

控制与决策

Control and Decision

基于偏差的图注意力神经网络推荐算法

王建芳, 文茜琳, 杨许, 张秋玲

引用本文:

王建芳,文茜琳,杨许,张秋玲. 基于偏差的图注意力神经网络推荐算法[J]. *控制与决策*, 2022, 37(7): 1705–1712.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1626>

您可能感兴趣的其他文章

Articles you may be interested in

结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

基于反时限混沌郊狼优化算法的BP神经网络参数优化

Parameter optimization of BP neural network based on coyote optimization algorithm with inverse time chaotic

控制与决策. 2021, 36(10): 2339–2349 <https://doi.org/10.13195/j.kzyjc.2020.0408>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

基于生物启发神经网络和DMPC的多机器人协同搜索算法

Multi-robot cooperative search algorithm based on bio-inspired neural network and DMPC

控制与决策. 2021, 36(11): 2699–2706 <https://doi.org/10.13195/j.kzyjc.2020.0959>

基于地标特征和元学习方法推荐最适用优化算法

Recommending best suitable metaheuristic based on landmarking feature and meta-learning approach

控制与决策. 2021, 36(5): 1223–1231 <https://doi.org/10.13195/j.kzyjc.2019.0993>

基于偏差的图注意力神经网络推荐算法

王建芳[†], 文茜琳, 杨 许, 张秋玲

(河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

摘要: 在推荐系统中, 基于知识图谱的神经网络与传统神经网络相比, 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推理和推荐. 然而, 现有基于图神经网络的推荐算法, 面临着知识表示不准确以及信息融合单一的问题. 对此, 将图神经网络与注意力机制相结合, 提出一种基于偏差的图注意力神经网络推荐算法. 该方法采用翻译模型对知识图谱信息进行特征表示的嵌入, 获取节点在同一投影空间下的三元组信息, 考虑到在三元组中预测值与真实值之间存在误差, 以及邻居节点在信息传播时权重的差异, 采用基于偏差的注意力计算机制以便更好地捕获节点间高阶连通性. 其次, 在神经网络的传播训练过程中, 通过多通道融合机制对节点和邻居信息进行聚合以提高模型的健壮性. 最后, 在 3 个真实数据集上与经典算法进行对比, 验证所提出算法的有效性.

关键词: 知识图谱; 注意力机制; 神经网络; 推荐; 偏差

中图分类号: TP273

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1626

引用格式: 王建芳, 文茜琳, 杨许, 等. 基于偏差的图注意力神经网络推荐算法[J]. 控制与决策, 2022, 37(7): 1705-1712.

A bias-based graph attention neural network recommender algorithm

WANG Jian-fang[†], WEN Xi-lin, YANG Xu, ZHANG Qiu-ling

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: In the recommender system, compared with the traditional neural network, the neural network based on the knowledge graph can combine node information and topological structure to infer and recommend with graphs as input. However, the existing recommender algorithm based on graph neural networks faces the problem of inaccurate knowledge representation and single information fusion. Combining graph neural network with attention, the bias-based graph attention neural network recommender (BGANR) algorithm is proposed. First, the translation model is used in the feature representation to gain the triple information of the node in the same projection space. Considering the error between the predicted value and the true value in the triple, and the difference in the weight of neighboring nodes during information dissemination, the bias-based attention method is used to better capture the high-order connectivity of nodes. Then, in the propagation training of the neural network, the node and neighbor information is aggregated through the multi-channel fusion mechanism to increase the robustness of the model. Finally, comparison with the state-of-the-art algorithms on two real datasets verifies the effectiveness of the BGANR algorithm.

Keywords: knowledge graph; attention mechanism; neural network; recommender system; bias

0 引言

随着信息技术, 尤其是电子商务的迅猛发展, 商品数量的增长远远超出个人的选择范围, 以满足用户个性化需求而著称的推荐系统越来越受到学术界和工业界的重视. 推荐系统是针对用户搜索提供一个较小且与兴趣高度关联的项目集合. 协同过滤算法作为推荐系统的经典方法, 通过分析用户的显示或隐式反馈, 过滤掉不感兴趣的物品, 进而达到预测推荐的目的. 然而, 基于传统的协同过滤方法通常存在用户项目交互数据稀疏的问题.

近年来, 深度学习在推荐系统领域取得较好的效果^[1], Google 团队^[2]利用用户历史交互行为及隐式特征(上下文信息等), 粗略地从所有视频集合中筛选出用户感兴趣的候选集合, 并对候选集进行细粒度的排序, 使用加权逻辑回归预测用户的观看时长, 将推荐问题转化为分类问题, 最终给用户提供更精准的项目信息. He 等^[3]设计了基于神经网络的协同过滤通用框架, 在该框架下结合矩阵分解和多层感知器, 提出了一种新的神经矩阵分解模型, 使用大量潜在因子来缓解由矩阵分解使用简单内积带来的限制. Ying

收稿日期: 2020-11-24; 录用日期: 2021-04-21.

[†]通讯作者. E-mail: wangjianfang@hpu.edu.cn.

等^[4]提出 PinSage 算法,将随机游走与图卷积相结合,通过对节点周围的邻域进行采样并动态构建计算图,执行局部卷积以减轻在训练时对内存的占用,并成功将图卷积神经网络应用于大规模真实的场景.尽管上述方法已成功融入推荐系统中,但也无法避免数据稀疏性的问题.随后一些学者对节点的属性信息研究指出,节点的属性不是孤立存在而是相互关联的,进而可形成一个相对完整的知识图谱^[5-6].

知识图谱作为辅助信息以缓解数据稀疏问题. Wang 等^[7]将用户项目交互图和属性信息融合到一个空间中,打破项目与其属性之间的相互独立的假设,在混合结构中以端到端的方式挖掘节点之间的高阶连通性,但在数据信息聚合时未考虑噪音带来的影响. Song 等^[8]将图注意力融入神经网络模型中,对用户社交圈进行了建模,使用循环神经网络探索用户的动态行为,从而根据用户的当前兴趣动态地推断影响因素,虽然利用注意力机制对用户当前兴趣进行了有效聚合,但却无法衡量用户真实值和预测兴趣值之间的误差.

为了解决上述问题,本文的贡献如下:

- 1) 提出基于知识图谱的神经网络优化方法,增强了模型对节点嵌入及聚合特性扰动的鲁棒性.
- 2) 在对知识图进行三元组嵌入表示时,将节点真实值与预测值之间存在的误差称为偏差,考虑到节点本身和邻居节点的关系,提出基于偏差的注意力计算方法,以便能够更好地捕获节点间高阶连通性.
- 3) 在神经网络信息传播过程中,使用多通道融合函数对节点及邻居信息进行聚合,以提高模型的健壮性.
- 4) 在3个真实数据集上与经典算法比较的实验结果表明了本文所提出算法的有效性.

1 相关工作

1.1 深度学习与推荐系统

深度学习在推荐系统获得了巨大的成功.例如, Wang 等^[9]提出用于图像分类的自动卷积神经网络(CNN)推荐系统,通过定量分析训练数据,能够准确地评估分类任务的复杂性和 CNN 模型的分类能力. He 等^[10]提出轻量级图卷积网络(LightGCN)的推荐模型,在用户项目交互图上线性学习节点的嵌入表示,实验验证其泛化能力更好.

基于深度学习的注意力机制与人类的视觉注意力有着相似的直觉,它只关注目标最重要的部分,最近的一些研究也表明了它在推荐任务中的能力,如 Veličković 等^[11]在卷积神经网络(GCN)的基础上引

入注意力机制(GAT),对不同的相邻节点分配相应的权重,避免重要数据在传播过程中丢失. Sun 等^[12]提出基于图神经网络和自注意力机制的图上下文自注意力模型,对相邻节点间存在的局部依赖关系进行改进.而本文工作与现有方法^[7]之间的主要区别是在神经网络嵌入传播中对三元组的预测节点和尾节点贡献程度加以区分,并使用多通道信息融合函数对节点信息进行聚合,以提高节点之间的聚合度和模型的健壮性.

1.2 知识图谱与推荐系统

知识图谱(knowledge graph, KG)用于表示多个领域的大规模信息的一种实用方法. KG 是一个异构网络,因为它在图中包含多种类型的节点和关系,并具有很强的表征能力,可跟踪图中不同的边以获得某个实体的多个属性,通过相应连接属性探索实体间的高阶关系.谷歌于2012年将 KG 引入搜索框架,挖掘搜索词语的潜在知识关系,呈现更加结构化的查询结果^[13]. Shi 等^[14]提出了基于异质信息网络表示的推荐算法,充分挖掘异质信息网络中用户和物品的潜在结构特征.基于翻译的推荐方法,是根据知识图的特征进行学习,将每个实体映射成低维向量并进行关系学习,以确保原始语义信息或结构依旧对应. Chen 等^[15]提出基于社交翻译深度学习模型,融合多层翻译模块来建模以表示用户的个人偏好,从而进行推荐.然而,现有的方法在对知识进行特征表示时,只是简单地处于不同空间中三元组的预测值和真实值赋予相同的权重,较少考虑到二者之间存在的偏差.

2 BGANR 算法

2.1 BGANR 算法框架

本文所提出的 BGANR 框架主要由嵌入层、BGANR 层和预测层组成.在嵌入层,引入项目属性的知识图谱,结合用户项目交互图组成协同知识图,在此基础上为每一个节点进行初始化嵌入, e_i 和 e_u 分别嵌入用来表示项目节点和用户节点;在 BGANR 层,基于图注意力网络,设计了基于偏差的三元组注意力计算模块(TriA),以便更好地区分在同一投影空间下真实节点和预测节点带来的误差,可通过 TriA 计算得到邻居节点不同的权重参数.考虑到节点信息高效率的聚合,设计了多通道融合函数对节点 e_i 和邻居信息 e_{N_h} 进行聚拢,并对协同知识图中的节点交互信息和高阶关系进行建模,获得每层的嵌入表示 e_h ;在预测层,分别级联用户 e_u 和项目 e_i 不同层的节点表示进行内积,输出预测分数 $y_p(u, i)$. BGANR 框架如图1所示.

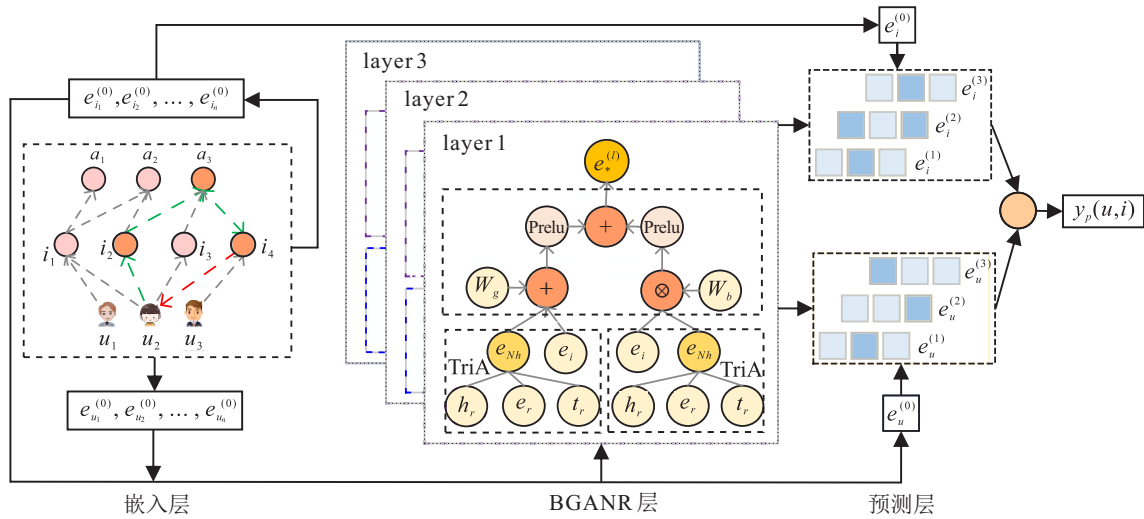


图1 BGANR框架结构

2.2 嵌入层

嵌入层主要对协同知识图中的实体和关系转化为低维表示,以最大程度上保持图的结构信息. 将知识图谱引入推荐系统的方法可分为基于多源信息的表示方法^[16-18]和基于翻译的表示方法^[19]. 基于多源信息的表示方法,通过引入知识图谱之外的文本信息来增强算法模型的学习能力,但是对于多源信息的获取增加了模型的复杂度,不易对其进行扩展. 而基于翻译的表示方法,可以在投影空间中保持关系平移不变的特性,最大程度上保证三元组的结构信息,且该模型相对简单,适合大规模的网络架构.

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^[20]. 用一个三元组(头实体 h , 关系 r , 尾实体 t)代表一条知识,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^[21]进行嵌入,即

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

在协同知识图中,一个实体包含了多种属性信息,即实体和属性之间存在多种关系. 一般认为,相似的实体在实体投影空间中应该彼此靠近,在对应的关系投影空间中应该彼此远离. 本文采用翻译模型 TransR 对实体关系分别进行嵌入表示,投影到不同实体和关系空间,选用 L_2 正则化以防止过拟合,对于给定的三元组 $(h, r, t) \in G$, 通过矩阵 M_r 将实体向量投影到关系空间,在对应的关系空间中完成翻译,即 $h_r = e_h M_r, t_r = e_t M_r$, 预测评分如下式所示:

$$\varsigma(h, r, t) = \|h_r + e_r - t_r\|_2^2. \quad (2)$$

其中 $\varsigma(h, r, t)$ 的得分代表了 (h, r, t) 是否真实存在,分值越低,表示在投影空间中的偏差越小. 根据知识图中节点的类型、项目或用户,可获得相应的嵌入表示

e_i, e_u , 如图1中嵌入层部分所示.

2.3 BGANR层

在获得项目节点的嵌入表示后, BGANR 模块首先基于注意力机制去学习某一实体的近邻节点的权重 e_{N_h} , 并聚合其邻居信息以更新项目的表示,旨在探索实体之间的高阶连通性. 考虑到实体自身与邻居的关系,设计多通道融合函数对数据信息进行整合,避免嵌入不相关的实体信息. 这里先讨论单层 BGANR 的嵌入传播,之后再推广到多层.

单层结构包括: 信息收集、TriA 注意力机制和信息整合.

1) 单层结构.

在信息收集阶段,一个实体可能包含多个属性信息. 用 $N_h = \{(h, r, t) | (h, r, t) \in G\}$ 来表示以头节点为 h 的三元组集合,其中 G 表示协同知识图. 考虑到不同属性对于头节点 h 的重要程度不同,在此阶段为其分配不同的权重,以构建节点 h 的一阶连通性并捕获邻居间的特征交互,有

$$e_{N_h} = \sum_{(h,r,t) \in N_h} \text{TriA}(h, r, t) \cdot e_t, \quad (3)$$

其中 $\text{TriA}(h, r, t)$ 是根据注意力机制计算邻居节点的权重系数. 由于在嵌入层使用 TransR 方法对节点嵌入,通过关系平移的方式对三元组依次学习,可能会导致在同一个投影空间中基于翻译表示方法对预测节点的 \tilde{t}_r 与 t_r 值之间存在偏差. 为了避免偏差对模型的影响,在此考虑提出基于偏差的注意力权重分配函数 $\text{TriA}(h, r, t)$ 来区分邻居节点对 h 的重要度. 假设 $(h_r + e_r)$ 在关系空间中近似于 \tilde{t}_r , 对 \tilde{t}_r 和 t_r 进行按位乘操作以获取二者的偏差,如下式所示:

$$\text{TriA}(h, r, t) = W_1^T \text{relu}[W_2[(h_r + e_r) \otimes t_r]]. \quad (4)$$

其中: $W_1 \in \mathbf{R}^n$ 是可训练的注意力参数向量; $W_2 \in \mathbf{R}^{n \times n}$ 是可训练权重矩阵参数; relu 为非线性激活函数; \otimes 表示按元素相乘, 通过点乘操作. 节点 \tilde{t}_r 和 t_r 某些共有的特征信号会被增强, 不关联的特征信号(如预测时噪声带来的偏差)则会被弱化. W_2 是将输入特征信号投影到高维输出特征空间的参数化线性变换矩阵, 对放大后的信号特征经过非线性的激活函数的映射, 使信号在神经网络传播时更加逼近真实的情况; W_1 将计算出的信号表示投影到输出注意力权重向量. 使用 softmax 函数对所有与 h 相连的三元组 $\text{TriA}(h, r, t)$ 进行系数归一化, 有

$$\text{TriA}(h, r, t) = \frac{\exp(\text{TriA}(h, r, t))}{\sum_{(h, r', t') \in N_h} \exp(\text{TriA}(h, r', t'))}. \quad (5)$$

此时, $\text{TriA}(h, r, t)$ 计算出来的得分就是节点邻居的不同权重系数, 得分越高, 其邻居节点应该被给予更多的关注, 以便更好地捕获协同信号.

在获得 e_{N_h} 的表示后, 采用多通道融合函数从不同方面聚合实体 h 及其邻居信息, 对节点自身和邻居节点进行累加操作, 使节点特征信号进行整体增强. 为了避免融合后 e_{N_h} 和 e_h 的相似信号特征被弱化, 对其进行按位乘操作, 减少噪声对信号融合时产生的影响, 即

$$f_{\text{Agg}}(e_h, e_{N_h}) = \text{Prelu}[W_g(e_h + e_{N_h}) + W_b(e_h \otimes e_{N_h})]. \quad (6)$$

其中: Prelu 为非线性激活函数, 给所有负值赋予了非零斜率, 且负值部分的斜率根据训练数据而定, 而非预先定义, 最大程度地传播数据信息; \otimes 表示按元素相乘; $W_g, W_b \in \mathbf{R}^{n \times n}$ 为可训练权重矩阵, W_g 和 W_b 分别为特征融合后的信号投影到高维的输出空间的参数化线性变化矩阵. 非线性激活函数

$$\text{Prelu}(x_i) = \begin{cases} x_i, & x_i > 0; \\ a_i x_i, & x_i \leq 0; \end{cases} \quad (7)$$

其中 i 表示不同的通道数.

2) 多层结构.

为了探索节点的高阶连通性, 对 BGANR 模块进行堆叠更多的传播层, 收集来自高阶节点传播信息. 类似地, 对于第 l 层节点 i 的嵌入可表示为

$$e_h^{(l)} = f_{\text{Agg}}(e_h^{(l-1)}, e_{N_h}^{(l-1)}), \quad l = 1, 2, 3. \quad (8)$$

其中: $e_h^{(l-1)}$ 是上一层节点 i 的表示, 包含了其 $(l-1)$ 层邻居节点的信息, 当 $l=1$ 时, $e_h^{(0)}$ 初始化为 e_h , 其中邻居节点聚合如下式所示:

$$e_{N_h}^{(l-1)} = \sum_{(h, r, t) \in N_h} \text{TriA}(h, r, t) \cdot e_t^{(l-1)}. \quad (9)$$

2.4 预测和优化

经过 l 层的传播, 获得了用户节点 u 的多个表示, 即 $\{e_u^{(1)}, e_u^{(2)}, \dots, e_u^{(l)}\}$ 集合. 类似地, 对于项目节点 i , 得到 $\{e_i^{(1)}, e_i^{(2)}, \dots, e_i^{(l)}\}$ 集合. 由于在不同传播层收集的邻居节点信息不同, 采用级联操作将每层节点表示整合为单一向量, 一方面可丰富初始化嵌入节点, 另一方面可通过调整参数 l 来控制传播强度. 最后对用户和项目进行评分预测, 有

$$y_p(u, i) = [e_u^{(0)} \parallel \dots \parallel e_u^{(l)}]^T [e_i^{(0)} \parallel \dots \parallel e_i^{(l)}], \quad (10)$$

其中 \parallel 表示级联操作. 为了优化 BGANR 模型, 对协同知识图采用贝叶斯个性化排名算法 (BPR) 进行优化^[7], 希望正实例的得分比负实例的得分高. 损失函数为

$$\text{loss}_{\text{BGANR}} = \text{loss}_k + \text{loss}_c + a \|\Theta\|_2^2, \quad (11)$$

其中 Θ 是模型的参数集合, 并采用 L_2 正则化防止过拟合.

2.5 时间复杂度分析

BGANR 算法时间复杂度主要来自于知识图谱嵌入表示层, 如式 (2) 所示, TransR 模型的时间复杂度为 $O(|G|d^2)$; 对于 BGANR 层中的注意力表示部分, 从第 $l-1$ 层到第 l 层的时间复杂度为 $O(|G|d_{l-1}d_l + |G|)$; 最终的预测层, 只是在第 l 层进行内积运算, 其时间复杂度为 $O(|G|d_l)$, 因此总的时间复杂度为 $O(|G|d^2 + \sum_{l=1}^L (|G|d_{l-1}d_l + |G| + |G|d_l))$. 本文提出的 BGANR 算法的时间复杂度, 与经典的文献 [7] 相比增加了计算偏差 $\sum_{l=1}^L (|G|)$. 以 Amazon-book 数据集为例, 在本文的运算平台上测试所花费的实际时间约为 680 s, 800 s, 830 s, 930 s, 870 s, 890 s. BGANR 相比训练时间稍有增长, 但从如下实验结果来看, 本文提出的 BGANR 算法能够提高模型的健壮性.

3 实验分析

实验编程语言采用 Python, 服务器配置为 Precision 7920 Tower (12 C 2.3 GHz)/128 G 内存.

主要解决以下 3 个问题:

RQ1: BGANR 算法与经典算法在推荐性能方面进行比较.

RQ2: 通过消融实验分析本文所提出的算法中不同模块对推荐结果的影响.

RQ3: 在本文提出的算法与经典的算法比较过程中给出可解释性.

3.1 数据集

为了评估本文所提出BGANR模型的性能,选用Amazon-book^[22]、Last-FM^[7]以及Yelp2018^[10]3个公开数据集作为基准数据集. 由于Amazon-book数据集密度为0.048%, Last-FM数据集密度为0.267%, Yelp2018数据集密度为0.466%, 进而对其进行相应预处理操作以保证数据集的质量. 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组成. 用户项目交互信息由用户数、项目数及交互总条数构成, 项目属性信息由实体数、关系以及三元组个数构成. 值得注意的是, 对于可映射的实体, 与通过标签匹配的方式将项目映射到Freebase中, 以获得结构化的数据.

表1 数据集描述

data category	information	Amazon-book	Last-FM	Yelp2018
User-Item Interaction	#User	70 679	23 566	45 919
	#Item	24 915	48 123	45 538
	#Interactions	847 733	3 034 796	1 185 068
	Density	4.81e-04	2.67e-03	4.66e-03
Attribute Graph	#Entities	88 572	58 266	90 961
	#Relations	39	9	42
	#Triplets	2 557 746	464 567	1 853 704

3.2 实验设置

3.2.1 基准算法

BPRMF^[23]: 针对用户的隐式反馈信息解决推荐成对排序问题.

NFM^[24]: 结合深度神经网络和因子分解机用于建模更高阶特征之间的关系.

CKE^[6]: 考虑节点和关系的异构性来提取项目的结构特征.

CFKG^[25]: 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^[7]: 提出协同知识图谱中显式建模高阶关系信息, 以提高推荐精度.

3.2.2 评价指标

推荐系统的优劣需从多个角度去衡量, 针对用户个性化推荐列表的Top-N推荐, 包括准确率(Precision)、召回率(Recall)和命中率(Hit)等评价指标, 使用归一化折损累计增益(NDCG)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.

3.3 实验性能分析

3.3.1 与经典算法比较(RQ1)

在本文中BGANR模型深度设置为3层, 为了保证每一层梯度大小相近使用Xavier初始化权重参数, Adam优化模型. 嵌入向量的大小设置为64, 批处理大小为1024, 学习率调整范围为 $\{10^{-3}, 10^{-2}, 10^{-1}\}$, α 表示为正则化参数系数, 初始化为0. 当开始训练模型时不进行正则化, 如果观察到存在过拟合现象, 则在 $\{10^{-5}, 10^{-4}, \dots, 10^{-1}\}$ 范围内微调. 丢包率可调整范围为 $\{0.0, 0.1, \dots, 0.5\}$, 对于本文提出 $f_{\text{Agg}}(\cdot)$ 中Prelu函数通道参数 a_i 初始化为0.5, a_i 会随着数据进行修正, 从而更好地融合数据信息. 基于Top-N推荐的不同指标比较可以发现K值为20或40, 其性能都是优于其他模型指标, 实验结果如表2所示.

以推荐序列长度等于20为例, 本文提出的BGANR与以上5个基准模型相比, 在4个指标上均取得最佳的效果, 其次为KGAT模型. 基于Top-N推荐评测, BGANR在两个数据集上的评测较KGAT有明

表2 算法指标对比

datasets	method	Recall@20	Recall@40	NDCG@20	NDCG@40	Precision@20	Precision@40	Hit@20	Hit@40
Amazon-book	NFM	0.1272	0.1814	0.0871	0.1065	0.0133	0.0098	0.2217	0.3064
	BPRMF	0.1314	0.1917	0.0889	0.1104	0.0138	0.0105	0.2308	0.3251
	CKE	0.1322	0.1951	0.0891	0.1113	0.0139	0.0107	0.2324	0.3300
	CFKG	0.1103	0.1645	0.0744	0.0937	0.0118	0.0091	0.1965	0.2823
	KGAT	0.1490	0.2110	0.1011	0.1230	0.0155	0.0114	0.2557	0.3498
	BGANR	0.1519	0.2141	0.1031	0.1251	0.0158	0.0116	0.2599	0.3547
Last-FM	NFM	0.0756	0.1049	0.1132	0.1439	0.0288	0.0225	0.3106	0.4192
	BPRMF	0.0729	0.1028	0.1177	0.1489	0.0305	0.0238	0.3236	0.4292
	CKE	0.0731	0.1029	0.1180	0.1493	0.0306	0.0236	0.3235	0.4305
	CFKG	0.0678	0.0959	0.1083	0.1374	0.0277	0.0216	0.2958	0.3962
	KGAT	0.0869	0.1144	0.1321	0.1657	0.0330	0.0262	0.3540	0.4639
	BGANR	0.0876	0.1199	0.1341	0.1672	0.0357	0.0276	0.3571	0.4660
Yelp2018	NFM	0.0672	0.0952	0.0813	0.1382	0.0236	0.0355	0.2590	0.2860
	BPRMF	0.0669	0.1020	0.0802	0.1324	0.0359	0.0478	0.2730	0.3110
	CKE	0.0653	0.1000	0.0803	0.1369	0.0324	0.0447	0.2770	0.3190
	CFKG	0.0542	0.0893	0.6270	0.1254	0.0263	0.0386	0.2580	0.3130
	KGAT	0.0707	0.1060	0.0892	0.1420	0.0283	0.0466	0.2870	0.3450
	BGANR	0.0711	0.1090	0.0893	0.1442	0.0286	0.0473	0.2870	0.3440

显的提升,其中在 Amazon-book 数据集中召回率、精确率以及命中率分别提升了 1.94%、1.87%、1.67%,在 Last-FM 数据集中召回率、精确率以及命中率分别提升了 0.70%、8.21%、0.88%,从而验证了 BGANR 的提出很有意义和价值.从衡量和评价搜索结果进行比较,BGANR 在两个数据集上较最优模型的 NDCG 值分别提高了 2.01% 和 1.45%.BGANR 模型性能最优,得益于使用注意力权重分配函数能够区分三元组在传播过程中差异化的贡献,使头尾节点更准确地投影到同一个关系空间,减少节点嵌入向量偏差的影响.在信息聚合阶段,利用多通道融合函数使得模型能够捕获基于特征交互的丰富语义,从而产生高质量的个性化推荐.

Recall@K 和 Precision@K 分别反映了算法在基于 Top-N 推荐成功的比例和推荐的精确率,为了验证 K 值对于模型的影响,在两个数据集上进行验证,由图 2~图 5 可知,BGANR 模型推荐效果最优,其次是 KGAT,CFKG 性能稍弱,这表明本文所提出的算法能够有效得到用户和项目的嵌入表示,给用户提供更加准确的推荐序列.虽然 KGAT 模型和 CFKG 模型在一定程度上都属于基于知识图谱的推荐算法,但 KGAT 模型使用的是协同知识图谱,且考虑了用户和项目之间的高阶交互,以增强其节点的表征;而 CFKG 模型基于用户物品交互图,其关系较为单一,导致推荐性能不佳,同时也验证了沿着知识图谱中的关系来探索用户的潜在偏好是有效的.总体来说,BPRMF 模型预测性能比 NFM 模型性能要好,一个对用户物品矩阵进行分解,学习特征之间的交互隐藏关系,一个探索多个特征之间的交互关系,说明在某种特定情况下矩阵分解对项目关系的探索优于因子分解机模型.但是 BPRMF 和 CKE 模型推荐性能非常接近,在两个数据集上的 Precision 指标中更为明显,CKE 模型是基于物品关系图搭建的模型,结合物品的文本视觉特征以及用户物品交互矩阵构成完整的物品表征,而 BPRMF 模型是通过构造三元组的方式进行训练,二者不属于同一类别的推荐模型,导致推荐性能相近

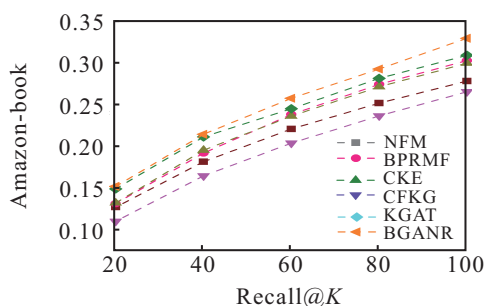


图 2 Amazon-book 数据集上不同算法召回率对比

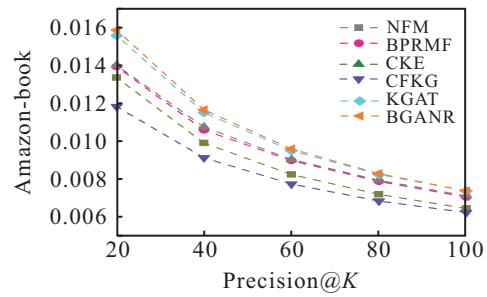


图 3 Amazon-book 数据集上不同算法准确率对比

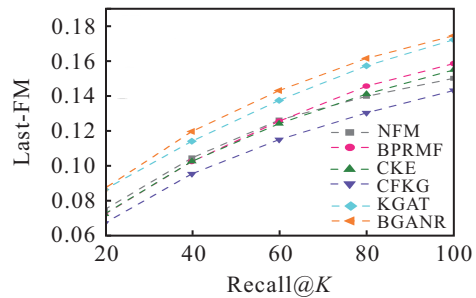


图 4 Last-FM 数据集上不同算法召回率对比

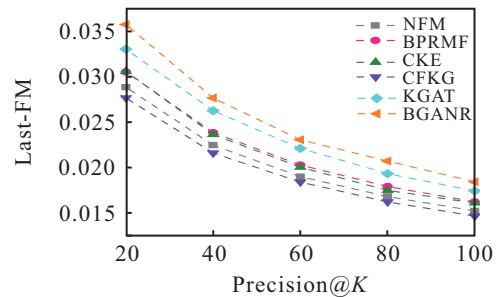


图 5 Last-FM 数据集上不同算法准确率对比

的原因可能是 CKE 模型在一定程度上忽略了图中信息的连通方式.

BGANR 在 100 次迭代的训练过程中,在两个数据集上的各个指标变化如图 6 和图 7 所示,本文选取 KGAT 模型作为参照.在第 100 次迭代时,BGANR 在两个数据集上的性能达到最佳;而 KGAT 在 Amazon-book 数据集上第 74 次迭代达到最佳性能,在 Last-FM 数据集上第 100 次迭代达到最佳.虽然二者都是借助知识图谱将物品属性信息注入节点表示向量,但 KGAT 模型未考虑实体节点之间的偏差,其在 Last-FM 数据集上的 Hit 指标对比更为明显.在嵌入传播

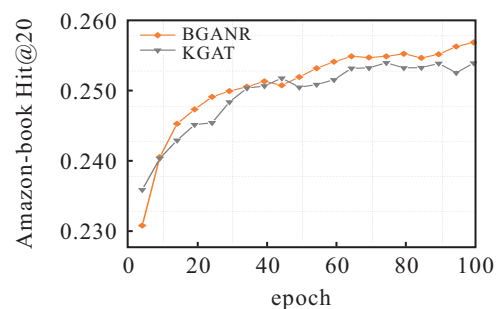


图 6 Amazon-book 数据集上迭代 100 次命中率指标变化

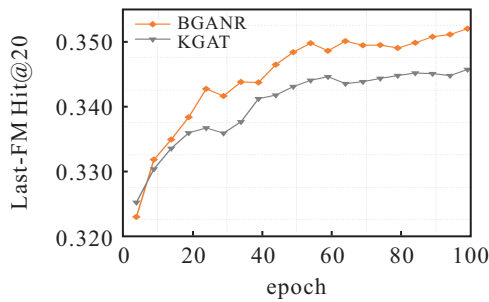


图7 Last-FM数据集上迭代100次命中率指标变化

融合信息阶段, 节点递归地获取相邻节点的信息并进行整合, 减少噪声对推荐性能的影响.

3.3.2 消融实验(RQ2)

为了验证本文提出的基于偏差的注意力算法和多通道融合函数在模型中的性能, 将BGANR与两个变体进行比较.

BGANR/A: 该模型不考虑节点嵌入时偏差对注意力得分的影响, 让节点的嵌入向量与关系向量经过非线性激活函数映射后, 再与邻居节点进行信息的融合得到最终的变体模型.

BGANR/P: 该模型在信息整合阶段不使用多通道融合函数, 假设在单一通道下节点与邻居节点间信息进行整合.

BGANR模型及其变体的NDCG@20和NDCG@100在不同数据集上的指标变化如表3所示. 由表3可知: 1) 多通道融合函数在一定程度上能够全方位地整合节点间信息. BGANR/P在Amazon-book数据集上的NDCG指标有明显的性能损失, 在Last-FM数据集上NDCG@20的指标性能不佳, 表明了信息在传播过程中使用多通道融合函数的重要性, 可防止部分信息在传播时的损耗. 2) 节点间偏差的注意力得分对BGANR也很重要, BGANR/A显示在两个数据集上的NDCG指标有中等程度的损失, 这并不意味着在BGANR模型中处理偏差没有意义. 若仅仅使用翻译模型预测尾节点, 则可能造成节点向量的泛化.

表3 BGANR变体模型指标对比

model	Amazon-book		Last-FM	
	NDCG@20	NDCG@100	NDCG@20	NDCG@100
BGANR/A	0.1026	0.1567	0.1335	0.2188
BGANR/P	0.1013	0.1557	0.1330	0.2197
BGANR	0.1031	0.1573	0.1341	0.2202

3.3.3 可解释性(RQ3)

本文基于用户物品协同知识图进行推荐, 会更加关注实体之间存在的关联关系, 采用基于翻译的方法对协同知识图中的实体和关系进行低维表示, 因为该嵌入更适用于大规模的网络架构, 并使模型在最大程

度上保持图结构的完整性. 在对注意力机制进行计算的过程中, 更加关注对实体邻居进行细粒度筛选操作, 以此避免在三元组翻译过程中预测尾实体和真实的邻居节点之间存在偏差, 会干扰对注意力机制的计算, 例如在图8中某用户听过 bad romance 这首歌, 已知歌曲在 music show 中被演唱过, 类型为 pop music, 旋律是 happy. 当实体对其邻居节点进行注意力权重计算时, bad romance(头节点)加类型(关系)未必能够准确预测为 pop music(尾节点). 为了避免使用翻译表示的方法对 bad romance 预测的尾节点和 pop music 节点存在的偏差对注意力权重产生的影响, 对其进行偏差处理操作, 即对预测的尾节点和真实的尾节点相同的信号进行增强, 不相关的信号进行弱化. 若不对偏差进行处理, 直接使用预测的尾节点计算实体的注意力权重, 则会带来一定程度的损失. 上述实验结果验证了本文提出的BGANR的有效性.

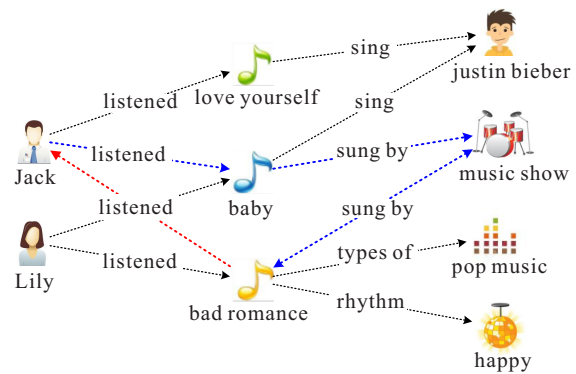


图8 音乐推荐系统协同知识图

4 结论

个性化推荐系统可以有效地提高推荐的准确率, 但随着研究的深入, 一些学者开始关注如何在用户项目交互图中加入辅助信息以丰富用户向量的表征, 然而基于图神经网络的嵌入传播往往忽略了节点之间对目标节点的贡献度不同. 因此, 本文提出了一种基于偏差的图注意力神经网络的推荐算法, 在节点信息嵌入传播过程中, 基于偏差注意力机制为邻居节点分配不同权重系数, 使节点在嵌入时能够关注到重要信息, 从而更能反映真实的情况. 通过多层网络嵌入传播捕获节点之间的高阶连通性, 并构造多通道融合函数对获取到的节点信息进行整合, 避免数据信息在嵌入时丢失. 最后在3个真实数据集上验证了BGANR模型的合理性和有效性.

参考文献(References)

[1] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
 [2] Covington P, Adams J, Sargin E. Deep neural networks for

- YouTube recommendations[C]. Proceedings of the 10th ACM Conference on Recommender Systems. New York, 2016: 191-198.
- [3] He X, Liao L, Zhang H, et al. Neural collaborative filtering[C]. Proceedings of the 26th International Conference on World Wide Web. Perth, 2017: 173-182.
- [4] Ying R, He R N, Chen K F, et al. Graph convolutional neural networks for web-scale recommender systems[C]. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, 2018: 974-983.
- [5] Wang Q, Mao Z D, Wang B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [6] Zhang F Z, Yuan N J, Lian D F, et al. Collaborative knowledge base embedding for recommender systems[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, 2016: 353-362.
- [7] Wang X, He X N, Cao Y X, et al. KGAT: Knowledge graph attention network for recommendation[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, 2019: 950-958.
- [8] Song W P, Xiao Z P, Wang Y F, et al. Session-based social recommendation via dynamic graph attention networks[C]. Proceedings of the 12th ACM International Conference on Web Search and Data Mining. Melbourne, 2019: 555-563.
- [9] Wang S, Sun L, Fan W, et al. An automated CNN recommendation system for image classification tasks[C]. 2017 IEEE International Conference on Multimedia and Expo (ICME). Hong Kong, 2017: 283-288.
- [10] He X N, Deng K, Wang X, et al. LightGCN: simplifying and powering graph convolution network for recommendation[C]. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, 2020: 639-648.
- [11] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[C]. International Conference on Learning Representations. Vancouver, 2017: 1-12.
- [12] Sun S M, Tang Y H, Dai Z M, et al. Self-attention network for session-based recommendation with streaming data input[J]. IEEE Access, 2019, 7: 110499-110509.
- [13] Ehrlinger L, Wöβ W. Towards a definition of knowledge graphs[J]. SEMANTiCS(Posters, Demos, Success), 2016, 48: 1-4.
- [14] Shi C, Hu B B, Zhao W X, et al. Heterogeneous information network embedding for recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31(2): 357-370.
- [15] Chen Q, Gu H, Yi L, et al. SocialTrans: A deep sequential model with social information for web-scale recommendation systems[J/OL]. (2020-05-09)[2021-04-06]. <https://arxiv.org/pdf/2005.04361v1.pdf>.
- [16] Xie R, Liu Z, Jia J, et al. Representation learning of knowledge graphs with entity descriptions[C]. Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016: 2659-2665.
- [17] Ma J W, Wen J H, Zhong M Y, et al. MMM: multi-source multi-net micro-video recommendation with clustered hidden item representation learning[J]. Data Science and Engineering, 2019, 4(3): 240-253.
- [18] Xie R, Liu Z, Sun M. Representation learning of knowledge graphs with hierarchical types[C]. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, 2016: 2965-2971.
- [19] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multirelational data[C]. Advances in Neural Information Processing Systems. Lake Tahoe, 2013: 2787-2795.
- [20] Zhang Z W, Cao L, Chen X L, et al. Representation learning of knowledge graphs with entity attributes[J]. IEEE Access, 2020, 8: 7435-7441.
- [21] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]. Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin: AAAI, 2015: 2181-2187.
- [22] Wang X, He X N, Wang M, et al. Neural graph collaborative filtering[C]. Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris, 2019: 165-174.
- [23] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[C]. The 25th Conference on Uncertainty in Artificial Intelligence. Montrea, 2009: 452-461.
- [24] He X N, Chua T S. Neural factorization machines for sparse predictive analytics[C]. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku Tokyo, 2017: 355-364.
- [25] Ai Q Y, Azizi V, Chen X, et al. Learning heterogeneous knowledge base embeddings for explainable recommendation[J]. Algorithms, 2018, 11(9): 137.

作者简介

王建芳(1979—),男,副教授,博士,从事推荐系统、人工智能等研究, E-mail: wangjianfang@hpu.edu.cn;

文茜琳(1997—),女,硕士生,从事推荐系统、人工智能的研究, E-mail: wenxilin7@gmail.com;

杨许(1997—),男,硕士生,从事推荐系统、人工智能的研究, E-mail: 212009020042@home.hpu.edu.cn;

张秋玲(1980—),女,讲师,从事数据挖掘和推荐算法的研究, E-mail: 309867230@qq.com.

(责任编辑:孙艺红)