有充分体现出信息的传播的完整性。

本文针对上述研究所存在的问题, 提出一种双向注意力机制的知识图谱神经网络推荐算法(BAGNR), 使得在不增加额外数据集中维度情况下, 通过双向嵌入翻译模型获取用户-项目间的高阶关系; 同时为了避免梯度消失问题, 采用多通道的激活函数针对不同的高阶关系进行修正, 从而增加反馈的信息量, 避免训练过程中的过拟合问题; 最终通过实验验证本文提出的算法有效地提高了推荐质量。

1 相关工作

协同过滤算法^[9]将用户和项目交互信息矢量化作为嵌入层, 并重新构造用户和项目的历史交互关系来学习嵌入参数。矩阵分解模型将用户或项目投射到嵌入的向量进行预测^[2]。基于神经协同过滤^[6]学习用户与项目潜在的交互关系, 并采用共享的嵌入层, 来增强神经网络的交互建模。Zhou 等人^[10]提出 DIN 模型, 从用户兴趣多样性和局部信息的相关性出发, 对每个用户的兴趣表示赋予不同的权值。Shi^[11]等人提出在推荐系统中使用自注意力机制, 把推荐问题当作一个序列的问题, 并对历史交互中的物品的序列的相关性进行加权, 来预测下一个可能发生交互的物品。

为了挖掘用户-项目间的高阶信息, 将注意力机制与知识图谱结合以抽取用户-项目之间的特征信息成为学术界研究的热点。Chen 等人^[7]提出将注意力机制与协同过滤相结合, 利用分层的神经网络, 对用户的偏好分数进行建模, 解决了项目隐式反馈问题。基于特征的推荐算法^[10]从知识图谱中抽取一些用户和物品的属性作为特征。基于异构信息网络的推荐方法^[11]是将知识图谱视为一个异构信息网, 然后构造物品之间的基于元路径或元图的特征。Wang 等人^[12]提出一种基于 RNN 的推荐模型, 在知识图谱中对用户和物品之间存在的关联路径进行建模, 并向用户提供可解释性的推荐。

然而, 上述方法虽然充分直观地利用了知识图谱的网络结构, 很少考虑项目属性之间的关系信息, 而且大多数需要人工来计算路径从而在时间中很难以达到最优。

2 BAGNR 模型

BAGNR 模型主要包括嵌入层、修正注意力传播层以及预测层等三部分组成, 如图 1 所示。(1) 嵌入层: 在知识表示过程中, 考虑到用户-项目之间的双向性, 通过三元组翻译模

型 TranE^[8]对实体进行嵌入表示以提高实体的度量精度；(2) 修正注意力传播层：从节点的邻域中进行嵌入以进行更新处理，再利用修正后的知识注意力机制来获取传播过程中每个邻居的权重，深层次挖掘用户-项目-用户、用户-项目-用户-项目等高阶关系，并结合注意力机制模型来揭示节点的高阶连通性；(3) 预测层：从所有传播层聚合用户和项目的预测的匹配值。

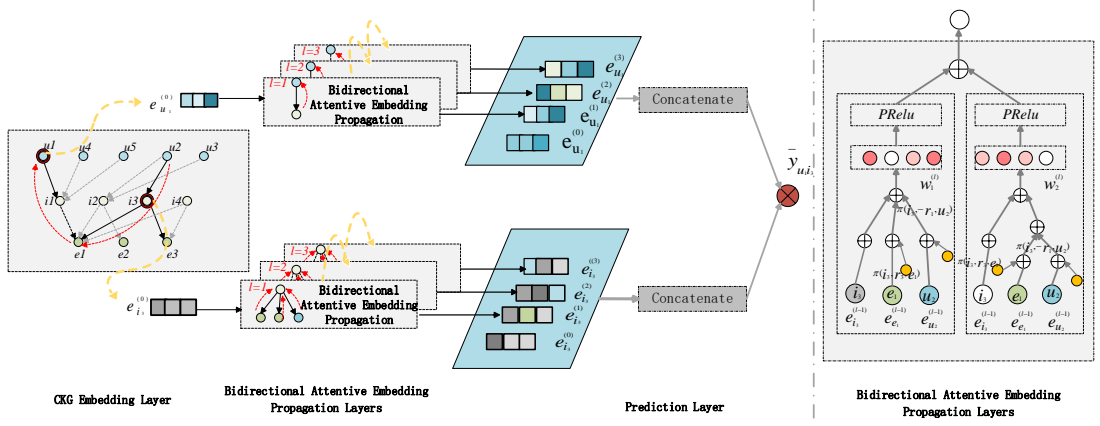


图 1 BAGNR 模型框架

2.1 嵌入层

知识特征表示是将知识图谱中的实体和关系嵌入到低维稠密的向量空间中通过对图中实体和关系的三元组结构进行嵌入表示，保留其在原始空间中拓扑结构，对于知识图谱的扩展及推理能力的提升具有重大意义。用一个三元组(头实体 h ，关系 r ，尾实体 t)代表一条知识，例如三元组(中国首都，地点，北京)表示中国的首都是北京，对于图中每一个存在的三元组(h, r, t), 实体和关系采用基于翻译的表示方法^{错误!未找到引用源。}进行嵌入, 如式(1)所示。

$$e_h^r + e_r \approx e_t^r \quad (1)$$

其中, e_h^r, e_t^r 为 e_h, e_t 在关系空间 r 中投影向量, $e_h, e_t \in \mathbb{R}^d$ 和 $e_r \in \mathbb{R}^n$ 分别表示 h, r, t 的初始化向量。

2.2 修正的注意力机制

知识图谱中蕴含着丰富的语义信息，使用 TransR 将知识图谱中的语义信息向量化，将实体投影到关系空间中。

$$\pi(h, r, t) = (W_r e_t)^T \tanh((W_r e_h + e_r)) \quad (2)$$

其中, e_h, e_t 是在关系 r 空间投影表示, $W_r \in \mathbb{R}^{k \times d}$ 是关系 r 的转换矩阵, k, d 分别表示评分矩阵的用户数和项目数。

由于 TransR 模型只考虑了节点的单向关系，而现实环境中存在反馈信息具有双向关系的特殊，同时借鉴文献[9]，本文提出了双向加性注意力机制，如公式(3)所示。

$$\pi(h, r, t) = \alpha(W_r e_t)^T \tanh((W_r e_h + e_r)) + (1 - \alpha)(W_r e_h)^T \tanh((W_r e_t + e_r)) \quad (3)$$

其中, α 是权重参数。改进后的注意力机制不仅关注于从头实体尾实体这种单向的关系，还对从尾实体到头实体可能存在的双向关系进行了关注，使得注意力机制在决策权重的时候，考虑关系更全面。

最终采用 *softmax* 函数对于 h 相连的所有三元组进行归一化，如公式(4)所示。

$$\pi(h, r, t) = \frac{\exp(\pi(h, r, t))}{\sum_{(h, r', t') \in N_h} \exp(\pi(h, r', t'))} \quad (4)$$

2.3 信息聚合的优化

信息聚合在最终模型的预测评分过程中起着重要的作用，考虑到不同路径的高阶关系具有不同的权重，非线性 $PRelu$ 作为激活函数，如公式（5）所示。

$$PRelu(x_i) = \begin{cases} x_i & \text{if } x_i > 0 \\ \beta_i x_i & \text{if } x_i \leq 0 \end{cases} \quad (5)$$

其中， i 表示不同的通道数。对于第 i 个通道拥有一个 β_i ，并且这些 β_i 会随着数据进行修正，从而增加反馈的信息量。同时也避免了引入过多的参数导致的计算量增加和过拟合问题，考虑划分通道，共享部分参数。本文考虑让每次聚合过程中线性变换的参数矩阵中每一列划为同一个通道。

3 实验

3.1 数据集

在 Amazon-book 和 Last-FM 的两个公开的数据集上评估本文所提出 BAGNR 模型的性能。为了保证数据集的质量，本文保留了用户和项目至少有过交互的数据信息，表 1 为数据集的统计信息，由用户项目交互数据信息和项目属性信息所组成。用户项目交互信息由用户数、项目数以及交互总条数构成，项目属性信息由实体数、关系以及三元组个数构成。除了用户和项目之间的交互信息外，还需要为每个数据集构造项目知识图谱。在数据集中，如果存在映射关系，则通过标题匹配的方式将项目映射到自由实体。与目前大多数只提供单跳的感知数据集不同，还将包含双跳相邻的实体的三元组考虑在内。为了保证知识图谱的高质量，筛选出了不常见或者交互信息量少的实体。实验过程中选择 80% 的交互历史信息作为训练集，其余作为测试集。同时采用了交叉 K 折交叉验证，从训练集中随机的选择 10% 的交互历史信息作为验证。

表 1 数据集描述

Tab.1 Statistics of the datasets

	Amazon-book	Last-FM
User-Item Interaction	#user	70679
	#Item	24915
	#Interactions	847733
	Density	4.81E-04
Knowledge Graph	#Entities	88572
	#Relations	39
	#Triplets	2557746

3.2 基准算法

- BPRMF^[13]针对用户的隐式反馈信息解决推荐排序问题。
- NFM^[6]错误!未找到引用源。结合深度神经网络和 FM 建模更高阶的特征之间的关系。
- CKE^[14]结合协同过滤学习节点的潜在表示。
- CFKG^[15]构造用户项目知识图并结合用户行为和物品信息进行推荐。
- KGAT^[9]强调协同知识图谱中显式建模高阶关系信息，以用于个性化推荐。

3.3 评价指标

推荐系统的优劣需从多个角度去衡量，本文针对用户个性化推荐列表的 Top-N 推荐，包括召回率 (Recall)和归一化折损累计增益 (NDCG)等评价指标，对整个测试集中的用户及其推荐列表结果进行评估。

Recall 如式(6)所示。

$$Recall = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (6)$$

NDCG 如式(7)所示。

$$NDCG @ k = \frac{DCG @ k}{IDCG @ k} \quad (7)$$

其中，DCG 如式(8)所示， rel_i 为单个相关性分数。

$$DCG @ k = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (8)$$

3.4 实验对比与分析

本部分主要将 BAGNR 算法与经典基准算法性能做以比较并进行分析，本文 BAGNR 模型深度设置为 3 层，采用 Adam 优化模型。嵌入向量的大小设置为 64，批处理大小为 1024，学习率调整范围在 $\{10^{-3}, 10^{-2}, 10^{-1}\}$ ，正则化参数系数，初始化为 0。对于本文提出 $f_{Agg}(\cdot)$ 中 $prelu$ 函数通道参数 i 初始化为 0.5， i 随着数据集及迭代过程会进行修正。

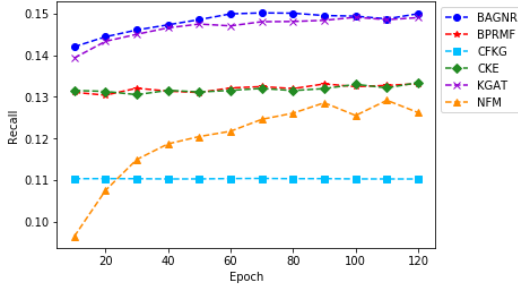


图 2 Amazon-book 数据集上不同算法 Recall 对比

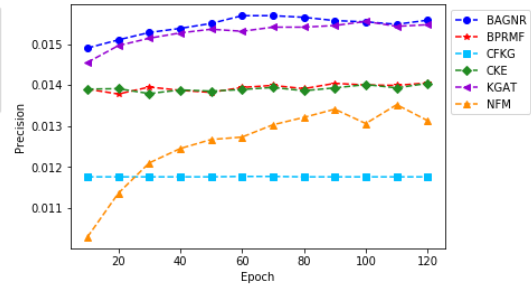


图 3 Amazon-book 数据集上不同算法 NDCG 对比

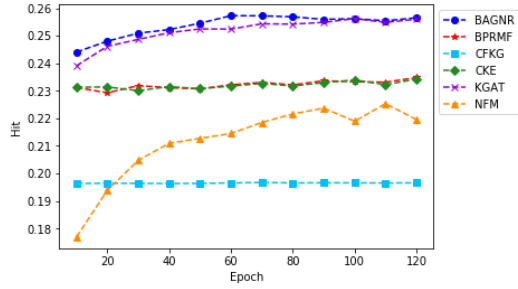


图 4 Last-FM 数据集上不同算法 Recall 对比

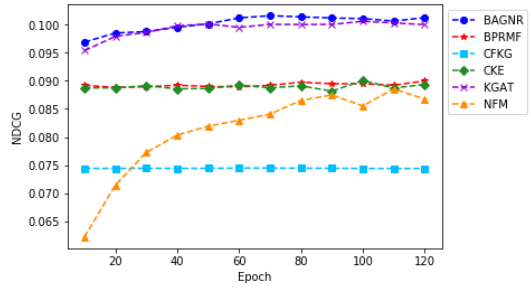


图 5 Last-FM 数据集上不同算法 NDCG 对比

图 2-图 5 是本文提出的 BAGNR 算法分别在 Amazon-book 及 Last-FM 数据集上与 5 个基准算法的 Recall 和 NDCG 随着迭代次数的变化情况。在 Amazon-book 数据集中，BAGNR 与 KGAT 相比，Recall 和 NDCG 提高并不明显，原因在于 Amazon-book 数据集中双向信息并不明显。但在 Last-FM 数据中 Recall、NDCG 指标相比其它 5 个基准算法有明显的提升，分别提高了 2.56%、1.96%。这验证了模型通过叠加双向注意力嵌入传播层，能够以显式的方式双向探索高阶连通性，以便捕捉有效的协同信号，可以捕获协作信号以传递知识。进一步与 KGAT 相比，BAGNR 证明了该方法的有效性，其可以进行双向信息挖掘，而不是单方向。

3.5 消融实验及最优参数确定

选取适合模型的参数是提高模型表现的一个重要指标，本算法通过在 Amazon-Book 数据集进行消融实验来确定 α 和 β 的最优值，在 Last-FM 数据集进行验证。

3.5.1 最优 α

为了寻找合适的参数值，首先将设置 β 为 0.5，对于 $PReLU$ 激活函数来说，是一个相对对称的一个参数。并在 $\{0.8, 0.6, 0.4, 0.2\}$ 调整参数 α 。Amazon-Book 数据集中以 Recall@20 和 NDCG@20 指标作为评测依据，与 KGAT 算法在不同的迭代次数（250 次）对比所绘曲线如图 6 和图 7 所示。各个分图的参数为 $\beta=0.5$ ， α 分别从 $\{0.8, 0.6, 0.4, 0.2\}$ 集合中对应取值。在各个指标的对比图中可以观测到，随着迭代次数的增加，BAGNR 算法表现优秀，并在 $\alpha=0.8$ 时，算法表现性能最佳。

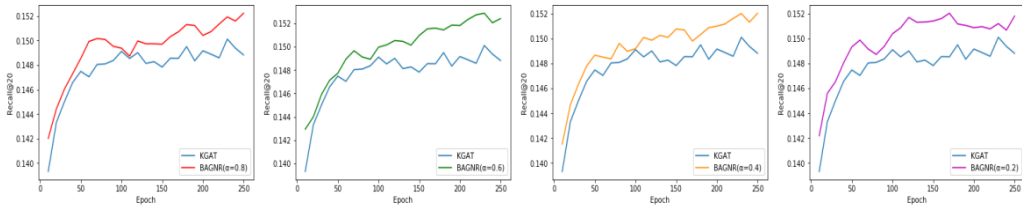


图 6 Recall@20 指标下 α 的变化

Fig.6 Changes in α under the Recall@20

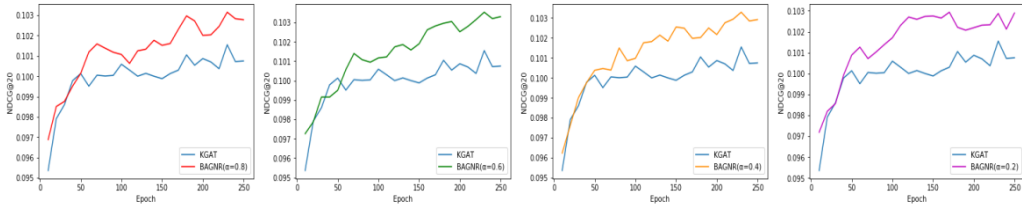


图 7 NDCG@20 指标下 α 的变化

Fig.7 Changes in α under the NDCG@20

表 2 是在设置不同 α 值下，不同指标的最优结果，其中第二行 Origin 是原始 KGAT 算法的各项指标，从表格中可以看出，当 $\alpha = 0.8$ ，在 Recall@20, Recall@100, NDCG@20, NDCG@100 均最优的，因此最终选择 $\alpha = 0.8$ 为最佳参数，在表 2 最后一行，列出了每项指标相对于原始 KGAT 算法改进提高率。

表 2 Amazon-book 不同评价标准下的 α 值

Tab.2 The best α value of Amazon-book in different metrics

α	Recall@20	Recall@100	NDCG@20	NDCG@100
KGAT	0.15001	0.31689	0.10155	0.15561
0.8	0.15315	0.31940	0.10362	0.15748
0.6	0.15287	0.31852	0.10353	0.15731
0.4	0.15214	0.31819	0.10320	0.15710
0.2	0.15204	0.31860	0.10293	0.15715
$\alpha = 0.8$	2%	0.5%	2%	1%

3.5.2 最优 β

基于上述实验确定的 α ，在本节固定 α 的值，调整 β 的值以寻找合适参数，选取了三个值进行了实验，

分别是 0.3、0.5 以及 0.7。在 Amazon-Book 数据集进行实验。实验结果如表 3 所示，通过对比发现，当 $\beta = 0.5$ 时各个指标的效果最好。

表 3 Amazon-book 不同评价标准下 β 值

Tab.3 The best β value of Amazon-book in different metrics

β	Recall@20	Recall@100	NDCG@20	NDCG@100
0.3	0.15052	0.31826	0.10296	0.15693
0.5	0.15015	0.31858	0.10362	0.15731
0.7	0.14992	0.31858	0.10297	0.15700

3.5.3 验证模型

在 3.5.1 和 3.5.2 中，通过多次实验找到了适合模型的参数 α 和 β ，为了验证提出的模型的可行性和准确性，在 Last-FM 数据集将本文提出的算法 BAGNR 与 4.3 中最优的模型 KGAT 进行对比，进而可以验证了 BAGNR 的可行性。各项指标对比结果如图 8 所示，BAGNR 算法优于 KGAT 算法，这也就验证了 BAGNR 算法选择参数的有效性和算法可行性。

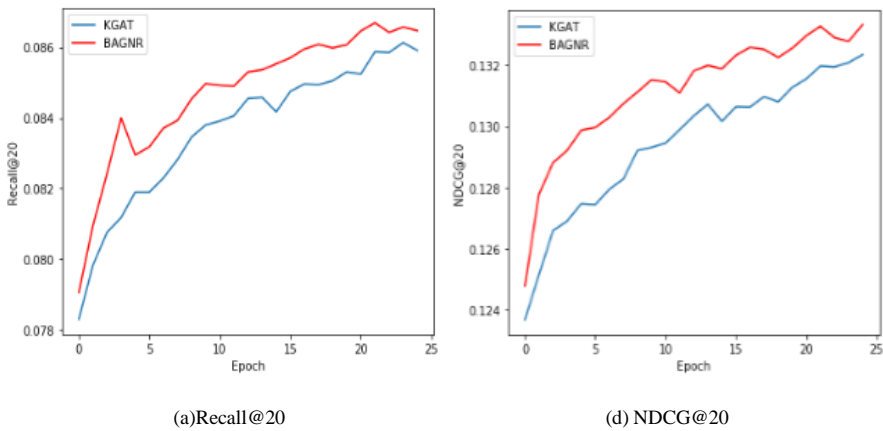


图 8 KGAT 与 BAGNR 算法评价标准对比

Fig8. Comparison of KGAT and BAGNR algorithm metrics

4 结论与展望

本文在基于知识图谱的推荐算法基础上，提出一种双向知识表示的嵌入方法，不仅考虑到从头实体到尾实体的距离，而且还考虑到从尾实体到头实体的距离。在传播层，首先获取项目与用户间的高阶关系，然后通过注意力机制来定义不同路径的权重关系。在信息聚合层，根据高阶关系设置 PReLU 激活函数不同通道避免梯度消失和过拟合。在 Amazon-Book 和 Last-FM 两个公共的数据集验证了本文提出算法的有效性。在未来的工作中，一方面将考虑异构数据间的用户/项目的高阶路径的获取，另一方面在推荐过程中，将用户的时序作为附加信息挖掘在不同时间段推荐的结果，提高推荐精度。

参考文献

[1] Travis Ebesu, Bin Shen, and Yi Fang. Collaborative Memory Network for Recommender Systems[C]. //In SIGIR 2018.515–524.

[2] Yehuda Koren, Robert M. Bell, and Chris Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems [J]. IEEE Computer,2009,42 (8), 30–37.

[3] 王留芳,刘镇镇等. 基于双因子混合加权相似度的协同过滤推荐算法[J]. 河南理工大学学报（自然科学版）, 2020,39(6):133-138.

WANG Liufang, LIU Zhenzhen, et.al. Collaborative filtering recommendation algorithm based on double-factor hybrid weighted similarity [J]. Journal of Henan Polytechnic University(Natural Science), 2020,39(6):133-138.

- [4] Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Latent relational metric learning via memory-based attention for collaborative ranking[C]. //In WWW, 2018, 729–739.
- [5] Y. Koren et.al.. Factorization meets the neighborhood: A multifaceted collaborative filtering model[C]. //In KDD, 2008, 426–434.
- [6] Xiangnan He, Lizi Liao, et.al. Neural Collaborative Filtering[C]. //In WWW, 2017. 173–182.
- [7] Jingyuan Chen, et.al.. Attentive Collaborative Filtering: Multimedia Recommender with Item- and Component-Level Attention[C]. // In SIGIR, 2017. 335–344.
- [8] 王建芳, 苗艳玲. 一种融合信任和项目卷积描述信息的PMF算法[J]. 控制与决策, 2020,35(8):1803-1812.
WANG Jian-fang, MIAO Yan-ling. A PMF algorithm based on trust and item convolutional description information [J]. Control and Decision, 2020,35(8):1803-1812.
- [9] Wang, H., et al. Knowledge Graph Convolutional Networks for Recommender Systems. arXiv (2019).
- [10] Zhou G, Song C, et al. Deep Interest Network for Click-Through Rate Prediction[J]. arXiv preprint arXiv:1706.06978,2017.
- [11] Shi C, Hu B, et al. Heterogeneous Information Network Embedding for Recommendation[J]. IEEE Transactions on Knowledge & Data Engineering, 2017:125-136.
- [12] Wang X, Wang D, et al. Explainable Reasoning over Knowledge Graphs for Recommendation[C]. //In AAAI 2008:565-668.
- [13] Rendle S, Freudenthaler C, et al. BPR: Bayesian personalized ranking from implicit feedback[C]// UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence,2009. Montreal, QC, Canada: AUAI Press, 2009,452-461.
- [14] Zhang F, Yuan N J, et al. Collaborative Knowledge Base Embedding for Recommender Systems[C]// the 22nd ACM SIGKDD International Conference. ACM, 2016, 236-245.
- [15] Y. Zhang, Q. Ai, X. Chen, and P. Wang, Learning over knowledge-base embeddings for recommendation, arXiv preprint arXiv:1803.06540, 2018.