

基于知识图残差注意力网络的推荐方法

范洪玉, 张永库, 孟祥福

引用本文

范洪玉, 张永库, 孟祥福. [基于知识图残差注意力网络的推荐方法](#)[J]. 计算机科学, 2023, 50(11A): 220900180-7.

FAN Hongyu, ZHANG Yongku, MENG Xiangfu. [Recommendation Method Based on Knowledge Graph Residual Attention Networks](#) [J]. Computer Science, 2023, 50(11A): 220900180-7.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[一种基于Meta-learning改进的特征交互算法](#)

Improved Feature Interaction Algorithm Based on Meta-learning

计算机科学, 2023, 50(11A): 230100087-8. <https://doi.org/10.11896/jsjcx.230100087>

[基于课程学习和图嵌入的协同推荐](#)

Collaborative Recommendation Based on Curriculum Learning and Graph Embedding

计算机科学, 2023, 50(11A): 221100030-8. <https://doi.org/10.11896/jsjcx.221100030>

[一种基于时效近邻可信选取策略的协同过滤推荐方法](#)

Time-effective Nearest Neighbor Trusted Selection Strategy Based Collaborative Filtering Recommendation Method

计算机科学, 2023, 50(11A): 220800199-11. <https://doi.org/10.11896/jsjcx.220800199>

[基于动态负采样的图卷积协同过滤推荐模型](#)

Dynamic Negative Sampling for Graph Convolution Network Based Collaborative Filtering Recommendation Model

计算机科学, 2023, 50(11A): 230200149-7. <https://doi.org/10.11896/jsjcx.230200149>

[融合物品关系的图神经网络推荐算法](#)

Graph Neural Network Recommendation Algorithm Based on Item Relations

计算机科学, 2023, 50(11A): 230100019-9. <https://doi.org/10.11896/jsjcx.230100019>

基于知识图残差注意力网络的推荐方法

范洪玉 张永库 孟祥福

辽宁工程技术大学电子与信息工程学院 辽宁 葫芦岛 125105

(fan_hy1997@163.com)

摘要 随着互联网的高速发展,推荐系统已成为缓解信息过载的重要手段之一。当前的推荐方法主要采用深度学习模型挖掘用户对项目的兴趣度,但是目前使用图神经网络的推荐方法无法有效表征用户和项目之间的交互行为,并且网络层数的增加会产生梯度消失问题。因此,文中提出了一种新的模型,基于正交变换和图上下文的知识图嵌入方式并融合残差网络和注意力机制的模型。首先,通过嵌入节点的邻居属性来表征用户与项目的交互行为,然后通过图神经网络和残差网络分析用户项目交互行为,最后利用注意力机制区分不同的邻域,提高推荐的准确性。通过在 Alibaba-fashion 和 Last-FM 两个真实数据集上进行实验,结果表明所提方法能够显著提升推荐效果。

关键词: 推荐系统; 图神经网络; 知识图谱; 协同过滤; 嵌入传播

中图法分类号 TP311

Recommendation Method Based on Knowledge Graph Residual Attention Networks

FAN Hongyu, ZHANG Yongku and MENG Xiangfu

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract With the rapid development of the Internet today, recommendation system has become an important means to relieve the information overload. Current recommendation methods mainly use deep learning model to mine users' interests in the project. However, the current recommendation methods using graph neural networks cannot effectively represent the interaction behaviors between users and items well, and the increase in the number of network layers will cause the problem of gradient disappearance. Therefore, this paper proposes a model that combines the GC-OTE knowledge graph embedding approach with residual networks and attention mechanisms. First, the interaction information of users or items is represented by embedding the neighbor attributes of nodes, then user-item interactions are analyzed by graph neural and residual networks, and finally, attention mechanisms are used to distinguish different neighborhoods. Experiments on two real-world datasets Alibaba-fashion and Last-FM demonstrate that the proposed method can significantly improve the recommendation performance.

Keywords Recommendation system, Graph neural network, Knowledge graph, Collaborative filtering, Embedding propagation

1 引言

近年来,随着移动网络的高速发展,物品数量日益增多、种类繁多。人们在做选择时往往需要过滤掉许多无用信息,因此几乎每一个提供内容给用户的网站都需要推荐系统,如商品推荐、歌曲推荐、电影推荐、兴趣点推荐等。推荐系统在网络中应用十分普遍,已经成为了一个热点的研究方向。现有的推荐系统中最常见的是基于协同过滤(Collaborative Filtering, CF)^[1]的推荐,其基本思想是利用用户-用户之间的相似性关系,挖掘用户与项目的交互信息,如浏览、购买等,为目标用户推荐喜爱的项目。基于CF的推荐方法利用用户-项目交互,根据用户-用户或项目-项目的相似性向用户提供建议^[2-5]。文献[6]提出了单类协同过滤模型,主要通过用户过去的行为来估计用户感知时尚的个性化排名功能。文献[7]提出了一种基于协同过滤去噪自编码器算法,利用向输入数据中增添用户的属性来提高推荐效果。协同过滤包括基于邻居的协同过滤、基于矩阵分解的协同过滤、基于神经网络的

协同过滤和最近常见的基于图神经网络的协同过滤。虽然这些方法具有高效性和通用性,但对用户-项目的历史交互的依赖性仍然很强,当交互数据稀疏时,性能会降低;并且,协同过滤方法无法对辅助信息建模^[8-9],例如社交网络^[10]、用户/项目属性^[11]、上下文信息^[12]等。其次,多数的推荐方法都独立的考虑每个项目的相关性,却忽略了推荐结果的多样性。

为了解决上述问题,本文提出一种使用知识图嵌入的推荐方法,首先建立用户-项目二部图,将用户和项目表示为一个直观的交互;同时构建知识图,其中包含了项目的属性信息。然后,耦合二部图与知识图,将耦合图经过图神经网络进行推荐。最近的一些研究使用深度学习模型进行推荐, Wang等^[13]提出了基于知识图的推荐注意力网络(Knowledge Graph Attention Network, KGAT)。上述方法尽管取得了较好的推荐效果,但其中的 TransR^[14]嵌入方式仍不能很好地解决用户和项目的关系问题,侧重于对单个三元组的建模,忽略了知识图结构信息,即有意义的图上下文。因此本文提出了使用 GC-OTE 的嵌入方法^[15],可以更好地解决用户和项目复杂的

基金项目:辽宁省教育厅科学研究项目(LJKZ0355)

This work was supported by the Liaoning Provincial Department of Education Scientific Research Project(LJKZ0355).

通信作者:孟祥福(marxi@126.com)

多对一(N-1)、一对多(1-N)、多对多(N-N)的问题,从而提升推荐性能。

本文的主要贡献包括:

- 1) 提高了使用 GC-OTE 的知识图谱嵌入方式解决推荐中的数据稀疏和冷启动问题。
- 2) 提出了融合残差注意力的方法,可以使训练过程中深度网络出现的梯度消失和梯度爆炸问题得到缓解。
- 3) 提出了使用注意力机制区分不同的邻域,使得推荐结果具有更好的可解释性。
- 4) 在 Alibaba-fashion 和 Last-FM 这两个数据集上进行测试,验证了所提方法的有效性和优越性。

2 相关工作

2.1 基于深度学习的推荐

随着近年来深度学习的快速发展,很多种方法提高了推荐系统的丰富性和准确性,对于冷启动和数据稀疏等问题,通常使用其本身提取特征的能力来缓解。文献[16]提出了 Wide&Deep 模型,该模型分为 3 部分,分别为 Wide 线性模型、Deep 非线性模型、Wide 部分输出和 Deep 部分输出联合预测部分,可以缓解数据稀疏的问题并且提高推荐的准确性。但仍有一些方面有待提高,比如 Wide 部分对人工特征工程的依赖性较强。文献[17]提出了 DeepFM 模型,结合了 Deep 和 FM 模型,其中的多层感知机对高阶属性交互建模,改进了 Wide&Deep 模型需要对低维特征做特征工程的不足,并且 Wide 和 Deep 部分共享特征向量。文献[18]提出了 NCF 模型,这种模型的建模灵感来源于多层感知机对项目以及客户之间的非线性关系,是神经网络附加过滤推荐的一种算法,是将低阶属性和高阶属性结合起来进行内积操作。但其也存在一些缺点,比如在嵌入项目和用户时没有关联性。文献[19]提出了 NFM 模型,它是 FM 与神经网络结合的尝试,使用非线性变换来提高模型的非线性表达能力并且能够学习到高阶的特征,该模型的结构更简单但是推荐性能更优。除了上述模型,其他的神经网络模型也可用于推荐。例如文献[17]和文献[20]都是基于神经网络的模型,文献[17]使用自编码和协同过滤结合的模型进行评分预测,文献[20]是基于 RNN 的序列协同过滤推荐算法,提供用户短期兴趣的推荐效果。

2.2 知识图谱

知识图谱属于异构图的分支,其中的节点对应头实体和

尾实体,以及知识谱图的边与尾实体和头实体联系。知识图谱中的语义关联关系可以帮助推荐系统挖掘出不同实体之间的潜在关联关系,通过潜在关系进而推出用户的关注点,提高推荐的准确性和多样性。并且,为解决用户-项目之间的数据稀疏问题,知识图谱充分利用其实体间的丰富关系来挖掘用户和用户之间的联系,项目和项目之间的联系,以增强推荐算法的效果。

基于知识图谱的推荐有 3 种方式,分别是基于嵌入的知识图谱推荐方法、基于路径的知识图谱推荐方法以及基于嵌入和基于路径相结合的知识图谱推荐方法。其中,基于嵌入的知识图谱推荐方法有:文献[21]提出的 CKE 模型,该模型中知识图谱的结构信息由 TransR 嵌入方法处理,以获得实体的结构化信息向量,结合与所产生的视图信息和文本信息向量,构成项目的潜在表示;文献[22]提出的 DKN 模型,该模型中知识图谱实体向量的一跳范围的上下文向量表示由 TransD^[23]嵌入方法处理,提取了自然属性和用户-项目的隐式关联关系,解决了新闻推荐的难题;文献[24]提出的 MKR 模型,该模型是一个端到端的推荐模型框架,由 3 个部分组成,分别是推荐部分、KGE 部分和交叉压缩单元。MKR 模型分别对推荐系统和知识图谱进行优化,提高推荐的灵活性和适应性。基于路径的知识图谱推荐方法有:文献[25]提出的 RKGE 模型,利用循环网络的技术自动学习并建模连接相同实体之间的路径语义,并且使用池化操作区分不同路径中用户对项目的偏好程度,最后将其嵌入到推荐中,挖掘用户对项目的最终偏好;文献[26]提出的 KPRN 模型,该模型是一个端到端的神经网络模型,利用实体和关系生成的路径表示向量来挖掘用户的偏好,提高推荐的可解释性。基于嵌入和基于路径相结合的知识图谱推荐方法有:文献[27]提出的 RippleNet 模型,该推荐模型结合知识图谱的辅助信息,利用消息传播将外层的项目作为用户的潜在偏好;文献[28]提出的 KTUP 模型,将基于翻译的用户偏好模型和知识图谱结合,共同学习用户、项目、实体和关系的向量表示,提高推荐结果的准确性和可解释性。

3 基于知识图谱残差注意网络的模型

本章主要对知识图谱残差注意力网络推荐算法模型 RK-GAT(Residual Knowledge Graph Attention Network)进行介绍,其结构如图 1 所示。

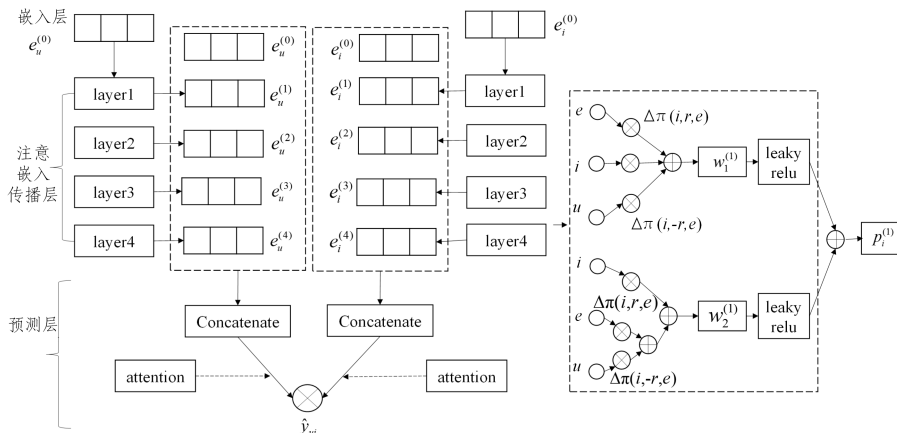


图 1 RKGAT 模型结构

Fig. 1 Structure of RKGAT model

该模型主要由3部分组成:1)嵌入层,是本文模型区别于其他神经网络的重要部分;2)注意力嵌入传播层,是实现推荐过程的核心部分;3)预测层,引入注意力机制对用户和项目加权区分。

3.1 嵌入层

知识图是一个拥有多种关系的有向图,知识图中的节点表示实体,边表示关系。学习低维向量的表示可以使用知识图嵌入模型,一个三元组 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 也可以用知识图来表示。多个三元组可以包含同一个实体,该实体作为连接这几个三元组的媒介。例如图2中三元组《百万美元宝贝》,导演,克林特·伊斯特伍德和三元组《百万美元宝贝》,演员,克林特·伊斯特伍德。《百万美元宝贝》为导演属性和演员属性提供了丰富的信息,从而能够挖掘出用户偏好,将属性信息传递给用户。

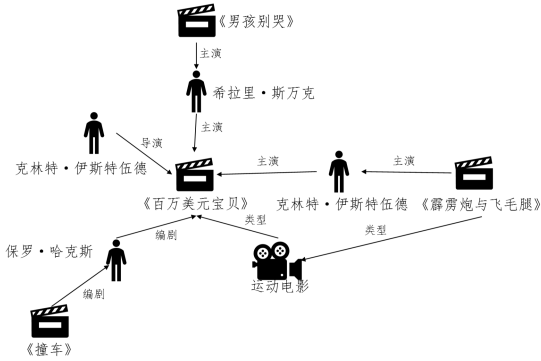


图2 一个知识图包含多个属性的例子

Fig. 2 Example of KG containing multiple attributes

知识图谱嵌入中使用GC-OTE的嵌入方法,GC-OTE嵌入方法使用正交变换将实体和项目之间的关系映射到高维空间,神经网络中使用正交变换的方法可使其更稳定、更有效。对于一个三元组 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$,其中 $\mathbf{h}, \mathbf{t} \in V$ (V 是图节点的集合), $\mathbf{r} \in R$, R 是图的边集合。三元组 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 对应的嵌入为 $\mathbf{e}_h, \mathbf{M}_r, \mathbf{e}_t$,其中 $\mathbf{e}_h, \mathbf{e}_t \in R_d$, d 是嵌入维数,头实体和尾实体的嵌入为 $\mathbf{e}_x, x = \{\mathbf{h}, \mathbf{t}\}$,对于 \mathbf{h} 和 \mathbf{r} 到 \mathbf{t} 的投影,计算式如下:

$$\mathbf{e}_t^*(i) = f_i(\mathbf{h}, \mathbf{r}) = \mathcal{O}(\mathbf{M}_r(i)) \mathbf{e}_h(i) \quad (1)$$

其中, \mathcal{O} 为施密特正交过程, $\mathcal{O}(\mathbf{M}_r(i))$ 是 $\mathbf{M}_r(i)$ 的正交矩阵。

对于 \mathbf{r} 和 \mathbf{t} 到 \mathbf{h} 的投影计算式如下:

$$\mathbf{e}_h^*(i) = \text{diag}(\exp(-S_r(i))) \mathcal{O}(\mathbf{M}_r(i))^T \mathbf{e}_t(i) \quad (2)$$

对应关系 \mathbf{r} 和头实体 \mathbf{h} 的预测值极端方法为:

$$\mathbf{e}_t^* = \text{diag}(\exp(S_r(i))) \mathcal{O}(\mathbf{M}_r(i)) \mathbf{e}_h(i) \quad (3)$$

3.2 注意嵌入传播层

注意嵌入传播层主要使用神经网络^[29]、图注意网络^[30]和残差网络^[31]。使用神经网络传播高阶连接,使用图注意网络的注意权重区分不同连接的重要程度,使用残差网络可以使得神经网络达到更深的层。一层的神经网络主要由3部分构成:信息传播、基于知识感知的注意力和信息聚合。

3.2.1 信息传播

对于一个实体 \mathbf{h} ,可以定义其三元组的集合计算式如下:

$$N_u = \{(\mathbf{h}, \mathbf{r}, \mathbf{t}) \mid (\mathbf{h}, \mathbf{r}, \mathbf{t}) \in G\} \quad (4)$$

其中, \mathbf{h} 表示头实体, \mathbf{r} 表示关系, \mathbf{t} 表示尾实体。

通过计算 \mathbf{h} 的自我网络的线性组合,表征实体 \mathbf{h} 的一阶

连通性计算式如下:

$$e_{N_u} = \sum_{(\mathbf{h}, \mathbf{r}, \mathbf{t}) \in N_u} \pi(\mathbf{h}, \mathbf{r}, \mathbf{t}) e_i \quad (5)$$

其中, $\pi(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 控制了每个边缘 $(\mathbf{h}, \mathbf{r}, \mathbf{t})$ 传播的衰减因子,表明从 \mathbf{t} 到 \mathbf{h} 传播的多少信息符合关系 \mathbf{r} 。

3.2.2 知识感知的注意力

得到用户和项目表示后,推荐系统通过用户点击的项目来增强其表示,这里采用平均池化,计算式如下:

$$\mathbf{e}_u^* = \mathbf{e}_u + \frac{1}{|N_u|} \sum_{t \in N_u} \mathbf{e}_t^* \quad (6)$$

其中, $N_u = \{t \mid (\mathbf{u}, i) \in O\}$ 和 \mathbf{e}_u 表示用于协同过滤的用户嵌入。但是平均池化没有考虑用户的个性化偏好,因此提出注意力机制来建模用户兴趣。

使用注意力机制为属性和用户分配注意分数 $f_{\text{att}}(\mathbf{h}, \mathbf{r}, \mathbf{t})$,并通过结合兴趣分数和交互项目生成用户表示。计算注意分数 $f_{\text{att}}(\mathbf{h}, \mathbf{r}, \mathbf{t})$,第一步是将不同的属性放置在相同的实体上,构造出 \mathbf{e}_n ,计算式如下:

$$\mathbf{e}_n = \mathbf{e}_u^{r_1} \parallel \mathbf{e}_u^{r_2} \parallel \dots \parallel \mathbf{e}_u^{r_m} \quad (7)$$

其中, $\mathbf{e}_u^{r_n} \in R^{d^m}$ 是 $\mathbf{e}_u \in R^D$ 的截断向量, \mathbf{e}_u 中 $\mathbf{e}_u^{r_n}$ 的开始索引是 $\sum_{p=1}^{m-1} d^p$,结束索引是 $\sum_{p=1}^m d^p$ 。

第二步,计算实体 \mathbf{e}_n 对实体 \mathbf{e}_i 的偏好得分,计算式如下:

$$f_{\text{att}}(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \sigma \left(\tau \frac{\frac{1}{|N_u|} \sum_{t \in N_u} \mathbf{e}_u^{r_1} \mathbf{e}_i^{r_m}}{\frac{1}{I} \sum_{t \in z} \mathbf{e}_u^{r_m} \mathbf{e}_i^{r_m}} \right) \quad (8)$$

其中, $\frac{1}{|N_u|} \sum_{t \in N_u} \mathbf{e}_u^{r_m} \mathbf{e}_i^{r_m}$ 表示用户对与属性 r_m 交互的项目的偏好程度, $\frac{1}{I} \sum_{t \in z} \mathbf{e}_u^{r_m} \mathbf{e}_i^{r_m}$ 表示对属性 r_m 中所有项的偏好程度。

最后,使用注意力机制, \mathbf{e}_u^* 表示的计算式如下:

$$\mathbf{e}_u^* = \mathbf{e}_u + \sum_{r_m \in R} \left(\frac{f_{\text{att}}(\mathbf{h}, \mathbf{r}, \mathbf{t})}{|N_u|} \sum_{t \in N_u} \mathbf{e}_t^{r_m} \right) \quad (9)$$

3.2.3 信息聚合

信息聚合,聚合了实体表示 \mathbf{e}_h 和自我网络 \mathbf{e}_{N_h} ,用以下3种类型的聚合器来实现。

1)GCN聚合器^[32]。两种表示可以用GCN聚合器相加,再对其进行非线性变换,计算式如下:

$$f_{\text{GCN}} = \text{LeakyReLU}(W(\mathbf{e}_u + \mathbf{e}_{N_u})) \quad (10)$$

其中,激活函数设置为LeakyReLU, $W \in R^{d \times d}$ 可以通过训练的权重矩阵来提取信息进行传播, d 是转换尺寸大小。

2)GraphSage聚合器^[33]。GraphSage聚合器连接两种表示,同样使用非线性变换,计算式如式(11)所示:

$$f_{\text{GraphSage}} = \text{LeakyReLU}(W(\mathbf{e}_u \parallel \mathbf{e}_{N_u})) \quad (11)$$

其中, \parallel 是连接运算。

3)Bi-Interaction聚合器^[34]。Bi-Interaction聚合器用来表示 \mathbf{e}_h 和 \mathbf{e}_{N_h} 的两种特征的交互情况,计算式如式(2)所示:

$$f_{\text{Bi-Interaction}} = \text{LeakyReLU}(W_1(\mathbf{e}_u \parallel \mathbf{e}_{N_u})) + \text{LeakyReLU}(W_2(\mathbf{e}_u \odot \mathbf{e}_{N_u})) \quad (12)$$

其中, $W_1, W_2 \in R^{d \times d}$ 可以通过训练的权重矩阵来提取信息进行传播, \odot 表示乘积。Bi-Interaction聚合器不同于GCN聚合器和GraphSage聚合器,它考虑了 \mathbf{e}_h 和 \mathbf{e}_{N_h} 之间的交互。

3.2.4 残差网络

残差网络是神经网络的另外一条通道,由一系列的

残差块组成,其中残差块定义如下:

$$x_{l+1} = h(x_l) + F(x_l, W_l) \quad (13)$$

由直接映射部分和残差部分两部分组成残差块。其中直接映射部分是 $h(x_l)$, 残差部分是 $F(x_l, W_l)$ 。

该残差结构用来解决“网络退化”问题,即当神经网络达到深层时,效果却不如浅层。通过残差结构可以将高层的网络传播到低层,缓解了网络层数加深导致的梯度消失的问题。

3.3 预测层

模型经过 K 层后,得到用户节点 u 的多个表示,即 $\{e_u^{(1)}, \dots, e_u^{(L)}\}$, 项目节点 i 的多个表示,即 $\{e_i^{(1)}, \dots, e_i^{(L)}\}$ 。将每一层的表示连接起来,计算式如下:

$$e_u^* = \sum_{L=0}^k \alpha_K e_u^{(K)} \quad (14)$$

$$e_i^* = \sum_{L=1}^k \alpha_K e_i^{(K)} \quad (15)$$

其中, $\alpha_K \geq 0$, 表示第 K 层嵌入的重要程度。

此处的注意力机制采用多层感知机实现,计算式如下:

$$f(p_i, q_i) = h^T \text{ReLU}(w(p_i \odot q_i) + b) \quad (16)$$

其中, $p \in R^{d \times 1}$ 是稀疏向量转换为密集向量的表示, $q^{n \times 1}$ 表示用户向量的初始表示, $q_j \in R^{d \times 1}$ 表示项目向量的嵌入表示, $j \in R_u^+$ 表示与用户有交互行为的项目集合, $w^{k \times d}$ 表示权重矩阵, $b^{d \times 1}$ 表示防止过拟合的偏置项。

然后对输出结果使用 softmax 函数进行归一化处理,计算式如下:

$$\alpha = \frac{\exp(f(p_i, q_j))}{[\sum \exp(f(p_i, q_j))]^\beta} \quad (17)$$

其中, $\beta \in [0, 1]$ 为平滑系数,防止对活跃的用户过度惩罚。

最后,对用户和项目表示进行内积,得到它们的预测得分,计算式如下:

$$\hat{y} = e_u^* \text{T} e_i^* \quad (18)$$

3.4 模型训练

在本文的模型中,首先依次对 L_{CF} 和 L_{KG} 采用 BPR^[35] 损失进行优化, BPR 损失假设观察到的交互,表明更多的用户偏好,应该比未观察到的交互分配更高的预测值,其计算式如下:

$$L_{CF} = \sum_{(u,i,j) \in O} -\ln \sigma(\hat{y}(u,i) - \hat{y}(u,i)) \quad (19)$$

其中, $O = \{(u,i,j) | (u,i) \in R^+, (u,j) \in R^-\}$ 表示训练集, R^+ 表示用户 u 与项目 j 之间有交互行为, R^- 表示用户 u 与项目 j 之前没有交互行为。

$$L_{KG} = \sum_{(h,r,t,t')} -\ln \sigma(g(h,r,t') - g(h,r,t)) \quad (20)$$

最后,将两个函数联合起来:

$$L_{KGGAT} = L_{CF} + L_{KG} + \lambda \| \mathfrak{R} \|_2^2 \quad (21)$$

其中, $\mathfrak{R} = \{E, W_l, \forall l \in R, W_l^1, W_l^2\}$, $\forall l \in \{1, 2, \dots, L\}$, E 是所有实体和关系嵌入表,在 \mathfrak{R} 上进行了由 λ 参数化的 L2 正则化,以防止过拟合。

使用 Adam^[36] 优化嵌入损失和预损失。用一批随机抽样 (h, r, t, t') , 来更新所有节点的嵌入。然后,随机抽样一批 (u, i, j) , 在 L 个传播步后检索它们的表示,最后利用预测损失的梯度更新模型参数。

4 实验及评价

本章首先介绍数据集和实验设置,然后介绍实验的对比算法,最后分析本文提出的模型和各个对比模型的效果。

4.1 数据集及实验设置

本文实验采用两个在不同的领域公开的数据集,分别是 Alibaba-fashion 和 Last-FM。这两个数据集具有不同的稀疏度。数据集均由两部分组成:用户-项目交互部分和知识图部分。为缓解数据的稀疏性,对数据集进行了简单的预处理,使签到次数少于 10 次的用户和项目被过滤掉。最终 Alibaba-fashion 数据集中包含了 114 737 个用户、30 040 个项目和 1 781 093 次交互, Last-FM 中包含了 23 566 个用户、48 123 个项目和 3 034 796 次交互。另外还要按照 ID 匹配的方式构建实体和关系的知识图。预处理之后的实验数据集的统计数据如表 1 所列。

表 1 实验数据信息统计
Table 1 Experimental data information statistics

		Alibaba-fashion	Last-FM
User-Item Interaction	Users	114 737	23 566
	Items	30 040	48 123
	Interactions	1 781 093	3 034 796
Knowledge Graph	Sparse/%	94.83	99.73
	Entities	59 156	58 266
	Relations	51	9
	Triplets	279 155	464 567

Alibaba-fashion: 是一个电子商务数据集,由用户-项目对应的知识图组成,包含用户点击服装的历史数据、每件衣服的不同类别属性等。

Last-FM: 是世界上最大的社会音乐平台,该数据集是从 2015 年 1 月到 2015 年 6 月的时间范围内收集而来的。

本文进行的实验采用 8:2 的比例划分训练集和测试集。RKGAT 模型使用 python 语言和 tensorflow 框架实现。模型的嵌入尺寸为 64, 采用 Adam 优化器进行优化,设置学习率为 0.0001, 一个批次数量为 1024, 偏置初始设置都为 0, 并采用 Xavier 初始化器对参数进行初始化,迭代次数为 100 时模型收敛,并设置早停机制。在 10 个训练轮数中,若提出的评价指标没有明显变化,则提前停止训练。

将 RKGAT 中 dropout 数值分别设置为 $\{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$, 实验测试表明(见图 3), dropout 取值为 0.5 时,推荐效果最优。

4.2 对比方法

将本文模型和以下最近比较流行的模型进行比较。

1) FM^[37]: 因子分解机模型,引入两个特征的二阶特征组合,考虑特征的交叉,使得每个特征都能学习到自己对应的嵌入,最后通过内积计算出权重。

2) NFM^[19]: 该方法通过特征交叉池化层,融合了 FM 模型和 DNN 模型,包括线性计算和交叉特征计算两部分,最后的结果是两部分加权的結果。

3) CKE^[21]: 该方法是结合知识图谱和协同过滤的方法,采用典型的知识图嵌入算法 TransR^[14] 对项目的表示进行规范化,并将其输入 MF 框架进行推荐。

4) CFKG^[38]: 该方法提出了一种新的翻译模型,使用知识表征学习对异构实体进行嵌入,将推荐任务转化成对三元组的合理性预测。

5) RippleNet^[26]: 该模型是基于嵌入的方法和基于路径的方法的结合,使用知识图传播用户的偏好属性,找到潜在联系,丰富用户表示后进行推荐。RippleNet 首先使用 TransE^[39] 嵌入

方式嵌入初始的知识图实体,自动发现用户潜在的兴趣。

6)KGAT^[40]:该方法将用户行为和项目知识编码为一个统一的知识图,以利用高阶连通性。KGAT 在一个整体图上应用了一种注意的邻域聚合机制,并引入了 TransR 来正则化表示。

4.3 评价标准

NDCG:一种基于排名情况的测试标准,项目分数的高低由排名的前后决定,排名靠前的项目得分更高,计算式如下:

$$NDCG@K = \sum_{i=1}^k \frac{2^{rel_i-1}}{\log_2(i+1)} \quad (22)$$

其中, rel_i 表明用户对第 i 个项目的评分, $\log_2(i+1)$ 是位置递减权重。

Recall:召回率反映了系统返回的项目是用户真正感兴趣的项目百分比,计算式如下:

$$recall@K = \frac{\sum |I \cap I^*|}{|I^*|} \quad (23)$$

其中, I 是训练数据集为用户作出的推荐集合, I^* 是测试数据集为用户作出的推荐集合。

4.4 实验结果分析

4.4.1 模型推荐效果对比

该实验的目的是将本文提出的基于知识图残差注意力网络的推荐方法与目前各种常见的推荐方法的推荐效果进行对比。分别在 Last-FM 和 Alibaba-fashion 两个数据集上的得到各个模型的推荐效果。表 2 列出了各个模型的对比结果。由实验结果可知,在两个数据集的两个评价指标上,本文的模型的表现都优于 3.2 节中的方法。其中在 Last-FM 上的推荐效果提升比在 Alibaba-fashion 上更好,主要原因是在数据集 Last-FM 上,平均每个用户进行过交互的数量多于 Alibaba-fashion 数据集。从模型的总体推荐效果来看,基于图神经网络的实验结果比基于路径的方法和基于嵌入的方法效果更好,原因是这 3 种推荐采用了不同的属性嵌入方法,基于图神经网络的推荐方法是将邻居的属性信息聚合到项目的节点中,来学习更丰富的表示。而另外两种方法都有一定的局限,它们都没有脱离知识图谱算法来建模属性信息。

表 2 本文模型和对比模型实验结果比较

Table 2 Comparison of experimental results between the proposed model and comparison models

	Last-FM		Alibaba-fashion	
	Recall@20	NDCG@20	Recall@20	NDCG@20
FM	0.0778	0.0768	0.0857	0.0713
NFM	0.0829	0.0810	0.1278	0.0735
CKE	0.0736	0.0805	0.1287	0.0674
CFKG	0.0723	0.0644	0.0692	0.0558
RippleNet	0.0842	0.0766	0.1121	0.0695
KGAT	0.0870	0.0845	0.1295	0.0816
RKGAT	0.1095	0.1544	0.1484	0.1173

4.4.2 不同聚合器对模型推荐效果的影响

该实验的目的是研究不同的聚合器对本文模型的推荐效果产生的影响,实验结果如表 3 所列。聚合器分别采用 GCN, GraphSage 和 Bi-Interaction。推荐结果表明聚合器 GCN 的表现优于聚合器 GraphSage 的效果,聚合器 Bi-Interaction 的表现优于聚合器 GCN 的效果。原因是聚合器 GraphSage 没有考虑实体表示 e_h 和自我网络 e_{N_h} 之间的交互。因此信息传播、信息聚合和特征的交互作用对推荐的

效果有积极作用。

表 3 聚合器测试

Table 3 Aggregator test

	Last-FM		Alibaba-fashion	
	Recall@20	NDCG@20	Recall@20	NDCG@20
GCN	0.1053	0.1385	0.1429	0.1087
GraphSage	0.1050	0.1378	0.1432	0.1068
Bi-Interaction	0.1061	0.1408	0.1464	0.1092

4.4.3 学习率对模型推荐效果的影响

该实验的目的是研究不同的学习率对本文模型的推荐效果产生的影响。图神经网络的收敛情况受模型学习率的影响,因此也会影响模型的推荐效果。本文模型中,将模型的学习率设置为超参数,手动进行调整来选择一个最优的学习率以提高训练模型的推荐效果。实验结果如图 3 所示,当学习率为 0.001 时,推荐的整体效果最优,当学习率为 0.05 时,推荐的整体效果最差。

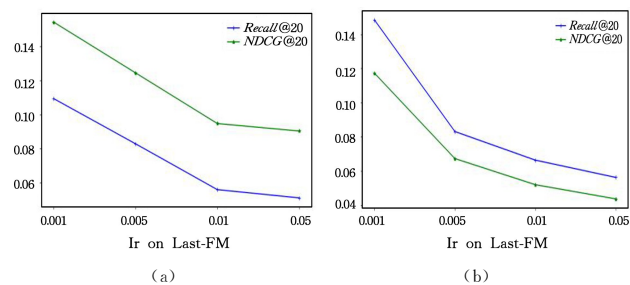
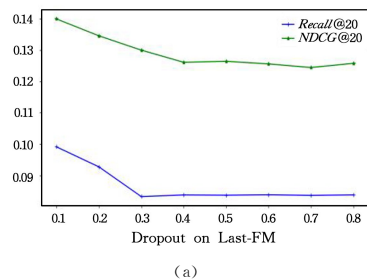


图 3 学习率测试

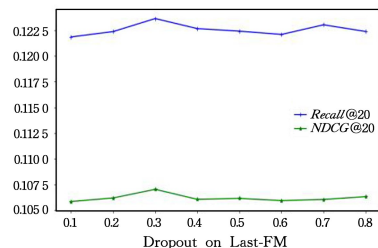
Fig. 3 Learning rate test

4.4.4 dropout 对模型推荐效果的影响

该实验的目的是研究 dropout 的不同取值对本文模型的推荐效果产生的影响。图神经网络模型训练时,dropout 的适当取值可以防止本文模型的过拟合。实验结果如图 4 所示。当 dropout 的值为 0.1 时,在 Last-FM 数据集上本文的推荐模型效果最好;当 dropout 的值为 0.3 时,在 Alibaba-fashion 数据集上本文的模型推荐效果最好。由于图 4(b)的实验结果相差很小,因此与图 4(a)设置不同的精确度,以更清晰地表明实验结果。



(a)



(b)

图 4 dropout 率测试

Fig. 4 dropout test

结束语 本文首先利用知识图谱和用户项目关系二部图,获得用户和项目的高阶向量表示;然后构建图神经网络,使用 GC-OTE 的嵌入方式将用户和项目的向量表示嵌入到网络模型中;最后通过注意力机制区分重要程度得到最终的预测得分。实验结果表明本文提出的方法能够提高 Recall 和 NDCG。

在未来的工作中,将从以下两个方面进行研究:1)在推荐系统中,社交关系如何影响推荐的结果? 2)如何基于长期行为序列建模用户兴趣的演变过程?

参 考 文 献

- [1] ZHANG Y, SONG W. A collaborative filtering recommendation algorithm based on item genre and rating similarity[C]// 2009 International Conference on Computational Intelligence and Natural Computing. IEEE, 2009, 2: 72-75.
- [2] SHAMBOUR Q, HUSSEIN A H, KHARMA Q, et al. Effective Hybrid Content-Based Collaborative Filtering Approach for Requirements Engineering [J]. *Comput. Syst. Sci. Eng.*, 2022, 40(1): 113-125.
- [3] ZHAO X, ZHANG W, WANG J. Interactive collaborative filtering[C]// Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. 2013: 1411-1420.
- [4] REZAIMEHR F, DADKHAH C. A survey of attack detection approaches in collaborative filtering recommender systems[J]. *Artificial Intelligence Review*, 2021, 54(3): 2011-2066.
- [5] JEONG B, LEE J, CHO H. Improving memory-based collaborative filtering via similarity updating and prediction modulation [J]. *Information Sciences*, 2010, 180(5): 602-612.
- [6] 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. 2016: 507-517.
- [7] WU Y, DUBOIS C, ZHENG A X, et al. Collaborative denoising auto-encoders for top-n recommender systems[C]// Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. 2016: 153-162.
- [8] WANG X, HE X, FENG F, et al. Tem: Tree-enhanced embedding model for explainable recommendation[C]// Proceedings of the 2018 World Wide Web Conference. 2018: 1543-1552.
- [9] WANG X, HE X, NIE L, et al. Item silk road: Recommending items from information domains to social users [C]// Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2017: 185-194.
- [10] JAMALI M, ESTER M. A matrix factorization technique with trust propagation for recommendation in social networks[C]// Proceedings of the fourth ACM conference on Recommender systems. 2010: 135-142.
- [11] KOREN Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2008: 426-434.
- [12] HAN W, ZHANG Z, ZHANG Y, et al. Contextnet: Improving convolutional neural networks for automatic speech recognition with global context[J]. *arXiv*: 2005. 03191, 2020.
- [13] WANG H, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems [C]// The World Wide Web Conference. 2019: 3307-3313.
- [14] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion [C]// Twenty-ninth AAAI Conference on Artificial Intelligence. 2015.
- [15] TANG Y, HUANG J, WANG G, et al. Orthogonal relation transforms with graph context modeling for knowledge graph embedding [J]. *arXiv*: 1911. 04910, 2019.
- [16] SHI S, ZHANG M, LU H, et al. Wide & deep learning in job recommendation: An empirical study [C]// Asia Information Retrieval Symposium. Cham: Springer, 2017: 112-124.
- [17] GUO H, TANG R, YE Y, et al. DeepFM: a factorization-machine based neural network for CTR prediction [J]. *arXiv*: 1703. 04247, 2017.
- [18] HE X, LIAO L, ZHANG H, et al. Neural collaborative filtering [C]// Proceedings of the 26th International Conference on World Wide Web. 2017: 173-182.
- [19] HE X, 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. 2017: 355-364.
- [20] SEDHAIN S, MENON A K, SANNER S, et al. AutoRec: Autoencoders meet collaborative filtering [C]// Proceedings of the 24th International Conference on World Wide Web. 2015: 111-112.
- [21] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 353-362.
- [22] WANG H, ZHANG F, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation [C]// Proceedings of the 2018 World Wide Web Conference. 2018: 1835-1844.
- [23] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix [C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long papers). 2015: 687-696.
- [24] WANG H, ZHANG F, ZHAO M, et al. Multi-task feature learning for knowledge graph enhanced recommendation [C]// The World Wide Web Conference. 2019: 2000-2010.
- [25] SUN Z, YANG J, ZHANG J, et al. Recurrent knowledge graph embedding for effective recommendation [C]// Proceedings of the 12th ACM Conference on Recommender Systems. 2018: 297-305.
- [26] WANG X, WANG D, XU C, et al. Explainable reasoning over knowledge graphs for recommendation [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 5329-5336.
- [27] WANG H, ZHANG F, WANG J, et al. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems [C]// Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018: 417-426.
- [28] CAO Y, WANG X, HE X, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences [C]// The World Wide Web Conference.

2019:151-161.

- [29] WU Z,PAN S,CHEN F,et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems,2020,32(1):4-24.
- [30] VELIČKOVIĆ P,CUCURULL G,CASANOVA A,et al. Graph attention networks[J]. arXiv:1710.10903,2017.
- [31] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.
- [32] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907,2016.
- [33] HAMILTON W L,YING R,LESKOVEC J. Inductive Representation Learning on Large Graphs[J]. arXiv:1706.02216,2017.
- [34] ZHU H,FENG F,HE X,et al. Bilinear graph neural network with neighbor interactions[J]. arXiv:2002.03575,2020.
- [35] RENDLE S,FREUDENTHALER C,GANTNER Z,et al. Bayesian personalized ranking from implicit feedback[C]//Proceedings of Uncertainty in Artificial Intelligence. 2014:452-461.
- [36] KINGMA D P,BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980,2014.
- [37] RENDLE S,GANTNER Z,FREUDENTHALER C,et al. Fast context-aware recommendations with factorization machines [C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2011:635-644.
- [38] AI Q,AZIZI V,CHEN X,et al. Learning heterogeneous knowledge base embeddings for explainable recommendation[J]. Algorithms,2018,11(9):137-137.
- [39] BORDES A,USUNIER N,GARCIA-DURAN A,et al. Translating embeddings for modeling multi-relational data[J]. Advances in Neural Information Processing Systems,2013,26:2787-2795.
- [40] WANG X,HE X,CAO Y,et al. Kgat: Knowledge graph attention network for recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:950-958.



FAN Hongyu, born in 1997, postgraduate. Her main research interests include recommendation system and so on.



MENG Xiangfu, born in 1981, Ph. D. professor, is a senior member of China Computer Federation. His main research interests include big data analysis and query, spatio-temporal data mining, and machine learning algorithms.