

DOI:10.19651/j.cnki.emt.2314114

融合实体与关系交互信息的知识感知推荐模型*

姚静¹ 吕腾²

(1.安徽建筑大学电子与信息工程学院 合肥 230601; 2.安徽新华学院大数据与人工智能学院 合肥 230088)

摘要: 由于知识图谱包含了丰富的项目属性及其关联信息,因此在推荐系统中引入知识图谱能在一定程度上解决数据稀疏和冷启动问题。如基于传播的推荐系统就利用了知识图谱的图结构学习用户及项目表示等相关特征。但在传播过程中,往往忽略了实体与关系之间的交互信息对特征表示的贡献,由此提出一种融合实体与关系交互信息的知识感知推荐模型。首先,将协同信息和知识关联整合,采用异构传播方式传播并扩展用户和项目的表示。其次,在传播过程中用注意力机制强化实体与关系之间的交互信息,增强语义关联,保证用户和项目基于知识的高阶交互的有效性。然后采用知识感知注意力机制来区分每层中实体邻居的重要性,更精确地生成用户和项目的表示。最后通过聚合器将多个表示结合得到用户和项目的最终表示,从而预测用户与项目进行交互的概率。通过添加KL散度损失函数对模型进行优化,以对齐模型的预测分布和真实分布之间的差异。在Last.FM、Book-Crossing和MovieLens-20M 3个数据集上进行的实验结果表明该模型在CTR预测性能中比其他基线模型有较大提升。

关键词: 推荐系统;知识图谱;交互信息;注意力机制

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6099

Knowledge-aware recommendation model fused with interaction information between entities and relations

Yao Jing¹ Lyu Teng²

(1. School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei 230601, China;

2. School of Big Data and Artificial Intelligence, Anhui Xinhua University, Hefei 230088, China)

Abstract: As knowledge graphs contain rich item attributes and their associated information, introducing knowledge graphs into recommendation systems can to some extent solve data sparseness and cold start problems. For example, recommendation systems based on propagation utilize the graph structure of knowledge graphs to learn relevant features such as user and item representations. However, the contribution of interactive information between entities and relationships to feature representation is often ignored in propagation, so this paper proposed a knowledge aware recommendation model that fused with interaction information between entities and relationships. Firstly, collaborative information and knowledge correlation were integrated, and heterogeneous propagation methods were used to propagate and expand the representation of users and items. Secondly, in the process of propagation, attention mechanism was used to strengthen the interaction information between entities and relationships, enhance semantic relevance, and ensure the effectiveness of knowledge-based high-level interaction between users and items. Then, a knowledge aware attention mechanism was used to distinguish the importance of entity's neighbors in each layer, and generate representations of users and items more accurately. Finally, to predict the probability of user interaction with item, multiple representations were combined to obtain the final representation of user and item by an aggregator. To optimize the model, KL divergence loss function was added to align the difference between the prediction distribution and the real distribution of the model. Experimental results on three datasets of Last.FM, Book-Crossing and MovieLens-20M show that the proposed model has a great improvement in CTR prediction performance compared with other baseline models.

Keywords: recommendation system; knowledge graph; interaction information; attention mechanism

0 引言

随着信息技术和互联网的广泛应用,网络中包含的信息

快速增长,带来了信息过载和信息混乱的问题。信息过载使得社交媒体中的信息量远远超过其所能承载的信息量,信息混乱往往使人们在搜索信息时无法找到有用的数据。此外,

收稿日期:2023-07-14

* 基金项目:安徽省高校自然科学基金重点项目(KJ2021A1157)资助

它使人们被不相关和混乱的数据所吸引^[1]。为了解决这些问题,研究人员提出了使用推荐系统来满足用户个性化需求的方法。协同过滤(collaborative filtering, CF)这种传统推荐模型已经得到了广泛的研究与应用,但在实际应用中,协同过滤通常存在着严重的数据稀疏和冷启动问题,导致其推荐效果不太理想。这些问题可以通过将各种辅助信息引入到推荐系统中来得到有效解决,例如项目属性^[2]、社交网络^[3]等。知识图谱(knowledge graph, KG)作为一种新兴的辅助信息,近年来逐渐受到广泛的关注。

知识图谱包含了丰富的项目属性信息和关联信息,在推荐系统中引入知识图谱不仅能有效缓解数据稀疏和冷启动问题,还能提升个性化推荐的准确性。知识图谱是一种基于图的数据结构,由三元组(头实体—关系—尾实体)形式组成。其中包含的信息由实体与关系表示,通过知识图谱提供的关系查找推荐项目可以提高推荐的多样性和可解释性。

知识图谱与推荐系统融合的核心是如何有效地将这些异构信息集成到用户和项目的潜在向量表示中。现有的基于传播的方法通过在知识图谱上迭代传播以找到有用的辅助信息来用于推荐,但该方法在传播过程中往往忽略了实体与关系之间的交互信息对推荐效果的贡献。例如 CKAN^[4](collaborative knowledge-aware attentive network)模型在传播时使用知识感知注意网络来区分不同邻居的重要性,但是没有利用到实体与关系之间的交互信息,使所得的用户和项目嵌入表示不够准确。在知识感知注意网络中结合实体与关系之间的交互信息能增强语义关联,可以更好的捕捉知识图谱中的潜在关系信息,丰富实体的特征表示。因此,本文基于 CKAN 模型首次提出一种融合实体与关系交互信息的知识感知推荐模型(knowledge-aware recommendation model fused with interaction information between entities and relations, KIER)。该模型增强了用户与项目的特征表示,解决了高阶传播过程中知识语义关联性降低的问题,提高了推荐的准确性与有效性。

本文基于 CKAN 模型做出的改进及主要工作如下:

1)对 CKAN 模型中的采用知识感知注意网络区分传播中每层实体邻居重要性的过程,本文模型通过使用注意力机制融合了实体与关系之间的交互信息,增强语义关联,更有效地得到用户和项目的嵌入表示。

2)在推荐模块的损失函数基础上添加了 KL 散度(Kullback-Leibler divergence)损失函数,对齐模型的预测分布和目标的真实分布之间的差异,对模型进行优化。

3)在 3 个不同推荐场景的数据集 Last.FM、Book-Crossing 和 MovieLens-20M 上进行了对比实验,实验结果显示在 CTR 预测任务中本文模型优于其他基线模型。

1 相关工作

1.1 知识图谱

知识图谱是表示来自多个领域的大规模信息的实用方

法^[5]。近年来,知识图谱在搜索引擎^[6]、问答系统^[7]以及词嵌入^[8]等研究领域得到广泛应用,同时也在推荐领域受到了极大的关注。在知识图谱中,节点表示实体,边表示实体之间的关系。其具有很强的表示能力,实体的多个属性可以通过跟踪图谱中不同边来获得,并且可以通过这些关系链接来发现实体的高级关系。如图 1 所示,用户 A 观看并喜欢电影“少年的你”,根据“少年的你”的属性信息可以为用户推荐“曾国祥”导演或“易烊千玺”主演或“爱情”类型的其他电影,比如用户可能喜欢“易烊千玺”主演的电影,则可以为他推荐“送你一朵小红花”和“满江红”,且“送你一朵小红花”与“少年的你”属于同一类型电影,则推荐“送你一朵小红花”给喜欢“少年的你”的用户更容易满足用户需求。

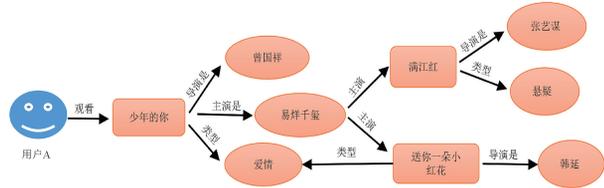


图 1 知识图谱推荐实例

将知识图谱引入推荐系统,通过丰富的项目属性和关系信息可以有效地改善推荐系统的性能。知识图谱推荐算法可以分为 3 类:1)基于嵌入的方法,主要使用知识图谱嵌入(knowledge graph embedding, KGE)^[9]预训练得到的实体嵌入到推荐算法中。例如,CKE^[10](collaborative knowledge base embedding)通过 TransR^[11]在知识图谱的参与下得到实体的向量特征表示。DKN^[12](deep knowledge-aware network)通过 TransD^[13]将实体嵌入和单词嵌入相结合从而对新闻实体建模。然而,这些模型中采用的 KGE 方法更适用于图内应用,如链接预测等,不足以进行项目推荐。2)基于路径的方法,主要是挖掘知识图谱中用户、项目之间多种连接关系以便于更好的进行推荐。例如,MCRec^[14](metapath based context for recommendation)通过学习元路径的显式表示以描述用户-项目对的交互上下文。基于路径的方法可以为推荐的结果提供可解释性,但其严重依赖手动设计的元路径,这在实践中很难优化。3)基于传播的方法,在知识图谱中迭代传播以找到有用的高阶辅助信息。例如,RippleNet^[15](multiple ripples networks)是一种很经典的分层传播模型。RippleNet 通过偏好传播方法很自然的将 KGE 融入到推荐系统中,将用户历史交互视为知识图谱中的种子集,不断自动地发现用户的潜在层级兴趣。KGCN^[16](knowledge graph convolutional networks)融合知识图谱与图卷积神经网络(graph convolutional network, GCN)^[17]模型,在计算知识图谱中某一个给定的实体表示时将邻居信息与偏差结合了起来,有效地提升了推荐模型的表达能力。但是 RippleNet 和 KGCN 模型都忽略了协同信息,导致用户和项目嵌入表示不足。KGAT^[18](knowledge graph attention network)模型

提出了协作知识图(collaborative knowledge graph, CKG), 使用图注意力网络递归地从节点的邻居传播嵌入表示, 以补充实体嵌入。但 KGAT 将用户视为节点, 这就使得一旦有了新的交互用户, 其需要重新训练整个模型来为新用户提出推荐建议。并且 KGAT 将处于不同潜在空间的协同信息与知识图谱信息视为同构节点。CKAN 采用异构传播策略来显式编码协同信息和知识图谱信息, 并通过知识感知注意网络来区分不同实体邻居的贡献, 但是在区分不同实体邻居重要性的过程中没有考虑到实体和关系之间的交互信息对用户和项目表示的贡献。本文模型引入知识图谱, 将协同信息与知识图谱中的知识关联有效地整合起来, 并且融合了实体与关系交互信息, 增强用户和项目节点的表达能力, 提高了推荐的准确性。

1.2 注意力机制

注意力机制^[19]在机器翻译^[20]、自然语言处理^[21]和图像识别^[22]等领域中被广泛应用, 可以提高模型的性能和精度。注意力机制的快速发展让研究者开始考虑将这项技术用于推荐系统。在注意力机制中, 模型会为每个时间步的输入数据赋予一个权重, 表示该输入数据在当前任务中的重要程度。该权重是通过计算输入数据与模型当前状态之间的相似度得到的, 相似度越高, 则权重越大。模型在计算输出结果时就会更加重视与当前任务相关的输入数据, 而忽略掉那些不相关的信息。在推荐系统中, 不同的用户和项目可以通过注意力机制来得到分配的不同权重, 从而获取具有一定代表性的用户对或项目对, 使推荐结果的准确性得到提升。

常见的注意力机制包括自注意力、键值对注意力和多头注意力等。本文使用多头注意力机制来强化实体与关系的交互信息, 增强它们之间的关联, 使获得的用户和项目表示更准确, 并使用知识感知注意力来区分实体邻居重要性。

2 推荐模型

本章介绍本文提出的 KIER 模型, 该模型是一种端端的模型。KIER 将协同信息与知识图谱中的知识关联有效整合起来, 并将实体与关系的高阶交互信息考虑进了推荐模型中。KIER 模型结构如图 2 所示, 模型由异构传播层、知识感知注意力嵌入层和预测层 3 个部分组成。异构传播层将用户由已交互的项目来表示、项目由自身加上与其交互的用户的历史交互项目表示, 通过用户-项目交互和知识图谱中的知识关联传播协作信号。知识感知注意力嵌入层将实体与关系的高阶交互信息融合进来以增强模型的整体性能, 并利用知识感知注意机制来学习同一实体集中实体的知识权重, 并生成实体的加权表示。预测层通过聚合器聚集来自不同传播层的用户和项目的表示, 最后输出预测结果。

2.1 符号说明

假设在某个推荐场景中, 存在 M 个用户 $U = \{u_1,$

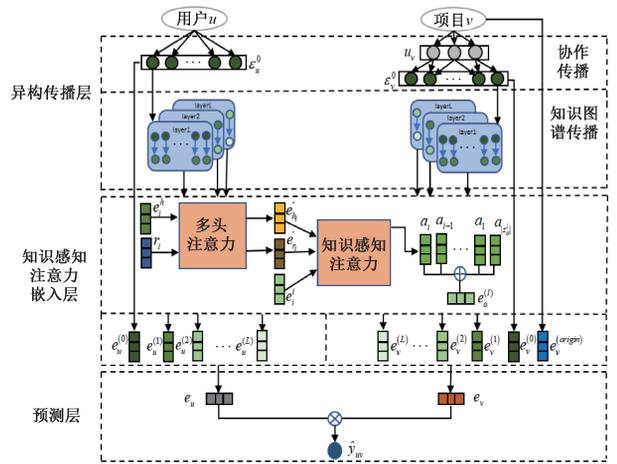


图 2 KIER 模型框架

$u_2, \dots, u_M\}$ 和 N 个项目 $V = \{v_1, v_2, \dots, v_N\}$ 。根据用户历史交互信息, 可以得到用户-项目交互矩阵 $Y \in R^{M \times N}$, 其中, $y_{uv} = 1$ 表示用户 u 与项目 v 发生交互, 否则 $y_{uv} = 0$ 。知识图谱中边的信息用 $G = \{(h, r, t) \mid h, t \in \epsilon, r \in \mathcal{R}\}$ 表示。其中每个三元组 (h, r, t) 表示头实体 h 和尾实体 t 之间存在关系 r , ϵ 和 \mathcal{R} 为知识图谱中实体和关系的集合。使用集合 $\mathcal{A} = \{(v, e) \mid v \in V, e \in \epsilon\}$ 表示项目与实体的对应关系, 其中 (v, e) 表示项目 v 和知识图谱中的实体 e 对应。

2.2 异构传播层

异构传播层由协作传播和知识图谱传播两个模块组成, 如图 2 所示。协作传播明确编码协作信号并生成用户和项目的表示。知识图谱传播沿着知识图谱中的边进行传播知识关联, 从而扩展用户和项目的表示。

1) 协作传播

用户交互过的项目在一定程度上表示用户的偏好。使用用户的历史相关项目来表示用户, 而不使用独立的潜在向量表示。用户 u 历史交互项目集与实体之间的对齐转换可以作为用户 u 在知识图谱上进行传播的初始种子集。用户 u 的历史交互项目集定义如下:

$$u_v^0 = \{v \mid y_{uv} = 1\} \quad (1)$$

将用户 u 历史交互项目集与实体之间对齐得用户 u 的初始实体集, 用户 u 的初始实体集定义如下:

$$\epsilon_u^0 = \{e \mid (v, e) \in \mathcal{A} \text{ and } v \in u_v^0\} \quad (2)$$

与同一项目交互的用户由于其相似的行为偏好可以对项目的特征表示做出贡献。将同一用户交互的项目称为协作邻居, 项目 v 的协作项目集表示公式如下:

$$V_v = \{v_u \mid u \in \{u \mid y_{uv} = 1\} \text{ and } y_{vv_u} = 1\} \quad (3)$$

由此将协作项目集和项目与实体之间的对齐集进行合并可以得到项目 v 的初始实体集, 项目 v 的初始实体集定义如下:

$$\epsilon_v^0 = \{e \mid (v, e) \in \mathcal{A} \text{ and } v_u \in V_v\} \quad (4)$$

其中, 项目 v 的初始实体集还包括直接与其关联的实体。

协作传播层将用户由已交互的项目来表示、项目由自身加上与其交互的用户的历史交互项目表示,这样的好处在于可以很方便的处理新增加的用户,并不需要重新训练新用户的嵌入表示。协作传播层在初始实体集中融合能够最有效的表达潜在语义的一阶交互信息,有效增强了用户和项目的特征表示并提高模型的整体性能。

2)知识图谱传播

知识图谱中相邻实体之间具有强关联性。通过知识图谱的链路传播可以得到扩展的不同层级的实体集与三元组集,从而有效地学习到高阶关联知识和用户偏好。不同传播距离得到的有效信息可以使用户与项目的向量表示更加丰富且准确。在某一层用户 u 和项目 v 实体集的定义递归地表示为:

$$\epsilon_o^l = \{t \mid (h, r, t) \in G \text{ and } h \in \epsilon_o^{l-1}\} \quad l = 1, 2, \dots, L \quad (5)$$

其中, l 为传播层次, o 为占位符, o 可表示用户 u 及项目 v 。用户 u 和项目 v 的第 L 个三元组集定义如下:

$$S_o^l = \{(h, r, t) \mid (h, r, t) \in G \text{ and } h \in \epsilon_o^{l-1}\} \quad l = 1, 2, \dots, L \quad (6)$$

在知识图谱上迭代传播,有效地获取用户与项目的高阶交互信息,并丰富用户与项目的特征表示。

2.3 知识感知注意力嵌入层

在知识图谱中不同的头部实体或关系对应的尾部实体具有不同的含义和潜在的向量表示,因此,在同一层的传播过程中,需要使用注意力机制来有效地区分相邻实体之间的重要程度,得到不同实体的注意力权重。此外,本文不仅限于单纯的从任务层面做特征融合,还加入了强化实体和关系之间的交互信息模块。从实体层面出发,在每一层的传播中融合实体和关系之间的交互信息以增强整个模型推荐效果。突出实体与关系之间的相关性从而增强知识图谱传播的质量,提高知识感知注意力网络的整体性能。

1)实体与关系的交互注意力

本文选择多头注意力机制来计算知识图谱传播中的每一层实体与关系之间的交互相关性。多头注意力机制的结构如图 3 所示。

在多头注意力机制中,每个“头”的注意力机制计算方式相同,但参数不同,所以能够从多个子空间中得到不同方向的信息,从而获得多个维度的特征。多头注意力机制将查询向量 Q 、键向量 K 和值向量 V 进行 n 次不同的线性映射,并且通过缩放点积注意力(如图 4 所示)对 n 个不同的映射结果并行地操作。然后将缩放点积注意力的结果拼接后输入线性映射层,最后得到多头注意力机制的输出结果。缩放点积注意力公式为:

$$Attention(Q, K, V) = \text{soft max}(\frac{QK^T}{\sqrt{d_k}})V \quad (7)$$

其中, d_k 是向量的维度,用来避免内积值过大。

假设在知识图谱传播中的第 l 层的第 i 个三元组为 $(h,$

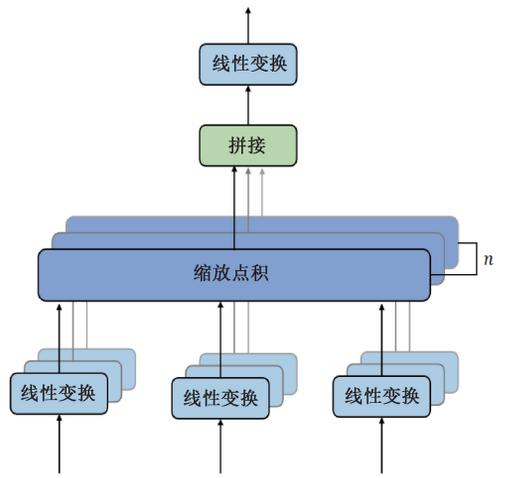


图 3 多头注意力机制结构

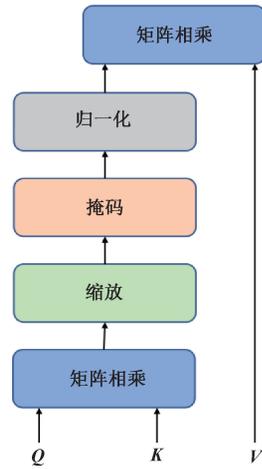


图 4 缩放点积注意力结构

$r, t)$, 将实体和关系进行交互相关性计算分别得到各自每个单头的输出:

$$Head_h^j = Attention(e_i^h, r_i, r_i) = \text{soft max}(\frac{e_i^h r_i^T}{\sqrt{d_k}})r_i \quad j = 1, 2, \dots, n \quad (8)$$

$$Head_r^j = Attention(r_i, e_i^h, e_i^h) = \text{soft max}(\frac{r_i (e_i^h)^T}{\sqrt{d_k}})e_i^h \quad j = 1, 2, \dots, n \quad (9)$$

其中, e_i^h 是第 i 个三元组的头实体的嵌入, r_i 是第 i 个三元组的关系的嵌入。

最后将各自所有单头的输出结果拼接起来,再进行一次线性变换得到多头注意力机制的输出结果:

$$e_{h_i}' = MultiHeadAttention(e_i^h, r_i, r_i) = Concat(Head_h^1, Head_h^2, \dots, Head_h^n)W^O \quad (10)$$

$$e_{r_i}' = MultiHeadAttention(r_i, e_i^h, e_i^h) = Concat(Head_r^1, Head_r^2, \dots, Head_r^n)W^O \quad (11)$$

其中, e_{h_i}' 是用交互相关性加权的实体向量, e_{r_i}' 是用交

互相关性加重的关系向量, W^O 为可训练的权重矩阵。

2) 知识感知注意力

通过将第 l 层的第 i 个三元组的头实体和关系进行交互注意力后得到了加权的头实体和关系向量, 再采用知识感知注意力机制来生成第 i 个三元组的尾实体节点的注意力嵌入向量 a_i , a_i 由尾实体嵌入表示 e'_i 以及注意力网络函数 $\pi(\cdot)$ 计算得到的注意力权重的乘积而来, 定义如下:

$$a_i = \pi(e'_{h_i}, e'_{r_i}) e'_i \quad (12)$$

注意力网络函数 $\pi(\cdot)$ 计算得到尾实体 e'_i 的注意力权重, 注意力网络函数定义如下:

$$\pi(e'_{h_i}, e'_{r_i}) = \sigma(W_2 \text{ReLU}(W_1 z_0 + b_1) + b_2) \quad (13)$$

$$z_0 = \text{ReLU}(W_0(e'_{h_i} \parallel e'_{r_i}) + b_0) \quad (14)$$

注意力网络通过拼接操作将用交互相关性加重的头实体向量 e'_{h_i} 和用交互相关性加重的关系向量 e'_{r_i} 拼接到一起, ReLU 作为非线性激活函数, σ 为 *Sigmoid* 激活函数, W 和 b 是可训练的权重矩阵和偏差。最后使用 *Softmax* 函数归一化三元组的系数, 得到最终使用的注意力网络函数定义如下:

$$\pi(e'_{h_i}, e'_{r_i}) = \frac{\exp(\pi(e'_{h_i}, e'_{r_i}))}{\sum_{(h', r', i') \in S'_o} \exp(\pi(e'_{h_i}, e'_{r_i}))} \quad (15)$$

其中, S'_o 为用户 u 或项目 v 在第 l 层的三元组集合。

最后对第 l 层三元组集中所有尾实体的注意力嵌入表示进行求和, 得到在传播中第 l 层用户 u 或项目 v 的嵌入表示 e'_o , 定义如下:

$$e'_o = \sum_{i=1}^{|S'_o|} a_i \quad l = 1, 2, \dots, L \quad (16)$$

其中, $|S'_o|$ 为第 l 层中三元组的数量。将实体与关系之间的交互信息整合到知识感知注意力网络中而不只是单纯的进行头实体与关系的拼接, 保证了用户和项目基于知识的高阶交互的有效性, 更好的捕捉了知识关联。

由于初始实体集中的实体作为在知识图谱上进行传播的初始种子集, 其最接近于用户和项目的原始表示, 所以初始实体集与原始用户和项目表示具有很强的相关性, 因此将初始实体集的表示加入到用户和项目表示中。用户和项目初始实体集表示定义为初始实体集中实体嵌入表示的平均, 公式如下:

$$e_o^{(0)} = \frac{\sum_{e \in \epsilon_o^0} e}{|\epsilon_o^0|} \quad (17)$$

此外, 项目 v 的原始表示具有相关实体, 而用户则没有。原始相关实体最接近项目 v 本身, 因此将其添加到项目 v 的表示中, 定义如下:

$$e_v^{(origin)} = \frac{\sum_{e \in \{e | (e, v) \in \mathcal{A}\}} e}{|\{e | (e, v) \in \mathcal{A}\}|} \quad (18)$$

在经过异构传播层和知识感知注意力嵌入层之后, 得到的用户 u 和项目 v 的最终表示集如下:

$$\begin{aligned} T_u &= \{e_u^{(0)}, e_u^{(1)}, \dots, e_u^{(L)}\} \\ T_v &= \{e_v^{(origin)}, e_v^{(0)}, e_v^{(1)}, \dots, e_v^{(L)}\} \end{aligned} \quad (19)$$

2.4 预测层

通过异构传播层和知识感知注意嵌入层得到用户和项目的表示集, 选用 *concat* 聚合器进行非线性变换, 将式(19)中的向量表示聚合为用户和项目的单一表示, 如下:

$$agg_{concat}^{(a)} = \sigma(W_a \cdot (e_o^{(i_1)} \parallel e_o^{(i_2)} \parallel \dots \parallel e_o^{(i_n)}) + b_a) \quad (20)$$

其中, $e_o^{(i_k)}$ 是用户和项目表示集 T_o 中的元素, \parallel 表示连接操作。

用户的聚合向量用 e_u 表示, 项目的聚合向量用 e_v 表示。最后, 采用内积方式来预测用户对该项目的偏好得分:

$$\hat{y}_{uv} = e_u^T e_v \quad (21)$$

2.5 损失函数

为了确保模型训练的效果, 将用户与项目进行了交互的为正样本, 从未交互中随机抽取负样本, 使正样本与负样本的大小相同。使用交叉熵损失函数评估推荐模型的效果, 损失函数定义如下:

$$L_{pred} = \sum_{u \in U} \left(\sum_{v \in \{v | (u, v) \in \mathcal{P}^+\}} \mathcal{J}(y_{uv}, \hat{y}_{uv}) - \sum_{v \in \{v | (u, v) \in \mathcal{P}^-\}} \mathcal{J}(y_{uv}, \hat{y}_{uv}) \right) \quad (22)$$

其中, \mathcal{J} 为交叉熵损失函数, \mathcal{P}^+ 表示用户与项目的正交互, \mathcal{P}^- 表示负交互。

再使用 KL 散度损失函数对真实值和预测值之间的分布差异进行约束, 损失函数定义如下:

$$L_{KL} = \|P \cdot T\|_2^2 \quad (23)$$

其中, P 为预测分布, T 为真实分布。

本文提出模型的整体损失函数是预测的交叉熵损失函数与对样本中对齐分布的 KL 散度损失函数加上正则化项, 计算如下:

$$L = c L_{pred} + (1-c) L_{KL} + \lambda \|\Theta\|_2^2 \quad (24)$$

其中, c 为两个损失函数的权重参数, $\Theta = \{E, R, W_i, b_i, \forall i \in \{0, 1, 2, a\}\}$ 是模型的参数集, $\|\Theta\|_2^2$ 是由 λ 控制的 L_2 正则化项。

3 实验与结果分析

在本章中, 通过在 3 个实际应用场景的公开数据集对模型 KIER 进行评估, 并与一些基线模型做对比分析。实验内容包括 1) 通过 CTR 点击率预测任务验证算法效果; 2) 通过实验选择模型的超参数设置; 3) 通过消融实验证明模型中改进模块的重要性。

3.1 数据集介绍

本文实验从音乐、书籍、电影推荐 3 个不同的场景下进行, 将以显式反馈方式组织的数据集 (Last.FM, Book-Crossing, MovieLens-20M) 转换为隐式反馈, 以便于更好的

反映模型性能。数据集的统计信息如表 1 所示。

表 1 实验数据集信息

数据集	Last.FM	Book-Crossing	MovieLens-20M
用户数	1 872	17 860	138 159
项目数	3 846	14 967	16 954
交互数	42 346	139 746	13 501 622
实体数	9 366	77 903	102 569
关系数	60	25	32
三元组数	15 518	151 500	499 474

3.2 实验对照

为了验证 KIER 模型的有效性,本文将其与以下几种基于传播的推荐模型进行对比。

1)RippleNet^[15]:一种经典的分层传播模型。将用户的历史交互信息作为 KG 中传播的种子集,迭代地扩展用户的兴趣以获得用户的层次兴趣,增强用户表示。

2)KGCL^[16]:基于非光谱图卷积网络的传播模型。通过有选择地且有偏差地将邻居信息聚合起来,捕捉用户的潜在兴趣,使推荐的结果具有个性化。

3)KGAT^[18]:在传播过程中使用注意力机制来区分协同知识图谱中邻居信息的重要性,有效地学习高阶关系表示。

4)KGCL^[23]:使用关系感知知识聚合机制来捕获用于项目表示的实体和关系相关的上下文信号,通过跨视图对比学习模式来减小知识图谱的噪声。

5)CKAN^[4]:将交互信息和知识关联视为两个不同空间中的信息,并应用知识感知注意网络来区分不同邻居的贡献。

3.3 参数设置

实验中将向量嵌入表示的维度设置为 64,学习率设置为 0.002,BatchSize 设置为 4096,传播层数设置为 3,多头注意力的头数设为 8,损失函数的权重参数设为 0.7,归一化系数设为 10^{-5} ,用户和项目三元组大小在 {8,16,32,64} 之间选择。使用 Xavier 初始化器来初始化模型参数,使用 Adam 优化模型。将所有数据集按照 6:2:2 的比例划分为训练集、验证集和测试集。

3.4 评价指标

KIER 模型在 CTR 预测任务过程中选择 AUC 与 F1 两个指标来评估模型。

AUC 指标反映模型对样本的排序能力,AUC 的值越接近 1,表明算法效果越好。AUC 的计算公式如下:

$$AUC = \frac{\sum_i^n (pos > neg) + 0.5 \times \sum_i^n (pos = neg)}{N_{pos} \times N_{neg}} \quad (25)$$

其中, n 为样本总数量, pos 为正样本预测得分, neg

为负样本预测得分, N_{pos} 、 N_{neg} 为正样本、负样本数量。

F1 的值是模型准确率与召回率的一种加权平均数,F1 取值为 [0,1],数值越接近 1,表明算法越准确。F1 的计算公式如下:

$$F1 = \frac{2 \times P \times R}{P + R} \quad (26)$$

其中, P 为模型精确率, R 为模型召回率。

3.5 实验结果与对比分析

在 CTR 预测实验中,KIER 模型与其他基线模型在 3 个数据集上的比较结果如表 2 所示。具体的实验结果以及分析如下:

表 2 CTR 预测实验结果

模型	Last.FM		Book-Crossing		MovieLens-20M	
	AUC	F1	AUC	F1	AUC	F1
RippleNet	0.777	0.702	0.721	0.649	0.974	0.927
KGCN	0.805	0.706	0.684	0.633	0.975	0.930
KGAT	0.831	0.742	0.731	0.654	0.976	0.926
KGCL	0.841	0.763	0.751	0.667	0.977	0.932
CKAN	0.842	0.769	0.753	0.673	0.976	0.929
KIER	0.847	0.778	0.757	0.680	0.984	0.953

1)从 CTR 预测任务中的 AUC、F1 指标来看,本文提出的 KIER 模型在 3 个不同数据集上均优于其他基线模型。其中,在 Last.FM 数据集上 AUC 指标提高了 0.6%~9.0%,F1 指标提高了 1.2%~10.8%,在 Book-Crossing 数据集上 AUC 指标提高了 0.5%~10.7%,F1 指标提高了 1.0%~7.4%,在 MovieLens-20M 数据集上 AUC 指标提高了 0.7%~1.0%,F1 指标提高了 2.3%~3.0%。

2)与未使用注意力机制的 KGCN 模型相比,加入注意力机制的 KGAT、CKAN 和 KIER 模型在 3 个数据集上的 AUC 指标上来看均有一定的提升,这表明加入注意力机制的推荐模型能够有效学习传播中用户和项目的特征表示。

3)与忽略了协同信息的 RippleNet 模型相比,CKAN 和 KIER 模型在性能上的表现证明将协同信息和知识关联整合起来可以有效地得到用户、项目的高阶信息,提升推荐的准确度。

4)与将用户作为节点,且将处于不同潜在空间的协同信息与知识图谱信息视为同构节点的 KGAT 模型相比,CKAN 和 KIER 模型将用户用与其交互过的项目来表示,将协同信息与知识关联视为两个不同空间中的信息,有效地提高了推荐的性能。

5)与忽略了实体与关系之间的交互信息的 KGCL 模型和 CKAN 模型相比,KIER 模型的优势在于传播过程中融合了强化实体和关系之间的交互信息模块,更精确地学习用户及项目表示,使得模型得到了优化,性能得以提升。

3.6 实验参数影响分析

本节将对 KIER 模型使用的一些超参数进行研究。

1)多头注意力头数。实验在 3 个数据集上设计了不同注意力头数来验证注意力头数对 KIER 模型性能的影响。实验结果如图 5 所示。

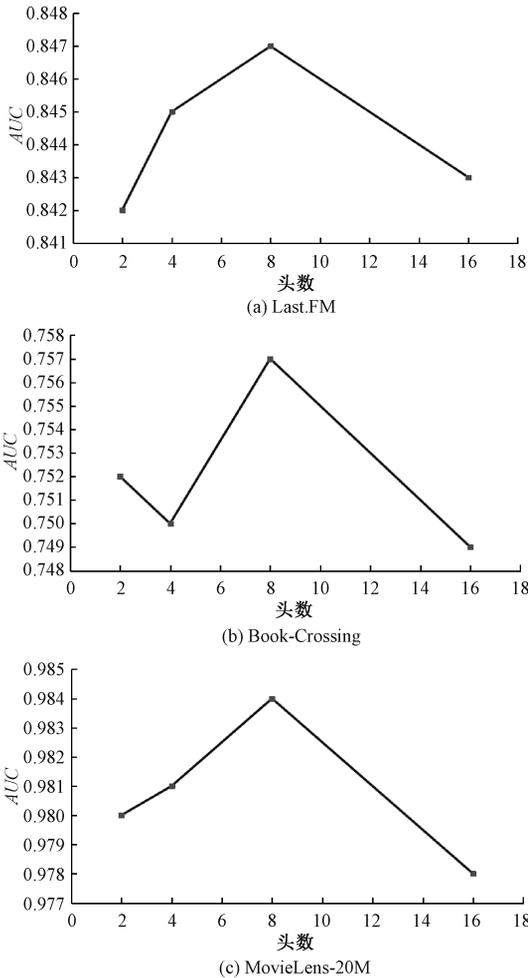


图 5 不同注意力头数的 AUC 结果

由图 5 可知,适当增加注意力头数可以提升模型性能,因为多个注意力头可以考虑到更多层面的信息,能够关注到不同特征组合的重要性。当头数为 8 时 AUC 达到最高值,继续增加头数后模型性能呈现下降趋势,因为头数越多,与真实目标间的累计误差越大。并且随着注意力头数增加,参数量也增加,会导致模型训练的时间成本加重。故本实验中将注意力头数设置为 8。

2)损失函数的权重参数。实验通过改变损失函数的权重参数大小来分析 KIER 模型性能的变化,实验结果如表 3 所示。

表 3 不同权重参数的 AUC 结果

权重参数	Last.FM	Book-Crossing	MovieLens-20M
0.6	0.835	0.700	0.982
0.7	0.847	0.757	0.984
0.8	0.841	0.706	0.983
0.9	0.843	0.707	0.983

由表 3 可以看出损失函数的权重参数为 0.7 时模型的性能最好,权值太小会导致权重不能提供足够的约束,而权值过高会干扰目标函数。故本实验中所设置的损失函数的权重参数为 0.7。

3.7 模型分析

本文以 CTR 预测任务的 AUC 值作为消融实验的评估指标,讨论强化实体和关系之间的交互信息模块和 KL 散度损失函数在模型中的有效性。KIER/I 为去除强化实体与关系之间交互信息模块的模型,直接进行实体知识感知注意力计算。KIER/L 为去除 KL 散度损失函数的模型,只使用针对模型预测的交叉熵损失函数。消融实验结果如表 4 所示,从表中可知:

1)对比 KIER/I 模型与 KIER 模型实验结果,可知添加了强化实体与关系之间交互信息的模块可以提升推荐效果。其原因是在知识感知注意网络中将实体与关系的交互信息融合起来可以保证高阶交互的有效性,更好地实现用户和项目的特征表示,提高推荐的准确率。

2)对比 KIER/L 模型与 KIER 模型实验结果,可知使用了 KL 散度损失函数的模型的推荐效果更好。其原因是 KL 散度损失函数可以对齐模型的预测分布和目标的真实分布之间的差异,对模型起到优化作用。

表 4 消融实验结果

模型	KIER/I	KIER/L	KIER
Last.FM	0.841	0.841	0.847
Book-Crossing	0.709	0.710	0.757
MovieLens-20M	0.983	0.983	0.984

4 结 论

本文提出了一种融合实体与关系交互信息的知识感知推荐模型,在异构传播层中进行协作传播和知识图谱传播,逐层传播用户和项目的表示。在传播过程中使用注意力机制强化实体与关系之间的交互信息,增强语义关联,提高用户与项目高阶表示的准确性。使用知识感知注意力区分每层实体邻居的重要性,聚合邻居节点表示,得到更丰富的用户与项目表示。并添加了 KL 散度损失函数对模型进行了优化,提升推荐效果。在 3 个公共数据集 (Last.FM, Book-Crossing, MovieLens-20M) 上与 RippleNet, KGCN, KGAT 和 CKAN 模型进行对比实验,本文模型的推荐效果有较大提升。在未来,计划进一步评估模型在更多数据集上的可行性,并将作为辅助信息的用户社交网络结合到本文的模型中,增强推荐模型的准确率与可解释性。

参考文献

- [1] YANG X, HUAN Z, ZHAI Y, et al. Research of personalized recommendation technology based on

- knowledge graphs [J]. Applied Sciences, 2021, 11(15): 7104.
- [2] 宣鹏程, 唐彦, 王汪送. 基于项目类型的群组推荐方法[J]. 电子测量技术, 2019, 42(7): 54-58.
- [3] DENG S, HUANG L, XU G, et al. On deep learning for trust-aware recommendations in social networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(5): 1164-1177.
- [4] WANG Z, LIN G, TAN H, et al. CKAN: Collaborative knowledge-aware attentive network for recommender systems [C]. Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, 2020: 219-228.
- [5] EHRLINGER L, WÖB W. Towards a definition of knowledge graphs[J]. SEMANTiCS(Posters, Demos, SuCCESS), 2016, 48(1-4): 2.
- [6] UYAR A, ALIYU F M. Evaluating search features of Google Knowledge Graph and Bing Satori: Entity types, list searches and query interfaces[J]. Online Information Review, 2015, 39(2): 197-213.
- [7] ZHANG Y, DAI H, KOZAREVA Z, et al. Variational reasoning for question answering with knowledge graph [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [8] SIMOV K, OSENOVA P, POPOV A. Comparison of word embeddings from different knowledge graphs[C]. Language, Data, and Knowledge: First International Conference, LDK 2017, Galway, Ireland, June 19-20, 2017, Proceedings 1. Springer International Publishing, 2017: 213-221.
- [9] WANG Q, MAO Z, 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] HU B, SHI C, ZHAO W X, et al. Leveraging meta-path based context for top-n recommendation with a neural co-attention model[C]. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018: 1531-1540.
- [15] 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.
- [16] WANG H, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems[C]. The World Wide Web Conference, 2019: 3307-3313.
- [17] YING R, HE R, CHEN K, 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, 2018: 974-983.
- [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 & Data Mining, 2019: 950-958.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, DOI: 10.48550/arXiv.1706.03762.
- [20] BROWN P F, COCKE J, DELLA PIETRA S A, et al. A statistical approach to machine translation[J]. Computational Linguistics, 1990, 16(2): 79-85.
- [21] CHOWDHARY K R, CHOWDHARY K R. Natural language processing [J]. Fundamentals of Artificial Intelligence, 2020: 603-649.
- [22] WU M, CHEN L. Image recognition based on deep learning[C]. 2015 Chinese Automation Congress (CAC), IEEE, 2015: 542-546.
- [23] YANG Y, HUANG C, XIA L H, et al. Knowledge graph contrastive learning for recommendation [C]. Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 1434-1443.

作者简介

姚静, 硕士研究生, 主要研究方向为数据处理和数据挖掘。

E-mail: 3476514399@qq.com

吕腾(通信作者), 博士, 教授, 主要研究方向为数据库与数据处理。

E-mail: LT0410@163.com