



# 计算机科学

COMPUTER SCIENCE

## 结合图注意力机制的知识图谱推荐算法

张小婉, 邓秋军, 柳先辉

引用本文

张小婉, 邓秋军, 柳先辉. [结合图注意力机制的知识图谱推荐算法](#)[J]. 计算机科学, 2023, 50(11A): 230100057-7.

ZHANG Xiaowan, DENG QiuJun, LIU Xianhui. [Knowledge Graph Recommendation Algorithm Combined with Graph Attention Mechanism](#) [J]. Computer Science, 2023, 50(11A): 230100057-7.

---

## 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)

[一种融合CNN和Swin Transformer的医学显微图像分割模型](#)

Medical Microscopic Image Segmentation Model Based on CNN Structure and Swin Transformer  
计算机科学, 2023, 50(11A): 230200119-8. <https://doi.org/10.11896/jsjcx.230200119>

[基于注意力机制和ConvLSTM的船舶交通流量预测算法](#)

Ship Traffic Flow Prediction Algorithm Based on Attention Mechanism and ConvLSTM  
计算机科学, 2023, 50(11A): 230800067-7. <https://doi.org/10.11896/jsjcx.230800067>

[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>

[基于图卷积网络和注意力机制的诊断预测](#)

Diagnosis Prediction Based on Graph Convolutional Network and Attention Mechanism  
计算机科学, 2023, 50(11A): 221100232-6. <https://doi.org/10.11896/jsjcx.221100232>

[基于知识蒸馏和高效通道注意力的异常检测](#)

Novelty Detection Method Based on Knowledge Distillation and Efficient Channel Attention  
计算机科学, 2023, 50(11A): 220900034-10. <https://doi.org/10.11896/jsjcx.220900034>

# 结合图注意力机制的知识图谱推荐算法

张小婉<sup>1</sup> 邓秋军<sup>2</sup> 柳先辉<sup>2</sup>

1 同济大学电子与信息工程学院 上海 201804

2 同济大学电子与信息工程学院 CAD 研究中心 上海 201804

(zhangxiaowan@tongji.edu.cn)

**摘要** 由于传统推荐算法存在数据稀疏性和冷启动问题,并且将物品作为单独的个体,没有考虑到物品之间存在的关系。为了解决这些问题,考虑引入知识图谱这一辅助信息。但现有的基于路径以及基于嵌入的知识图谱推荐算法没有考虑不同实体对于用户的重要性不同,导致重要性更低的实体对推荐结果的影响反而更大。针对这类局限性,文中提出了一种结合图注意力机制的知识图谱推荐系统,该推荐系统首先使用图嵌入方法生成用户和项目的初始表示,然后在表示传播时采用注意力机制区分不同邻居实体的重要性,通过权值加和来生成用户和项目的向量表示,最后预测层生成用户和项目的最终表示,并根据最终表示预测用户和项目交互的概率。在两个公开数据集 Amazon-book 和 Last-fm 上与其他算法进行对比实验,实验结果表明,该模型在指标 recall,ndcg,precision,HR 上均有提高,证明其能有效提高推荐的准确度。

**关键词:** 推荐系统;知识图谱;注意力机制;高阶连通性;实体传播

**中图法分类号** TP391

## Knowledge Graph Recommendation Algorithm Combined with Graph Attention Mechanism

ZHANG Xiaowan<sup>1</sup>, DENG QiuJun<sup>2</sup> and LIU Xianhui<sup>2</sup>

1 College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

2 CAD Research Center, College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

**Abstract** Due to the problems of data sparsity and cold start in traditional recommendation algorithms, and the item is regarded as a separate individual, the relationship between items is not considered. In order to solve these problems, recommender systems start to introduce auxiliary information. However, the existing path-based and embedding-based knowledge graph recommendation algorithms do not consider the importance of different entities to users, resulting in entities with lower importance having a greater impact on the recommendation results. Aiming at such limitations, this paper proposes a knowledge graph recommendation system combining graph attention mechanism, which firstly uses graph embedding method to generate initial representations of users and items, and then employs an attention mechanism to distinguish the importance of different neighbor entities during representation propagation, and generates user and item sums through weight summation. The final prediction layer generates the final representation of the user and item, and predicts the probability of user and item interaction based on the final representation. Compared with other algorithms on two public datasets Amazon-book and Last-fm, and experimental results show that the model has improved in indicators recall, ndcg, precision, HR, indicating that the model can effectively improve the accuracy of recommendation.

**Keywords** Recommendation system, Knowledge graph, Attention mechanism, Higher-order connectivity, Embedding propagation

## 1 引言

随着电子商务等数字化互联网平台的快速发展,引发数据过载、信息爆炸等问题,用户可选择的物品和服务越来越多。为了让物品和服务的提供者更精确地为用户做出推荐,方便用户对物品和服务进行筛选,推荐系统由此产生。推荐系统是信息过滤系统的子集,用于预测用户对物品的“评价”或“偏好”<sup>[1]</sup>。其根据用户历史点击行为来分析用户,预测未来的点击行为,向用户推荐物品或服务,为用户提供更多的信息帮助<sup>[2]</sup>。

知识图谱属于有向异构图,知识图谱中的节点表示实体,边表示实体之间的关系。知识图谱中的每条有向边都可由三元组(头实体,关系,尾实体)的形式表示。现有的比较流行的知识图谱有 Freebase<sup>[3]</sup>, Dbpedia<sup>[4]</sup>, YAGO<sup>[5]</sup>, NELL<sup>[6]</sup>, Bio2RDF<sup>[7]</sup>等。目前知识图谱已经被引入到多种应用中,包括搜索引擎、推荐系统、问答系统<sup>[8]</sup>、信息提取<sup>[9]</sup>等。如图 1 所示,《西游记》、四大名著、《水浒传》、《红楼梦》、《三国演义》都是知识图谱中的实体,用户 A 与实体《西游记》存在交互,根据路径《西游记》→ 四大名著 → 《水浒传》,用户 A 可能对《水浒传》存在偏好。

基金项目:国家重点研发计划(2022YFB3305700)

This work was supported by the National Key Research and Development Program of China(2022YFB3305700).

通信作者:柳先辉(xianhui\_l@163.com)

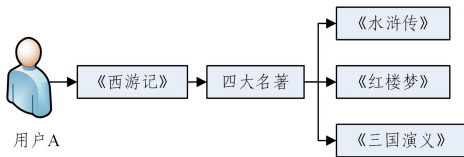


图1 知识图谱推荐示例

Fig. 1 Example of knowledge graph recommendation

传统的推荐方法有基于内容的推荐方法和基于协同过滤的推荐方法<sup>[10]</sup>。基于内容的推荐方法根据物品的自身属性学习用户和物品的表示,只考虑了物品的本身性质属性。基于协同过滤的推荐方法可以分为基于用户的协同过滤算法,给用户推荐与其有高相似度的其他用户偏好的物品或服务;基于物品的协同过滤算法,给用户推荐与其历史偏好相似度高的其他物品。但是传统的推荐方法存在数据稀疏性和冷启动的问题,为了缓解这些问题,推荐系统引入知识图谱这一辅助信息。知识图谱推荐方法分为3种,分别是基于嵌入的方法、基于路径的方法以及基于嵌入传播的方法。其中基于嵌入传播的方法将前两种方法结合,使用注意力机制区分邻居实体的重要性,递归地将与实体更相关的邻居实体的语义信息嵌入到实体,增强了知识图谱中的实体表示。但该方法在使用注意力机制来区分实体重要性时区分得不明显,例如KGAT模型<sup>[11]</sup>在使用注意力机制时,为了简化只考虑一种情况下的相似性。注意力机制<sup>[12]</sup>最初是在基于编解码器框架下处理序列到序列的转换任务,后来被应用于机器学习、推荐系统<sup>[13]</sup>。

为解决以上问题,本文提出了一种基于对偶思想和向量相似性注意力机制的知识图谱推荐算法。本文模型对高阶关系进行建模,递归地对用户和项目的表示进行传播,根据邻居的表示来更新本节点的表示;采用注意力机制,结合对偶的思想,分别从三元组的头实体和尾实体的两个方向考虑关系对邻居权重的影响;使用向量相似度的衡量方法,在关系空间下考虑两种相似情况来衡量实体的重要性。将该模型与其他基于知识图谱的推荐算法在数据集 Amazon-book 和 Last-FM 上进行对比实验,实验结果验证了该模型提高推荐结果的准确性。本文的贡献如下:

- 1) 本文在推荐算法的基础上引入了知识图谱,缓解了传统推荐模型数据稀疏与冷启动问题,提高了推荐结果的准确性。
- 2) 本文考虑了高阶信息对用户和项目的表示的影响,在聚合邻居节点特征时,采用基于对偶思想和向量相似性的注意力机制,来区分邻居节点的重要性。

## 2 相关工作

现有的融合知识增强的推荐模型,根据使用知识图谱信息的方式分成3类:基于嵌入的方法、基于路径的方法和基于嵌入传播的方法。

1) 基于嵌入的方法(Embedding-based Method)指利用知识图谱嵌入方法 KGE(Knowledge Graph Embedding)对知识图谱中的实体与关系进行嵌入表示。Zhang 等<sup>[14]</sup>提出 CKE(Collaborative Knowledge Base Embedding)模型,分别计算项目的结构嵌入、文本嵌入、视觉嵌入,使用贝叶斯框架将嵌入表示与传统的协同过滤推荐算法相结合。但是在 CKE 中,传统的协同过滤推荐和 KGE 模块在贝叶斯框架下的耦合较为

松散,使得该模型没有有效使用知识图谱的特征信息。Wang 等<sup>[15]</sup>提出的 DKN(Deep Knowledge-aware Network for News Recommendation)是新闻推荐模型,其设计了一个 CNN 框架,将实体嵌入和词嵌入相结合。但是由于知识图谱嵌入表示模块和推荐系统模块相互独立,因此无法做到端到端的训练。基于嵌入的方法虽然可以提升推荐系统的准确性,但是没有对知识图谱进行高阶建模,影响系统性能。

2) 基于路径的方法(Path-based Method)利用知识图谱结构化的特征,挖掘图谱中实体之间的关系。PER(Personalized Entity Recommendation)模型<sup>[16]</sup>和 FMG(Factorization Machine with Group lasso)模型<sup>[17]</sup>将知识图谱视为异构的信息网络,并在不同类型的关系路径或关系图下提取基于元路径或元图的潜在特征,用于表示用户和项目之间的连接性。基于路径的方法以一种更为直观的方式嵌入知识图谱。但是此类方法严重依赖手动设计的元路径或元图,在实践中很难优化;并且当实体和关系不在同一个领域中时,手工设计元路径的工作量大到几乎不可能完成。

3) 基于嵌入传播的方法(Embedding Propagation-based Methods)利用知识图谱的高阶信息,极大地提高了推荐准确度。该方法在知识图谱中以迭代传播关联信息的方式挖掘关联信息,利用获取的关联信息进行推荐,从而提高推荐准确性。Wang 等<sup>[18]</sup>提出了 RippleNet(Multiple Ripples Networks)模型,RippleNet 通过向外传播用户的历史兴趣偏好来对用户进行建模,在知识图谱中迭代用户历史兴趣偏好,以挖掘用户的潜在偏好。RippleNet 将用户的历史交互作为初始输入,以“水波”的形式传播以获得用户的潜在兴趣。Wang 等<sup>[19]</sup>提出了 KGCN(Knowledge Graph Convolutional Networks)模型,该模型为基于图卷积网络的端到端框架,在知识图谱的邻居节点之间传播并计算项目的嵌入表示,同时使用注意力机制区分邻居节点的重要性,聚合邻居节点的表示。该模型将图卷积网络应用到知识图谱推荐算法中,以有效捕获项目间的相关性,提高了推荐性能。CGAT(Contextualized Graph Attention Network)模型<sup>[20]</sup>在获取近邻节点信息的基础上,引入带偏随机游走策略,提取非近邻实体节点信息,显式地提取图谱中实体近邻节点和非近邻节点的上下文信息。CRM-SBKG<sup>[21]</sup>模型是一种基于 Siamese BERT 和知识图谱的引文推荐模型,该模型使用 BERT 模型和 Siamese 网络进行文本特征提取,并构建了一个提取实体特征的引文知识图谱,是一种挖掘文本特征和实体特征以获得推荐列表的混合模型。

综上所述,基于路径的方法需要手动定义有效的元路径,对于大规模的知识图谱来说工作量巨大,是不切实际的。基于嵌入的方法虽然避免了大量的元路径的定义工作,但其通过隐式的方式建模,没有充分挖掘用户的潜在偏好。基于嵌入传播的方法有效结合了基于嵌入的推荐方法和基于路径的推荐方法的优点,该方法沿着知识图谱中的路径,通过嵌入传播、多跳关系挖掘用户的潜在偏好,取得了良好的推荐效果。

## 3 问题描述

在此推荐系统中有两个集合,分别为用户集和物品集。用户集为  $U = \{u_1, u_2, \dots\}$ ,项目集为  $V = \{v_1, v_2, \dots\}$ 。用户与项目的交互矩阵  $\mathbf{Y} = \{y_{uv} \mid u \in U, v \in V\}$ 。当  $y_{uv} = 1$  时,表示

用户  $u$  与项目  $v$  之间存在交互,当  $y_{uv} = 0$  时,表示用户  $u$  与项目  $v$  之间没有观察到交互。本文的推荐系统使用的是隐式反馈,即  $y_{uv}$  的值为 1 时不能确定用户  $u$  的偏好项目  $v$ ,仅表示用户  $u$  和项目  $v$  之间存在交互行为,交互行为不仅可以是购买、点击、收藏,还可以是浏览等; $y_{uv}$  的值为 0 时并不能确定用户  $u$  对项目  $v$  不感兴趣。根据交互矩阵  $\mathbf{Y}$  将交互数据表示为用户-项目的二部图  $G_1$ ,定义  $G_1$  为  $\{(u, r_{\text{Interact}}, v) \mid u \in U, v \in V\}$ ,其中  $U$  表示用户集,  $V$  表示项目集,  $r_{\text{Interact}}$  表示用户  $u$  和项目  $v$  之间的交互关系。除了交互数据,已知内容还包括知识图谱  $G_2$ ,它由大量的项目-关系-项目三元组组成,定义  $G_2 = \{(h, r, t) \mid h \in E, r \in R, t \in E\}$ ,其中  $E$  表示项目集合,  $R$  表示项目间的关系集合。

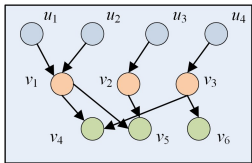


图 2 联合图示例

Fig. 2 Example of collaborative graph

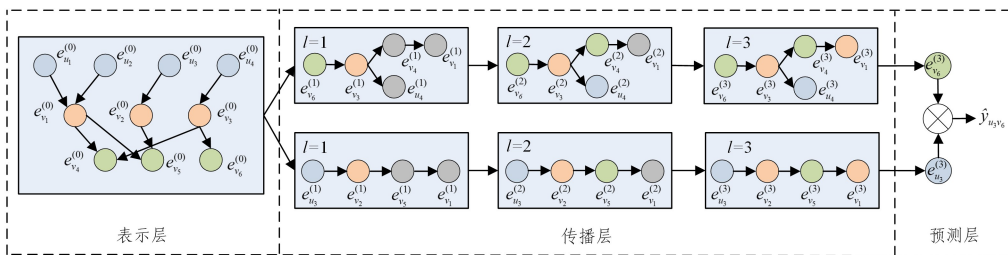


图 3 本文模型的结构

Fig. 3 Framework of the proposed model

#### 4.1 表示层

知识图谱嵌入是在保留知识图谱结构下,将实体和关系表示为向量的方法。本文使用 TransR 模型表示<sup>[20]</sup>,Trans 知识图谱嵌入是在保留知识图谱结构的前提下,将实体和关系表示为向量的方法。本文使用 TransR 模型表示<sup>[22]</sup>,TransR 在单独的实体空间和关系空间中构建实体和关系嵌入,对于一个三元组  $(h, r, t) \in G$ ,TransR 首先将实体从实体空间投影到相应的关系  $r$  空间,通过优化翻译原则  $e_h + e_r \approx e_t$  学习每个实体和关系的表示,其中  $e_h, e_t \in R^d, e_r \in R^k$ ,分别是  $h, t, r$  的表示; $e_h$  和  $e_t$  是  $e_h$  和  $e_t$  在关系  $r$  空间的投影表示。对于一个三元组,其评分函数为:

$$g(h, r, t) = \|W_r e_h + e_r - W_r e_t\|_2^2 \quad (1)$$

其中,  $W_r \in R^{k \times d}$  是关系  $r$  的变换矩阵,它将实体从  $d$  维实体空间投影到  $k$  维关系空间。 $g(h, r, t)$  的分数越低,表明三元组的向量表示越准确。TransR 的损失函数为:

$$L_{KG} = \sum_{(h, r, t, t') \in T} -\ln \sigma(g(h, r, t') - g(h, r, t)) \quad (2)$$

其中,  $T = \{(h, r, t, t') \mid (h, r, t) \in G, (h, r, t') \notin G\}$ ;  $(h, r, t')$  是随机替换有效三元组中的一个实体得到的无效三元组; $\sigma(\cdot)$  是 sigmoid 函数。

#### 4.2 传播层

在传播层中,使用  $N_h = \{(h, r, t) \mid (h, r, t) \in G\}$  表示以  $h$  为头实体的三元组集合,  $t$  表示  $h$  的邻居节点,计算  $h$  的所有邻居节点

由于  $G_1$  中的项目集  $V$  是知识图谱  $G_2$  的实体集  $E$  的子集,因此可以将  $G_1$  和  $G_2$  进行联合,形成新的联合图  $G = \{(h, r, t) \mid h, t \in E', r \in R'\}$ ,其中  $E' = E \cup U, R' = R \cup \{r_{\text{Interact}}\}$ 。图 2 给出了一个联合图  $G$  的示例,其中  $u$  属于  $G_1$  的用户集  $U$ ,与用户  $u$  直接相连的  $v_1, v_2, v_3$  属于  $G_1$  的项目集  $V$ ,不与用户  $u$  直接相连的  $v_4, v_5, v_6$  属于  $G_2$  的项目集  $E$ 。本文算法的目的是预测用户和项目交互的概率。

### 4 基于对耦与向量相似性的注意力机制的知识图谱推荐算法

本文提出了一种基于对耦思想和向量相似性的注意力机制的知识图谱推荐算法。本文模型的框架如图 3 所示,由 3 个部分组成:1)表示层,在保留知识图谱结构的条件下,将实体和关系参数化为向量的方法;2)传播层,递归地对实体的表示进行传播,根据邻居的表示来更新本节点的表示;采用注意力机制,结合对耦的思想,分别从三元组的头实体和尾实体的两个方向考虑关系对邻居权重的影响;使用向量相似度的衡量方法,在关系空间下考虑两种相似情况来衡量实体的重要性;3)预测层,输出用户对项目的预测偏好。

的线性组合  $e_{N_h}$ ,此线性组合表示实体  $h$  的一阶连通性结构:

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

其中,  $\pi(h, r, t)$  表示每个邻居节点的传播权重,  $e_t$  表示  $t$  的向量。注意力权重  $\pi(h, r, t)$  的计算方式如下。

1)仅使用最简单的注意力机制:

$$\pi(h, r, t) = (W_r e_t)^T \tanh(W_r e_h + e_r) \quad (4)$$

2)基于对耦的思想,从头实体和尾实体两个方向考虑关系对权值的影响:

$$\pi(h, r, t) = \lambda_1 (W_r e_t)^T \tanh(W_r e_h + e_r) + (1 - \lambda_1) (W_r e_h)^T \tanh(W_r e_t + e_r) \quad (5)$$

3)仅基于向量相似性,在使用 TransR 算法向量化知识图谱的实体和关系后,由 TransR 算法的翻译关系可知头实体与关系向量之和,与尾实体向量越接近,注意力权值越大。

$$\pi(h, r, t) = (W_r e_h)^T \tanh(W_r e_t - e_r) \quad (6)$$

4)同时使用耦合方式与向量相似性方式:

$$\pi(h, r, t) = \lambda_1 (W_r e_t)^T \tanh(W_r e_h + e_r) + \lambda_1 (W_r e_h)^T \tanh(W_r e_t + e_r) + (1 - 2\lambda_1) (W_r e_h)^T \tanh(W_r e_t - e_r) \quad (7)$$

其中,  $W_r \in R^{k \times d}$  是关系  $r$  的变换矩阵,它将实体从  $d$  维实体空间投影到  $k$  维关系空间; $e_h, e_t \in R^d$  和  $e_r \in R^k$  分别是本节点、邻居节点、两者关系的表示; $\lambda_1$  为超参数。

通过 softmax 函数对以实体  $h$  为头实体的所有三元组的权值  $\pi(h, r, t)$  进行归一化:

$$\pi(h, r, t) = \frac{\exp(\pi(h, r, t))}{\sum_{(h, r', t') \in N_h} \exp(\pi(h, r', t'))} \quad (8)$$

因此,在消息传播时,注意力权值越高,三元组传播的权重越大,可以筛选出有效信息,弱化无效信息。

在计算完实体  $h$  的线性组合  $e_{N_h}$  后,通过 Bi-Interaction 聚合器对  $e_h$  和  $e_{N_h}$  进行聚合:

$$f_{\text{Bi-Interaction}} = \text{LeakyReLU}(W_1(e_h + e_{N_h})) + \text{LeakyReLU}(W_2(e_h \otimes e_{N_h})) \quad (9)$$

得到新的实体  $h$  的表示,其中  $W_1, W_2 \in R^{d \times d'}$  是可训练的权重矩阵,  $\otimes$  表示向量点积。

以上是一阶图谱信息的提取方式,重复传播层即可递归地提取图谱的高阶信息,传播至  $l$  层时实体  $h$  的向量表示为:

$$e_h^{(l)} = f(e_h^{(l-1)}, e_{N_h}^{(l-1)}) \quad (10)$$

其中,  $e_h^{(l)}$  表示传播至第  $l$  层时,实体  $h$  的向量表示,  $e_{N_h}^{(l-1)}$  的计算式为:

$$e_{N_h}^{(l-1)} = \sum_{(h, r, t) \in N_h} \pi(h, r, t) e_t^{(l-1)} \quad (11)$$

其中,  $e_t^{(l-1)}$  是从先前的信息传播步骤生成的实体  $t$  的表示,它包含了来自其  $(l-1)$  跳邻居的信息;  $e_h^0$  为传播层的初始输入向量。

### 4.3 预测层

在传播  $L$  层后,  $G$  中的每个实体都可以得到  $L$  个表示。由第 3 节的问题描述可知  $G$  中的实体包括用户集  $U$  和项目集  $V$ ,因此对于一个用户  $u$  和一个项目  $v$ ,可以得到  $L$  个  $u$  的表示  $\{e_u^{(1)}, \dots, e_u^{(L)}\}$ , 和  $L$  个  $v$  的表示  $\{e_v^{(1)}, \dots, e_v^{(L)}\}$ , 不同层的输出表示不同阶的图谱连通信息。采用层聚合机制<sup>[23]</sup> 将每一层的输出表示连接成一个向量:

$$e_u^* = e_u^{(1)} \parallel \dots \parallel e_u^{(L)}, e_v^* = e_v^{(1)} \parallel \dots \parallel e_v^{(L)} \quad (12)$$

其中,  $\parallel$  是连接操作。最后,进行用户和项目表示的内积,以预测用户  $u$  对实体  $v$  的偏好分数:

$$\hat{y}(u, v) = e_u^{*T} e_v^* \quad (13)$$

### 4.4 优化

为了优化推荐模型,本文选择 BPR 损失<sup>[24]</sup>。即与某用户存在交互的项目的预测值,应当比不存在交互的项目的预测值更高。

$$L_{CF} = \sum_{(u, i, j) \in O} -\ln \sigma(\hat{y}(u, i) - \hat{y}(u, j)) \quad (14)$$

其中,  $O = \{(u, i, j) | (u, i) \in R^+, (u, j) \in R^-\}$  表示训练集,  $R^+$  表示用户  $u$  和项目  $j$  之间有(正)交互,  $R^-$  表示采样未观察到的(负)交互集;  $\sigma(\cdot)$  是 sigmoid 函数。最后,使用目标函数来联合学习方程(2)和方程(11):

$$L = L_{KG} + L_{CF} + \lambda \|\Theta\|_2^2 \quad (15)$$

其中,  $\Theta = \{E, W_l, \forall l \in R, W_l^{(d)}, W_l^{(d')}, \forall l \in \{1, \dots, L\}\}$  是模型参数集,  $E$  是对于所有实体和关系的嵌入表;使用  $\lambda$  对  $\Theta$  进行 L2 正则化以防止过拟合。

在训练时,为了最小化目标函数,本文使用 mini-batch Adam<sup>[25]</sup> 交替优化  $L_{KG}$  嵌入损失和  $L_{CF}$  预测损失。随机采样一批  $(h, r, t, t')$ , 以更新所有节点的嵌入表示;训练模型时,随机采样一批  $(u, i, j)$ , 经过  $L$  步传播后获得它们的表示,用预测损失梯度更新模型参数。

## 5 实验

### 5.1 数据集

为了评估模型的推荐性能,本实验使用了两个数据集:

Amazon-book<sup>[26]</sup> 和 Last-FM。Amazon-review 是记录亚马逊网站商品信息及用户评价的数据集,该数据集包含书籍、电影、音乐等子数据集,从这个集合中选择 Amazon-book 作为本文的数据集。Last-FM 是从 Last.fm 在线音乐系统收集的音乐收听数据集,其中的曲目被视为项目。本文为了确保数据集的质量,每个数据集都保留至少有 10 次交互的用户和项目。本文通过标题匹配的方式,将两个数据集的物品项目与 Freebase 的实体匹配映射。表 1 列出了 3 个数据集的统计数据。

表 1 两个数据集的统计数据

Table 1 Statistics of two datasets

数据集	Amazon-book	Last-FM
用户数	70 679	23 566
物品数	24 915	48 123
交互记录数	847 733	3 034 796
KG 实体数	88 572	58 266
KG 关系数	39	9
KG 三元组数	2 557 746	464 567

随机选择数据集的 80% 的历史交互数据构成训练集,将剩余的 20% 的交互数据构成测试集。随机挑选训练集的 20% 的交互数据构成验证集,用于调节超参数。用户和物品之间存在交互即视为正样例,采用负采样策略。

### 5.2 评估指标

本模型采用召回率 recall, NDCG, precision 和 HR 作为性能评价指标。

对于测试集中的每种用户,将用户没有与之交互的所有项目都视为负样例,除了训练集中的正样例项目,每种方法输出用户对其他项目的偏好分数。采用召回率  $recall@K$ 、归一化折损累计增益  $NDCG@K$ 、精确率  $precision@K$  和  $HR@K$  作为评估 Top-K 推荐和偏好排名的评估指标<sup>[27-28]</sup>, 设置  $K=20$ 。recall@K 表示正确分类的个数占全部正确数量的个数,是在用户的推荐列表中的项目与测试集项目总数的比值,其值越高,证明推荐性能越高。召回率  $recall@K$  的计算式为:

$$recall@K = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |T(u)|} \quad (16)$$

其中,  $R(u)$  是实验生成的用户  $u$  的推荐项目列表;  $T(u)$  是测试集中用户  $u$  交互的正样例项目列表。NDCG@K 用于评估推荐模型的排序性能,测试集中用户交互的正样例在推荐列表的位置越靠前,评估值越高。NDCG@K 的计算式为:

$$NDCG@K = \frac{1}{IDCG} \sum_{i=1}^K \frac{2^{rel_i - 1}}{\log_2(i+1)} \quad (17)$$

其中,  $rel_i$  是待推荐项目与用户历史交互项目之间的相似度,即项目在推荐列表位置  $i$  的相似度,取值为  $[0, 1]$ , IDCG 的计算式为:

$$IDCG@K = \sum_{i=1}^{|REL|} \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (18)$$

其中,  $|REL|$  是将待推荐的项目按照相关性从大到小排序,取前  $K$  个项目组成的序列。精确率  $precision@K$  表示正确分类的个数占预测正确数量的个数,其值越高,证明推荐性能越高,其计算式为:

$$precision@K = \frac{\sum_{u \in U} |R(u) \cap T(u)|}{\sum_{u \in U} |R(u)|} \quad (19)$$

$HR@K$  (Hit rate) 衡量召回率,该指标的分子是所有用户的测试集的项目总和,分子是每个用户前  $K$  个中属于测试集的个数的总和,该指标越大,证明推荐性能越高。

### 5.3 对比模型

为了证明本文模型的有效性,本文使用 NFM,CKE,CFKG,RippleNet,GC-MC 方法作为对比模型。

1)NFM<sup>[29]</sup>:该模型是较为先进的分解模型,在传统 FM 模型的基础上,使用神经网络获取用户和项目之间的非线性特征。本文按照文献[15]在输入特征时加入一层隐藏层。

2)CKE:该模型是一种代表性的基于正则化的推荐方法,该方法将协同过滤与结构知识、文本知识、图像知识统一在贝叶斯框架中<sup>[24]</sup>。在嵌入结构信息时,使用 TransR 算法将结构知识嵌入以获得项目的表示。

3)CFKG<sup>[30]</sup>:该模型构建用户-项目联合图,使用 TransE<sup>[31]</sup>算法将联合图中的用户和项目嵌入表示,以此利用知识图谱的辅助信息进行推荐。

4)RippleNet:该模型将正则化方法和基于路径的方法相结合,将表示以“水波”的形式传播,以获得用户的兴趣偏好。

5)GC-MC<sup>[32]</sup>:该模型在图结构数据上应用 GCN<sup>[33]</sup>编码器,本文将 GCN 编码器应用于用户-项目联合图,按照 GC-

MC 加入一层图卷积层,其维度设置与向量的维度相同。

### 5.4 实验设置

本文在实验时使用 Tensorflow 实现本文模型。用户、物品和知识图谱中的实体、关系的表示向量维度都固定为 64, batch size 大小固定为 1024。本文模型的注意力传播层的深度为 3,维度分别为 64,32,16。本文实验使用 Adam 优化器进行优化,使用 Xavier 初始化器<sup>[34]</sup>初始化模型参数。学习率在{0.05,0.01,0.005,0.001}之间调整,L2 归一化的系数在{ $10^{-5}$ , $10^{-4}$ , $\dots$ , $10^1$ , $10^2$ }之间调整,丢失率 dropout 和节点丢失率 node\_dropout 在{0,0.1, $\dots$ ,0.8}之间调整,消息丢失率 message\_dropout 为 0.1。本文实验采用提前终止策略,如果验证集上的 recall@20 在连续 50 个 epoch 后没有增加,则实验提前终止。

### 5.5 实验结果及分析

#### 5.5.1 实验结果对比

所有基线模型与本文模型在 Amazon-book 和 Last-fm 两个数据集上的对比实验结果如表 2 所列。

表 2 推荐模型的性能对比

Table 2 Performances comparison of different recommendation model

	Amazon-Book				Last-FM			
	recall	ndcg	precision	HR	recall	ndcg	precision	HR
NFM	0.1366	0.0913	0.0135	0.2218	0.0829	0.1214	0.0323	0.3148
CKE	0.1326	0.0893	0.0140	0.2333	0.0730	0.1181	0.0305	0.3236
CFKG	0.1171	0.0793	0.0124	0.2064	0.0676	0.1083	0.0275	0.2950
RippleNet	0.1336	0.0910	—	—	0.0791	0.1238	—	—
GC-MC	0.1316	0.0874	—	—	0.0818	0.1253	—	—
本文算法	<b>0.1489</b>	<b>0.1143</b>	<b>0.0154</b>	<b>0.2546</b>	<b>0.0863</b>	<b>0.1724</b>	<b>0.0343</b>	<b>0.3542</b>

由表 2 所列的对比结果可以看出,本文模型在两个数据集上的两个指标都表现最佳。指标 recall@20 在 Amazon-book 和 Last-fm 上分别提高了 21.4%和 21.7%;指标 ndcg@20 在 Amazon-book 和 Last-fm 上分别提高了 30.6%和 37.2%;指标 precision@20 在 Amazon-book 和 Last-fm 上分别提高了 19.5%和 19.8%;指标 HR@20 在 Amazon-book 和 Last-fm 上分别提高了 18.9%和 16.7%。

1)NFM 的表现仅次于本文模型,比其他模型的表现都更好。因为 NFM 的隐藏层可以捕获用户表示和项目表示之间非线性的特征关系。但是 NFM 的表现不如本文模型,其原因在于注意力表示传播的方法对知识图谱中的高阶连通信息的利用更加充分,以此丰富用户表示和项目表示,从而得到更好的实验结果。

2)CKE 和 CFKG 的表现不如 NFM,可能是因为基于正则化的推荐方法无法充分利用知识图谱辅助信息来丰富用户

和项目的表示。

3)RippleNet 的表现验证了相邻实体的表示的整合对于丰富用户表示的重要性,验证了高阶连通性和相邻实体建模对推荐结果的重要性。

4)GC-MC 使用的是固定权重值,本文模型根据三元组语义关系,结合对耦的思想,考虑三元组的关系对头、尾实体的影响来决定权重值;并且依据向量相似性来决定权重值,相似性越高,权重值越高。本文模型比 GC-MC 的效果更好,证明了结合对耦思想和基于向量相似性的注意力机制对推荐效果有一定作用。

#### 5.5.2 消融实验

为了验证本文提出的基于耦合思想与向量相似性的知识图谱推荐算法的改进对模型推荐性能提升的有效性,本文设计了消融实验进行对比,实验结果如表 3 所列。

表 3 消融实验的结果

Table 3 Results of ablation experiment

	Amazon-Book				Last-FM			
	recall	ndcg	precision	HR	recall	ndcg	precision	HR
Attention-only	0.14861	0.10076	0.01544	0.25476	0.08574	0.13325	0.03442	0.35380
Coupled-att	0.14877	0.11427	0.01543	0.25471	0.08594	0.17159	0.03427	0.35360
Similarity-att	0.14836	0.11401	0.01539	0.25418	0.08586	0.17168	0.03425	0.35400
本文算法	<b>0.14876</b>	<b>0.11434</b>	<b>0.01541</b>	<b>0.25463</b>	<b>0.08630</b>	<b>0.17241</b>	<b>0.03429</b>	<b>0.35421</b>

Attention-only 表示仅使用最简单的注意力机制,对应式(4);Coupled-att 表示基于对耦的思想,从头实体和尾实体两个方向考虑关系对权值的影响,对应式(5);Similarity-att 表示仅基于向量相似性,在使用 TransR 算法向量化知识图谱的实体和关系后,头实体与关系之和与尾实体向量越接近,注

意力权值就越大,对应式(6);本文算法同时使用耦合方式与向量相似性方式,对应式(7)。

由表 3 可以看出,本文算法在两个数据集上的两个指标都表现最佳。在 Amazon-Book 数据集上,recall@20,ndcg@20,precision@20 和 HR@20 分别提升了 0.27%,11.9%,

0.13%和0.18%;在 Last-FM 数据集上, recall@20, ndcg@20, precision@20 和 HR@20 分别提升了 0.65%, 22.71%, 0.12%和 0.17%。由此可知, 本文利用交互单元可以有效提取用户的偏好信息, 建模用户之间的相关性, 提高模型的推荐性能。由此可知, 从头实体和尾实体两个方向考虑关系对权值的影响, 并且根据头尾实体向量相似性越高, 注意力权值越大的思想, 可以有效提升推荐的有效性。

**结束语** 本文提出了一种基于注意力机制的知识图谱推荐算法, 该算法包括表示层、传播层、预测层。其中传播层是核心, 使用邻居节点的向量表示更新节点的向量表示, 在更新时使用图注意力机制, 区分邻居节点的重要性, 以更精确地传播表示。基于对耦的思想, 从三元组正反两个方向考虑关系对传播权值的影响; 同时基于向量相似性, 相似性越高的邻居, 其节点权值越大。该算法将知识图谱作为辅助信息, 缓解了将物品作为单独的个体, 没有考虑到物品之间存在关系的问题, 在一定程度上缓解了传统推荐算法的数据稀疏性和冷启动问题。在数据集 Amazon-book 和 Last-FM 上的实验结果证明了本文算法的有效性, 提高了推荐结果的准确性, 探索了神经网络和注意力机制在推荐应用中的潜力。

## 参 考 文 献

- [1] KHAN N, MA Z M, ULLAH A, et al. Categorization of knowledge graph based recommendation methods and benchmark datasets from the perspectives of application scenarios: A comprehensive survey[J]. *Expert Systems with Applications*, 2022, 206: 117737.
- [2] YU C L, WANG W Q, LIU X, et al. FolkScope: Intention Knowledge Graph Construction for E-commerce Commonsense Discovery[J]. *arXiv*:2211.08316, 2023.
- [3] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]// *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*. ACM, 2008: 1247-1250.
- [4] LEHMANN J, ISELE R, JAKOB M, et al. Dbpedia-a large-scale, multilingual knowledge base extracted from wikipedia[J]. *Semantic Web*, 2015, 6(2): 167-195.
- [5] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]// *Proceedings of the 16th International Conference on World Wide Web*. ACM, 2007: 697-706.
- [6] CARLSO N, BETTERIDGE J, WANG R C, et al. Coupled semi-supervised learning for information extraction[C]// *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. ACM, 2010: 101-110.
- [7] BELLEAU F, NOLIN M A, TOURIGNY N, et al. Bio2rdf: towards a mashup to build bioinformatics knowledge systems[J]. *Journal of Biomedical Informatics*, 2008, 41(5): 706-716.
- [8] ZHOU L, GAO J F, LI D. The Design and Implementation of XiaoIce, an Empathetic Social Chatbot[J]. *Computational Linguistics*, 2020, 46(1): 53-93.
- [9] GAUR M, GUNARATNA K, SRINIVASAN V, et al. ISEQ: Information Seeking Question Generation Using Dynamic Meta-Information Retrieval and Knowledge Graphs[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022, 36(10): 10672-10680.
- [10] WU L, HE X, WANG X, et al. A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(5): 4425-4445.
- [11] WANG X, HE X N, CAO Y X, et al. KGAT: Knowledge graph attention network for recommendation[C]// *Proc of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM Press, 2019: 950-958.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [13] TAN T Z, YAN J B. Hybrid recommendation system based on self-attention model[J]. *Computer Engineering and Applications*, 2020, 56(13): 172-180.
- [14] 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*. ACM Press, 2016: 353-362.
- [15] WANG H W, ZHANG F Z, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation[C]// *The 27th International Conference on World Wide Web*. ACM, 2018: 1835-1844.
- [16] YU X, REN X, SUN Y Z, et al. Personalized entity recommendation: A heterogeneous information network approach[C]// *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. ACM, 2014: 283-292.
- [17] ZHAO H, YAO Q M, LI J D, et al. Metagraph based recommendation fusion over heterogeneous information networks[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017: 635-644.
- [18] WANG H W, ZHANG F Z, WANG J J, et al. Ripplenet: Propagating user preferences on the knowledge graph for recommender systems[C]// *Proc of the 27th ACM International Conference on Information and Knowledge Management*. ACM Press, 2018: 417-426.
- [19] WANG H W, ZHAO M, XIE X, et al. Knowledge Graph Convolutional Networks for Recommender Systems[C]// *The World Wide Web Conference*. ACM Press, 2019: 3307-3313.
- [20] LIU Y, YANG S, XU Y, et al. Contextualized Graph Attention Network for Recommendation With Item Knowledge Graph[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 181-195.
- [21] YUAN M, WAN J, WANG D. CRM-SBKG: Effective Citation Recommendation by Siamese BERT and Knowledge Graph [C]// *International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*. 2023: 909-914.
- [22] LIN Y K, LIU Z Y, SUN M S, et al. Learning Entity and Relation Embeddings for Knowledge Graph Completion[C]// *Twenty-ninth AAAI Conference on Artificial Intelligence*. AAAI, 2015: 2181-2187.
- [23] XU K, LI C T, TIAN Y L, et al. Representation Learning on Graphs with Jumping Knowledge Networks[C]// *International Conference on Machine Learning*. ACM, 2018: 5449-5458.

- [24] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]// International Conference on Uncertainty in Artificial Intelligence. AUAI, 2009; 452-461.
- [25] KINGMA D P, JIMMY B A. Adam: A Method for Stochastic Optimization[J]. International Conference on Learning Representations. arXiv: 1412. 6980, 2014.
- [26] HE R N, MCAULEY J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering[C]// International World Wide Web Conferences. ACM, 2016; 507-517.
- [27] HE X N A, LIAO L Z, ZHANG H W, et al. Neural Collaborative Filtering [C] // The World Wide Web Conference. ACM Press, 2017; 173-182.
- [28] 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. ACM, 2018: 140-144.
- [29] HE X N, CHUA T S. Neural Factorization Machines for Sparse Predictive Analytics[C]// International Conference on Research on Development in Information Retrieval. ACM, 2017; 355-364.
- [30] AI Q Y, AZIZI V, CHEN X, et al. Learning Heterogeneous Knowledge Base Embeddings for Explainable Recommendation [J]. Algorithms, 2018, 11(9): 137.
- [31] BORDES A, USUNIER N, GARCÍA-DURÁN A, et al. Translating Embeddings for Modeling Multi-relational Data [C] // NeurIPS. MIT Press, 2013; 2787-2795.
- [32] VAN DEN BERG R, KIPF T N, WELIN G. Graph Convolutional Matrix Completion[J]. arXiv: 1706. 02263, 2017.
- [33] THOMAS N K, MAX W L. Semi-Supervised Classification with Graph Convolutional Networks. In ICLR [J]. arXiv: 1609. 02907, 2016.
- [34] XAVIER G, YOSHUA B. Understanding the difficulty of training deep feedforward neural networks [C] // Artificial Intelligence and Statistics. JMLR, 2010: 249-256.



**ZHANG Xiaowan**, born in 1997, master candidate. Her main research interests include recommender systems and knowledge graph.



**LIU Xianhui**, born in 1979, Ph.D, associate researcher, associate professor. His main research interests include machine learning, data mining and big data, networked manufacturing.