

融合知识图谱多维度信息的 电力科研成果推荐算法

徐晓轶[†], 毛艳芳, 吕晓祥

(国网江苏省电力有限公司南通供电分公司, 江苏 南通 226300)

摘 要:随着电力行业研究的迅速发展,科研工作者需要从海量文献中筛选相关文献。现有的推荐系统往往忽视了文献间深层次的语义联系和专业领域的内在差别,导致推荐结果的准确性和个性化程度不足。针对这一问题,本文提出了一种基于知识图谱的电力科研成果推荐算法。首先,本文构建了用于电力科研领域推荐算法数据集,具体为构建出电力科研成果知识图谱,其中电力关键词作为推荐主体,文献标题作为推荐对象,并从标题中提取研究内容和方法信息以丰富知识图谱结构,利用查询的电力科研关键词来推荐所需的电力科研文献。其次,本文在基于知识图谱的意图网络(KGIN)模型的基础上,聚合了电力关键词和推荐文献邻居节点的信息,提出一种改进的 KGIN 模型,该模型可以更好地表达关键词和推荐文献之间的关系,融合了意图感知信息、关系感知信息、语义信息和高阶结构信息,实现了关键词和推荐文献更丰富的嵌入表示,获得了更好的推荐效果。最终实验结果表明,相较于本文所述的一些推荐系统基线模型,本文模型明显提升了对电力科研成果推荐的效果。

关键词:知识图谱;推荐系统;图卷积神经网络;电力科研成果

中图分类号:TP391

文献标识码:A

A Recommendation Algorithm for Electric Power Research Results by Fusing Multi-dimensional Information from Knowledge Graphs

XU Xiaoyi[†], MAO Yanfang, LYU Xiaoxiang

(Nantong Power Supply Branch, State Grid Jiangsu Electric Power Company Limited, Nantong, Jiangsu 226300, China)

Abstract: With the rapid development of research in the electric power industry, researchers need to filter relevant literature from a huge amount of literature. Existing recommendation systems often ignore the deep semantic links between the literature and the intrinsic differences within the professional field, resulting in insufficient accuracy and personalisation of the recommendation results. To address this problem, this paper proposes a knowledge graph-based recommendation algorithm for electric power research results. Firstly, this paper constructs a recommendation algorithm dataset for the field of electric power research, specifically to construct a knowledge graph of electric power research results, in which the electric power keywords are used as the recommendation subject, the title of the literature as the recommendation object, and the research content and method information is extracted from the title in order to enrich the structure of the knowledge graph, and the query electric power research keywords are used to recommend the required electric power research literature. Secondly, based on the knowledge graph-based intent network (KGIN) model, this paper aggregates the information of the neighbour nodes of the electric power keywords and the recommended literature, and puts forward an improved KGIN model, which can better express the relationship between the keywords and the recommended literature, and integrates the intent-aware information, the relationship-aware information, the semantic information, and the higher-order structural information, and

收稿日期:2024-07-05

基金项目:国网江苏省电力有限公司科技项目(J2023051)

作者简介:徐晓轶(1984—),男,江苏南通人,硕士,高级工程师,研究方向:电气工程。

[†] 通信联系人, E-mail: 2840328472@qq.com

achieves a richer embedding representation of the keywords and the recommended literature, and obtains a better representation of the keywords and the recommended literature. rich embedding representation, and obtains better recommendation effect. The final experimental results show that compared with the baseline models of some recommendation systems described in this paper, the model in this paper obviously improves the effect of recommending scientific research results on electric power.

Key words: knowledge graph; recommender system; graph convolutional neural network; electric power research results

电力作为现代社会运行的重要支撑。对于经济社会发展与安全具有重要意义。同时,科研活动在电力领域是推动能源转型、确保供电安全、促进环境保护的关键。通过科研,可以为构建新型电力系统增添动力,为社会经济发展提供坚实支撑。然而,电力科研领域涵盖广泛,研究人员需要花费大量时间来查找和阅读相关文献、论文和数据。因此,针对电力科研成果进行推荐是十分必要的,这不仅可以提高研究效率,促进学术交流和合作,激发新的研究思路和创新,也可以为决策者和管理者提供科学依据和参考,支持科研战略规划和决策制定。

推荐系统是一种信息过滤系统,通过分析用户的历史行为、兴趣和偏好,为用户提供个性化的推荐内容^[1]。推荐系统主要应用于电子商务、社交媒体、音视频娱乐、新闻阅读等领域。协同过滤算法(CF)是经典的推荐算法之一,它假设具有相同交互行为的用户对推荐项目具有相似的兴趣。在广泛的应用中,协同过滤算法被验证是有效的^[2]。然而,由于数据的稀疏性以及缺乏历史数据时的冷启动问题,在电力科研领域科研成果推荐过程中,随着科研文献数量的迅速增加,传统的基于协同过滤算法的推荐方式越来越难以满足研究人员对科研成果推荐的需求^[3]。由此,基于知识图谱进行科研成果推荐的研究是必要的。

本文提出的基于融合用户意图感知信息、图谱语义信息和高阶结构信息的改进 KGIN 模型的电力科研成果推荐算法,将每个用户的意图建模为知识图谱中关系的注意力组合,通过注意力策略实现意图嵌入,通过多个潜在意图描述用户项目关系,然后依托用户潜在意图和电力科研成果知识图谱,对意图信息和知识图谱信息进行聚合,得到集合意图感知信息、关系感知信息、语义信息和高阶结构信息的用户和项目的多阶嵌入表示,实现了对电力科研成果推荐效果的提高。

1 相关工作

协同过滤算法是常用的推荐算法之一,文献[4]提出了一种用于个性化排序的通用优化准则(Bayesian Personalized Ranking, BPR),通过对问题的贝叶斯分析得出的最大后验估计值,性能优于传统的矩阵因式分解^[5]方法。为解决协同过滤算法数据稀疏与冷启动问题,目前许多研究者将推荐系统与知识图谱相结合^[6]。知识图谱是一种存储实体及其相互关系的结构化知识库,主要由实体和关系组成。基于知识图谱的推荐算法可以分为基于嵌入的方法^{[7][8]}、基于路径的方法^{[9][10]}和基于混合的方法。基于路径的推荐利用知识图谱中实体间的关系路径实现项目推荐,但可能面临计算复杂、对稀疏数据敏感导致推荐多样性和准确性下降的问题^[11]。基于嵌入的推荐通过将用户和推荐目标映射到低维空间以发现潜在关联,但可能忽略复杂的上下文信息,导致推荐的解释性不足^[12]。

基于混合的方法结合上述两种方法的优点,采用图神经网络^{[13][14]},将实体和关系的语义表示与连接信息相融合,通过在知识图谱上的聚合与传播获取更多高阶信息。文献[15]采用一种异构网络嵌入方法 TransR^[16],提出了一种协同知识库嵌入(Collaborative Knowledge Base Embedding, CKE)框架,联合学习协同过滤中的潜在表征以及知识库中的推荐对象的语义表征,是一种基于嵌入的方法。基于统一的方法充分利用实体和关系的语义表示以及连接信息。文献[17]提出的知识图谱卷积网络(Knowledge Graph Convolutional Network, KGCN)是一个端到端的框架,通过挖掘知识图谱上的相关属性,有效捕捉物品间的关联性。文献[18]提出了知识图谱关注网络(Knowledge Graph Attention Network, KGAT),它以端到端的方式对知识图谱中的高阶连接关系进行了明确建模,递归传播来自节点邻居的嵌入,以完善

节点的嵌入,并采用注意力机制来区分邻居的重要性。文献[19]提出了一个“意图”的概念,并构建了一种基于知识图谱的意图网络(Knowledge Graph-Based Intent Network, KGIN),它将“意图”定义为知识图谱中关系的注意力组合,提高了建模能力,同时设计了一种新的信息聚合方案,递归地整合了长距离连接的关系路径。

为了实现对电力科研成果推荐的有效性,本文的主要贡献在于,一方面基于文献检索平台的文献信息构建出电力科研成果知识图谱,并通过提取文章标题中的信息,实现了对知识图谱信息的补充;另一方面,依托电力科研成果知识图谱,利用融合了意图感知信息、关系感知信息、语义信息和高阶结构信息的改进 KGIN 模型,实现了信息集成更加丰富的关键词和文章的多阶嵌入表示,获得了优于其他模型的推荐效果。

2 基于知识图谱的电力科研成果推荐

2.1 电力科研成果知识图谱与推荐数据集构建

2.1.1 电力科研成果知识图谱构建

为了构建电力科研成果知识图谱,首先,通过爬虫技术从学术文献检索和数据库平台中获取电力领域的相关科研文献;然后,通过去重、去噪、缺失值处理等手段对获取的数据进行预处理,最终得到包含对应文献标题、作者、关键词的电力科研成果数据,可以直接构成电力科研成果知识图谱。同时,文献标题本身包含丰富的信息,能反映研究内容和研究方法,为提升推荐效果,对知识图谱进行数据增强。在文献标题中,相当一部分的标题具有特定的结构,以本文标题“融合知识图谱多维度信息的电力科研成果推荐算法”为例,通过文本规则匹配可以获取研究方法为“知识图谱”,研究内容为“电力科研成果推荐算法”。由此在知识图谱中可以增加研究方法和研究内容信息。

如图 1 所示,电力科研成果知识图谱中包含 5 类实体:文章、关键词、作者、研究方法和研究内容。同时对应的还包含 5 种关系:文章—作者、文章—关键词、文章—研究方法、文章—研究内容、关键词—作者。

2.1.2 电力科研成果推荐数据集构建

本文将关键词作为用户,将文章作为项目,按照上述构建知识图谱方法获取到的关键词—文章数据可能会出现一个关键词只对应少量甚至一篇

该现象时以该关键词为检索词从检索平台中进一步获取一定数量的文章,同时为避免引入更多关键词,将这些文章的关键词信息替换为该关键词,最终保证构建的数据集中,一个关键词可以对应到足够数量的文章。构建电力科研成果推荐数据集的具体过程如图 2 所示。

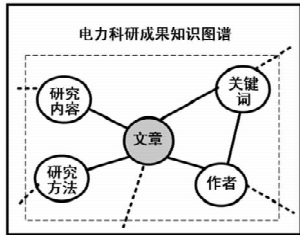


图 1 电力科研成果知识图谱示意图

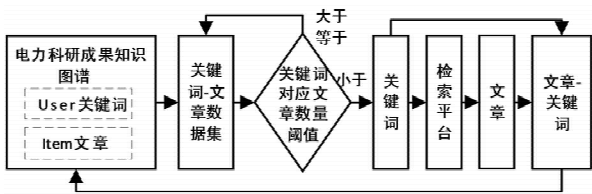


图 2 电力科研成果推荐数据集构建

2.2 基于知识图谱的电力科研成果推荐算法

为了实现对电力科研成果的有效推荐,本文提出了一种融合知识图谱实体邻域信息的改进 KGIN 模型。如图 3 所示,模型包括关键词意图建模、关键词嵌入、文章嵌入和模型预测 4 个部分。其中关键词嵌入在本文即为电力关键词嵌入表示,文章嵌入即为推荐文献嵌入表示。

首先,模型通过关键词意图建模部分学习意图的表示,将每一个意图建模为电力科研成果知识图谱中的关系的注意力组合,构建本文电力关键词潜在意图图;其次,在电力关键词嵌入部分中,通过构建的意图图聚合电力关键词与推荐文献的意图感知历史,此外,为了进一步丰富电力关键词嵌入信息,通过电力科研成果知识图谱聚合关键词和文章的实体邻域信息,集成关键词对应文章的意图感知信息以及实体邻域信息,在多阶连通信息的基础上实现关键词的多阶嵌入表示;同时,在文章嵌入部分,基于多阶连通信息利用电力科研成果知识图谱也可以构建出文章的多级嵌入表示;最后,在模型预测部分,集成关键词和文章的多阶嵌入表示,获取包含意图感知信息和电力科研成果知识图谱语义及高维结构信息的最终编码并预测出关键词对应该文章的可能性。

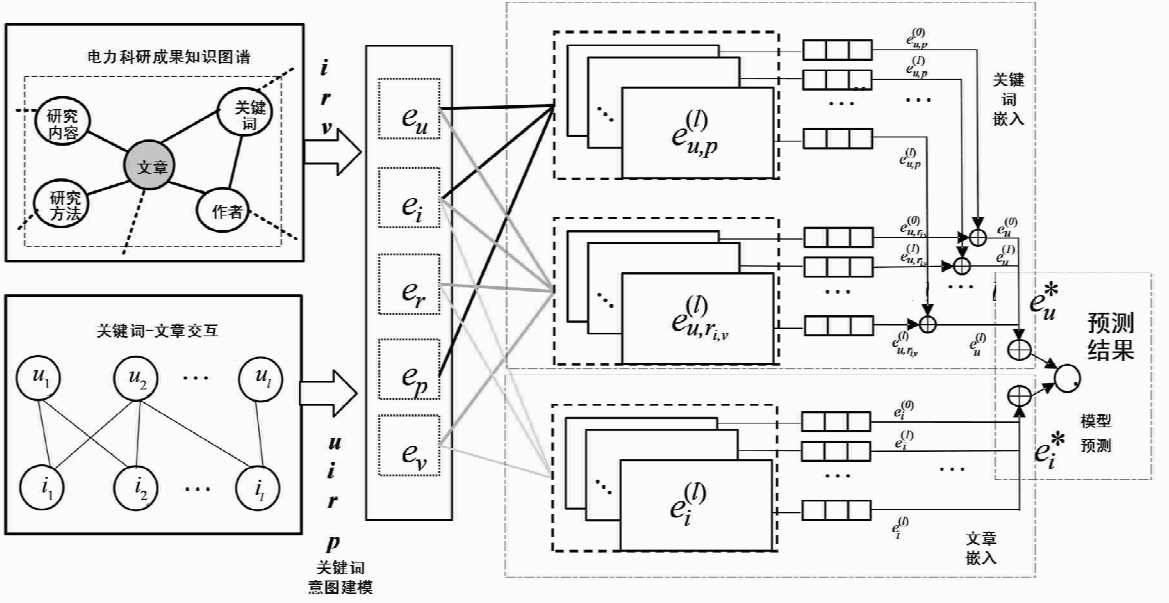


图3 模型结构

2.2.1 关键词意图建模

在科研活动中,电力关键词(本文中为推荐系统中的用户)对某个文章(本文中为推荐系统中的项目)的选择通常受多重因素影响,这些因素合在一起形成了关键词的特定偏好,即选择该文章的动机。对此,可以将关键词的不同选择模式抽象为不同的意图,在更细分的维度中增强关键词和文章的协同过滤效果,即有相同意图的关键词有相似的文章选择模式。

在关键词意图建模过程中,假设 P 是意图集,由所有关键词共享,将关键词 u 和文章 i 的关系分成 $n = |P|$ 个意图,即 $\{(u, p, i) | p \in P\}$ 。因此,可以利用关键词和文章的交互数据构建一个新的异构图,称为意图图(IG)。最后,可以使用注意力策略实现意图嵌入,如式(1)所示。

$$e_p = \sum_{r \in R} \alpha(r, p) e_r \quad (1)$$

式中: e_p 是意图嵌入; e_r 是关系嵌入; $\alpha(r, p)$ 是 e_r 的注意力得分,计算方式如式(2)所示。

$$\alpha(r, p) = \frac{\exp(w_{rp})}{\sum_{r' \in R} \exp(w_{r'p})} \quad (2)$$

式中: w_{rp} 是关系 r 与意图 p 之间的可训练权重。

在意图嵌入过程中,意图的独特性对于揭示关键词行为模式至关重要,若某意图可以被其他意图推断出,则认为该意图是冗余的。因此,为了提升意图嵌入的表达能力,应确保各意图之间的独立性,本文利用互信息方式量化意图嵌入的独立性,

如式(3)所示。

$$L = - \sum_{p \in P} \log \frac{\exp(s(e_p, e_p)/\tau)}{\sum_{p' \in P} \exp(s(e_p, e_{p'})/\tau)} \quad (3)$$

式中: L 是独立性指标; s 是余弦相似度计算函数; τ 是超参数。

2.2.2 关键词嵌入

在构建好的 IG 图中,具有相同意图的关键词有相似的文章选择模式, C 是关键词-文章交互集合,可以使用 $N_u = \{(p, i) | (u, p, i) \in C\}$ 表示关键词 u 在关键词-文章交互图中的一阶连通信息。然后,通过该一阶连通信息可以集成关键词对应文章的意图信息,实现关键词 u 的意图感知嵌入表示,如式(4)所示。

$$e_{u,p}^{(1)} = \frac{1}{|N_u|} \sum_{(p,i) \in N_u} \text{attention}(u, p) e_p \cdot e_i^{(0)} \quad (4)$$

$$\text{attention}(u, p) = \frac{\exp(e_p \cdot e_{u,p}^{(0)})}{\sum_{p' \in P} \exp(e_{p'} \cdot e_{u,p}^{(0)})} \quad (5)$$

式中: $e_{u,p}^{(1)}$ 是关键词的一阶意图感知嵌入表示; e_p 是意图嵌入; $e_i^{(0)}$ 是文章对应 ID 的嵌入表示; $e_{u,p}^{(0)}$ 是关键词对应 ID 的嵌入表示; $\cdot$ 表示相同位置元素逐个相乘; $\text{attention}(u, p)$ 是关键词 u 对意图 p 的注意力分数,用来区分不同意图的重要性,计算方法如式(5)所示。

为了提升关键词嵌入信息的表示能力,本文融入关键词文章的实体邻域信息。对一对候选关键词 u 和文章 i , K 表示知识图谱中“实体-关系-实体”集合,可以使用 $N_v = \{(i, v) | (i, r, v) \in K\}$ 表

示知识图谱中直接连接到 i 的实体集,即文章 i 在知识图谱上的一阶连通信息。通过该一阶连通信息可以集成文章实体的邻域信息,实现关键词 u 的实体邻域嵌入表示,如式(6)所示。

$$e_{u,r_{i,v}}^{(1)} = \text{ReLU}(W(e_{u,r_{i,v}}^{(0)} + \sum_{(1,v') \in N_v} \text{score}(u, r_{i,v'}) e_{u,r_{i,v'}}^{(0)}) + b) \quad (6)$$

$$\text{score}(u, r_{i,v'}) = \frac{\exp(e_u^{(0)} \cdot e_{r_{i,v'}})}{\sum_{(i,v') \in N_v} \exp(e_u^{(0)} \cdot e_{r_{i,v'}})} \quad (7)$$

式中: $e_{u,r_{i,v}}^{(1)}$ 是关键词的一阶实体邻域嵌入表示; $e_{u,r_{i,v}}^{(0)}$ 是文章 i 对应 ID 的嵌入表示; $e_{u,r_{i,v'}}^{(0)}$ 是实体 v' 对应 ID 的嵌入表示; $r_{i,v'}$ 是文章 i 与实体 v' 对应的关系 ID 的嵌入表示; $r_{i,v}$ 是文章 i 与知识图谱中相邻实体的关系; $e_u^{(0)}$ 是关键词对应 ID 的嵌入表示; $\cdot$ 表示相同位置元素逐个相乘; $\text{score}(u, r_{i,v})$ 是关键词 u 对应关系 $r_{i,v}$ 的归一化关键词关系得分,计算方法如式(7)所示。

同时,在实际的知识图谱中, $|N_v|$ 的值对于不同的实体存在差异。为了保持每个批次的计算模式固定且更高效,为每个实体均匀采样一组固定大小的邻居实体,即使用固定的感受野。本文中设置感受野为 4。

最终,基于关键词—文章交互图和知识图谱中的一阶连通信息,可以获得关键词 u 最终的嵌入表示,如式(8)所示。同样地,对于关键词 u 的高阶嵌入表示也可以通过类似的方式获得,如式(9)~式(11)所述,这种方式使得它们记住从它们的 $(l-1)$ 跳邻居传播的关系信号。

$$e_u^{(1)} = e_{u,p}^{(1)} + e_{u,r_{i,v}}^{(1)} \quad (8)$$

$$e_u^{(l)} = e_{u,p}^{(l-1)} + e_{u,r_{i,v}}^{(l-1)}, l \geq 2 \quad (9)$$

$$e_{u,p}^{(l)} = \frac{1}{|N_u|} \sum_{(p,i) \in N_u} \frac{\exp(e_p \cdot e_{u,p}^{(l-1)})}{\sum_{p' \in P} \exp(e_{p'} \cdot e_{u,p}^{(l-1)})} e_p \cdot e_i^{(l-1)}, l \geq 2 \quad (10)$$

$$e_{u,r_{i,v}}^{(l)} = \text{ReLU}(W(e_{u,r_{i,v}}^{(l-1)} + \sum_{(1,v') \in N_v} \text{score}(u, r_{i,v'}) e_{u,r_{i,v'}}^{(l-1)} + b), l \geq 2) \quad (11)$$

2.2.3 文章嵌入

对于文章的嵌入表示同样基于知识图谱中的实体连通信息,同上所述, K 表示知识图谱中“实体—关系—实体”集合, $N_i = \{(r, v) | (i, r, v) \in K\}$ 表示文章 i 的连通信息,通过该一阶连通信息可以集成与文章连接的实体的关系感知信息,实现文章 i 的嵌入表示,如式(12)所示。

$$e_i^{(1)} = \frac{1}{|N_i|} \sum_{(r,v) \in N_i} e_r \cdot e_v^{(0)} \quad (12)$$

式中: $e_i^{(1)}$ 文章 i 的嵌入表示; e_r 是关系嵌入; $e_v^{(0)}$ 是实体 v 对应 ID 的嵌入表示。

同时,对于文章 i 的 l 跳路径,它包含一系列相连的三元组 $i \xrightarrow{r_1} s_1 \xrightarrow{r_2} \dots s_{l-1} \xrightarrow{r_l} s_l$, 它的关系路径可以被表示为关系的序列 $(r_1, r_2, \dots, r_l)$, 则文章 i 的高阶嵌入表示如式(13)所示。

$$e_i^{(l)} = \sum_{s \in N_i^l} \frac{e_{r_1}}{|N_{s_1}|} \cdot \frac{e_{r_2}}{|N_{s_2}|} \cdot \dots \cdot \frac{e_{r_l}}{|N_{s_l}|} \cdot e_{s_l}^{(0)}, l \geq 2 \quad (13)$$

式中: N_i^l 文章 i 的 l 跳路径的集合; $e_{s_l}^{(0)}$ 是实体 s_l 对应 ID 的嵌入表示。

2.2.4 模型预测

模型通过 L 层对关键词和文章分别进行聚合,得到关键词和文章的多阶嵌入表示,然后对各层的信息进行组合得到最终的关键词和文章的嵌入表示。如式(14)、式(15)所示。最后,在关键词和文章表示上使用内部乘积来预测关键词采用该文章的可能性,如式(16)所示。

$$e_u^* = \sum_{u_u} e_u^{(L)} \quad (14)$$

$$e_i^* = \sum_{\text{所有 } L} e_i^{(L)} \quad (15)$$

$$\text{Predict} = e_u^* \cdot e_i^* \quad (16)$$

3 实验与结果

3.1 电力科研成果推荐数据集

本文通过爬虫技术从知网获取文献数据,通过去重、去噪、缺失值处理等手段对获取的数据进行预处理,最终得到 2012 篇电力领域文献的标题、作者、关键词信息,同时这些文章的标题都满足“基于…的…”的格式。此时构建出的数据集包括 6095 个关键词,2012 篇文章,7669 个实体,3 种实体关系,11577 次关键词—文章交互。经过图 2 所示的数据扩充过程,并从具有特定结构格式的标题中获取研究内容和研究方法信息完善电力科研成果知识图谱后,构建的电力科研成果知识图谱结构如图 1 所示,而最终构建出的电力科研成果推荐数据集信息如表 1 所示。

表 1 电力科研成果推荐数据集

关键词数量	文章数量	关键词-文章交互数量	实体数量	实体关系数量	三元组数量
6095	13982	35828	21934	5	36809

在本次实验中,对关键词文章交互数据进行随

机切割,按照 7 : 3 的比例划分训练集和测试集,最终训练集中包含 28818 条关键词—文章交互数据,测试集中包含 7010 条关键词—文章交互数据。

3.2 对比模型与评价标准

基于电力科研成果推荐数据集通过模型对比实验验证本文提出的方法的有效性,本文所有模型均在训练集上进行训练,在测试集上进行评估。

本文使用的对比模型包括 BPR 模型^[4]、CKE 模型^[15]、KGCN 模型^[17]、KGAT 模型^[18]和 KGIN 模型^[19]。

此外,在评价推荐系统性能的指标中,召回率(Recall)衡量的是模型检索到的相关项目占总相关项目的比例,反映了模型的全面性。归一化折损累计增益(Normalized Discounted Cumulative Gain, NDCG)是用于评估推荐系统质量的一种指标,尤其适用于排名型推荐系统。NDCG 考虑了推荐列表中项目的相关性以及它们在列表中的排名顺序。

在推荐系统中,通常会生成一个推荐列表,其中包含了系统预测用户可能喜欢的项目,并根据一定的排序准则进行排名。Recall@K(K 代表排名前 K 个项目)表示在推荐的前 K 个项目中,模型能够检索到的相关项目的数量占总相关项目的比例,Recall@K 越高,表示模型在前 K 个推荐中覆盖的相关项目越多,反映了推荐结果的全面性。NDCG@K(K 代表排名前 K 个项目)衡量的是前

K 个项目的相关性和排序质量。这两个指标可以更加全面地推荐系统的性能,它们的计算方法如式(17)~式(19)所示。因此,本文设置推荐的排名列表长度 K 为:(20,40,80),并在 top-K 推荐中使用 Recall@K 与 NDCG@K 作为模型性能的评价指标。

$$\text{Recall@K} = \frac{\sum_{u \in N} |\text{predict} \cap N(u)|}{\sum_{u \in N} |N(u)|}$$

(17)

式中,N 是用户集合;predict 为用户 u 的前 K 个推荐结果集合;N(u)是实际与用户 u 存在交互的项目集合。

$$\text{NDCG@K} = \frac{\text{DCG@K}}{\text{IDCG@K}}$$

(18)

$$\text{DCG@K} = \sum_{i=1}^K \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)}$$

(19)

式中,rel_i是第 i 个推荐项的相关性评分,通常是二元的,即相关或不相关;IDCG@K 是理想情况下的 DCG@K 值。

3.3 实验结果

本文中实验参数设置为:优化器固定为 Adam 优化器,批次处理大小固定为 256,学习率设置为 0.0005,嵌入维度设置为 64,KGIN 与本文提出的模型的独立性建模方案都设置为互信息建模方案,GCN 聚集层数为 4 层,用户意图粒度为 12。

最终实验结果如表 2 所示。

表 2 模型性能

模型	K=20		K=40		K=80	
	Recall@K	NDCG@K	Recall@K	NDCG@K	Recall@K	NDCG@K
BPR ^[4]	0.0188	0.0212	0.0524	0.0316	0.1352	0.0564
CKE ^[15]	0.0464	0.0332	0.0948	0.0464	0.1616	0.068
KGCN ^[17]	0.0635	0.0434	0.1134	0.0597	0.1901	0.0819
KGAT ^[18]	0.0528	0.0496	0.0968	0.0620	0.2020	0.0764
KGIN ^[19]	0.0948	0.0632	0.1360	0.0684	0.2552	0.1080
our	0.1245	0.0817	0.2084	0.1089	0.3047	0.1370

表 2 展示了不同模型在电力数据集上的表现,包括本文提出的改进模型(“our”)以及其他几个基线模型,如 KGIN、KGCN、KGAT 等。评估指标为 Recall@K 和 NDCG@K,其中 K 取值分别为 20、40 和 80。

由该表格可以看出,除了本文模型外,KGIN 与 KGCN 的表现最好,这是因为 KGIN 模型将关键词意图建模为关系,并采用关系路径感知聚合,

从而有效捕捉了知识图谱上的丰富信息,为关键词匹配最适合的结果。KGCN 模型利用图卷积网络深入挖掘知识图谱中的高阶关联信息,捕捉实体之间的高阶关联关系,通过聚合邻居节点的信息来丰富文章的表示。最后将本文模型与 KGIN、KGCN 模型进行对比,可以看出我们的模型(“our”)相较于 KGIN 和 KGCN 模型,在所有指标上都表现出了更好的性能。以 K=20 为例,我们的模型在

Recall@20 上达到了 0.1245,明显高于 KGIN 的 0.0948 和 KGCN 的 0.0635。在 NDCG@20 上,我们的模型得分为 0.0817,也优于 KGIN (0.0632)和 KGCN(0.0434)。在 K 值 40 与 80 时,我们模型在 Recall 和 NDCG 两个指标上依然表现最好,领先于 KGIN 和 KGCN。这些结果清楚地表明,在这个电力数据集上,提出的模型相对于 KGIN 模型展现出了更加出色的性能表现。

这是因为本文构建的改进 KGIN 模型融合了意图感知信息、关系感知信息、语义信息和高阶结构信息,在电力科研成果推荐中能够通过对于不同关键词意图的精确捕捉和利用,更好地表达关键词和文章之间的关系,实现关键词和文章更丰富的嵌入表示,获得更好的推荐效果。

4 结 论

本文构建了基于电力关键词、文章、作者等信息的知识图谱与适用于推荐系统的数据集,在此基础上提出了一种融合知识图谱多维信息的电力科研成果推荐算法,依据利用标题信息进行知识增强的电力科研成果知识图谱,构建的改进 KGIN 模型融合了意图感知信息、关系感知信息、语义信息和高阶结构信息,在电力科研成果推荐中能够通过对于不同关键词意图的精确捕捉和利用,实现电力科研成果的有效推荐。通过实验验证,本文所提模型相较于传统的推荐算法和其他基于知识图谱的推荐方法,在推荐效果上表现更好,召回率明显提升,验证了其在电力科研领域的应用潜力和实际价值。

参考文献

[1] 陈钢,常笑,胡枫. 基于隐语义模型的学生选课推荐算法[J]. 计算技术与自动化,2021,40(3):88—93.

[2] 张茂宇,李海明. 融合双重语义相似度的混合协同过滤推荐算法[J]. 计算机工程与设计,2023,44(9):2671—2678.

[3] 陈珊珊,姚苏滨. 基于知识图谱与邻域感知注意力机制的推荐算法研究[J]. 计算机科学,2024,51(8):313—323.

[4] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: bayesian personalized ranking from implicit feedback[J]. International Conference on Uncertainty in Artificial Intelligence, 2012.

[5] KOREN Y, BELL R, VOLINSKY C. Matrix factorization

techniques for recommender systems[J], Computer, 2009, 42(8): 30—37.

[6] 毛骞,谢维成,乔逸天,等. 推荐系统冷启动问题解决方法研究综述[J]. 计算机科学与探索,2024,18(5):1197—1210.

[7] 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, Lyon France, 2018: 1835—1844.

[8] CAO Y, WANG X, HE X, et al. Unifying knowledge graph learning and recommendation: towards a better understanding of user preferences[C]//Proceedings of the 2019 World Wide Web Conference, 2019: 151—161.

[9] WANG X, WANG D, XU C, et al. Explainable reasoning over knowledge graphs for recommendation[C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 5329—5336.

[10] 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.

[11] 柳啸峰,林广艳,于九阳,等. 基于知识图谱的多任务推荐算法[J]. 燕山大学学报,2024,48(4):349—355,376.

[12] JAGVARAL B, LEE W K, ROH J S, et al. Path-based reasoning approach for knowledge graph completion using cnn-bilstm with attention mechanism[J]. Expert Systems with Applications, 2020, 142: 112960.

[13] 娄铮铮,朱军娇,张万闯,等. 用户生成内容场景下角角色导向图神经推荐方法[J]. 计算机学报,2024,47(6):1288—1303.

[14] 周俊,胡斌斌,张志强,等. MoGE:基于图上下文增强的多任务推荐算法[J]. 电子学报,2023,51(11):3377—3387.

[15] 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.

[16] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2015: 2181—2187.

[17] WANG H, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems[C]// Proceedings of the 2019 World Wide Web Conference, 2019: 3307—3313.

[18] 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 and Data Mining, 2019: 950—958.

[19] WANG X, HUANG T, WANG D, et al. Learning intents behind interactions with knowledge graph for recommendation[C]// Proceedings of the 2021 World Wide Web Conference, 2021: 878—887.