文章编号:1671-6833(2025)06-0015-08

融合多模态信息的知识感知推荐方法

王海荣[1,2], 王怡梦[1], 周北京[1], 易之航[1]

(1.北方民族大学 计算机科学与工程学院,宁夏 银川 750021;2.北方民族大学 图像图形智能处理国家民委重点实验室, 宁夏 银川 750021)

摘 要:图片、文本等多模态信息具有语义互补性,能够有效增强知识图谱中的实体表示,从而提高推荐的准确率和可解释性。通过分析推荐系统中具有语义相关性的多模态数据特点,提出了一种融合多模态信息的知识感知推荐方法。在知识图谱传播的基础上,整合与图谱中实体语义相关的多模态信息,并将其与对应的实体进行特征融合,用来丰富实体表示,以便探索用户潜在的兴趣偏好。该方法充分考虑了多模态信息间的依赖性和交互性,采用模态间注意力关注各模态的重要信息,获取具有语义关联的多模态嵌入特征;通过门控注意力将实体对应的多模态嵌入特征与实体表示融合,进一步丰富实体的多模态语义信息,从而增强用户和项目的表示。为了验证方法的有效性,在 MovieLens-1M 和 Book-Crossing 数据集上进行实验,并与 RippletNet、KGAT、CKAN、LKGR、COAT、CKE、KGCN、SKGCR 和 KGCL 这 9 种方法进行对比分析,实验结果表明:所提方法在 *AUC* 和 *ACC* 上均优于对比方法;在 MovieLens-1M 和 Book-Crossing 数据集上,所提方法的 *AUC* 分别为 0.936 6 和 0.763 7,与其他模型的平均值相比,增幅为 0.027 2 和 0.029 1;所提方法的 *ACC* 分别为 0.862 3 和 0.708 9,与其他模型的平均值相比,增幅为 0.028 3 和 0.030 5。

关键词:知识图谱; 推荐系统; 多模态信息; 特征融合; 嵌入传播

中图分类号: TP391;TP18;TN912 **文献标志码**:A **doi**:10.13705/j.issn.1671-6833.2025.03.010

推荐系统(recommendation system, RS)中普遍使用的协同过滤(collaborative filtering, CF)[1]方法根据用户的共同兴趣偏好进行推荐,但存在着用户与项目交互数据稀疏和冷启动等问题[2]。为此,研究人员通常利用辅助信息来解决上述问题,其中知识图谱(knowledge graph, KG)因其富含大量语义信息和结构信息,可以有效辅助提升推荐准确率、增强推荐的可解释性[3-4]而受到人们关注。现有的基于 KG 的推荐系统大致可以分为基于嵌入的方法[5]、基于路径的方法[6]和基于传播的方法[7-9]。

基于传播的方法同时结合基于嵌入方法和基于路径方法的优点,已经成为基于 KG 的推荐系统中的主流方法。其不仅采用嵌入传播机制,同时充分利用 KG 中的语义信息和结构信息,沿着 KG 中的关系路径传播并聚合多跳的邻居信息,从而增强用户和项目的表示[10],典型的方法有 RippleNet[11]、KGAT[12]、CKAN[13]、LKGR[14]等。但上述的基于 KG 的推荐方法只专注于对 KG 中包含知识的挖掘,没有考虑多模态信息的语义知识可能包含着用户的个性偏好[15]。例如,用户在点餐时可能会因为菜品图片或美食文字介绍对菜品产生兴趣。将多模态信息[16]引入 KG 中,可以丰富图谱中实体的知识表示,有助于实现更精准的个性化推荐。MKGCN[17]提取了音乐相关的 7 种模态信息,通过使用多模态聚合器处理各模态信息的融合,增强实体表示,提高推荐性能。MKGAT[18]将多模态信息经过预处理后直接作为实体并引入新的关系整合进 KG 中,采用改进后的图注意力网络来聚合多模态邻居实体,并扩展到多跳来获取多模态知识图谱的高阶信息。

将多模态信息引入 KG 可以增强图谱中的实体表示,从而提升推荐性能,但上述的多模态的方法仍存在着多模态语义信息挖掘不完全和融合各模态信息时引入大量噪声等问题。多模态语义信息挖掘不完全是指未能充分挖掘多模态信息所具有的语义互补性和语义关联性,这些信息对于探究用户的深层兴趣偏好有重要意义。融合各模态信息时引入大量

收稿日期:2025-01-06;**修订日期**:2025-02-19

基金项目:宁夏自然科学基金资助项目(2023AAC03316);宁夏回族自治区教育厅高等学校科学研究重点项目(NYG2022051)

作者简介:王海荣(1977—),女,宁夏石嘴山人,北方民族大学教授,博士,主要从事大数据知识工程与智能信息处理的研究,E-mail:bmdwhr@163.com。

引用本文:王海荣,王怡梦,周北京,等. 融合多模态信息的知识感知推荐方法[J]. 郑州大学学报(工学版),2025,46(6):15-22. (WANG H R, WANG Y M, ZHOU B J, et al. Knowledge-aware recommendation method integrating multi-modal information[J]. Journal of Zhengzhou University (Engineering Science),2025,46(6):15-22.)

噪声问题是指在进行模态信息融合时未将各模态信息中的无用信息或干扰信息进行丢弃,而是直接融合进特征表示中,影响推荐的准确性。为此,本文提出了一种融合多模态信息的知识感知推荐方法(knowledge-aware recommendation method integrating multi-modal information,KRIM),在 KG 传播的基础上,整合与 KG 中实体相关的多模态信息,分析多模态数据特点,采用模态间注意力关注各模态交互信息的细粒度融合,挖掘多模态数据的深层语义信息;同时设计门控注意力来动态控制多模态信息与图谱中实体信息之间的数据融合,减少融合过程中的噪声数据干扰,进而丰富实体表示,提高推荐性能。

1 KRIM 方法架构

KRIM 将用户和项目交互图、KG 和多模态信息作为输入,首先,通过异构传播得到具有协同信号的多跳实体嵌入集,采用模态间注意力融合不同模态数据之间的语义信息,获得多模态嵌入特征集;其次,利用门控注意力将实体嵌入与多模态嵌入进行融合,在多跳实体嵌入集上沿着 KG 中的关系进行多跳传播来捕获高阶邻居信息,得到最终的嵌入表示;最后,通过内积计算最终的用户表示和项目表示,输出用户对项目的偏好得分,KRIM 模型结构如图 1 所示。

KRIM 由异构传播、模态间信息融合、模态外信息融合与传播、预测与学习 4 个模块构成。异构传播模块旨在将协同信号与图谱的结构信息相结合,获取反映用户偏好的多阶实体集。模态间信息融合模块使用预处理模型提取各模态的特征,采用模态间注意力整合多模态特征,实现多模态语义信息的深层挖掘。模态外信息融合与传播模块使用门控注意力来减少实体信息和多模态信息融合时的噪声数据影响,并通过知识感知注意力多跳传播聚合邻居实体信息,捕获图谱的高阶语义信息。预测与学习模块主要实现用户偏好的预测。

1.1 异构传播

异构传播模块由交互传播和知识传播两部分组成。交互传播考虑将协作信号显式地编码到用户和项目表示中,挖掘用户潜在偏好。知识传播是实体沿着 KG 中的关系路径多跳传播挖掘图谱中的语义信息和结构信息。

1.1.1 交互传播

点击同一个项目的不同用户之间可能存在着相同的兴趣偏好,这种现象称之为协作信号。通过交互传播可以捕捉用户和项目之间的协作信号,根据用户的历史交互数据构建用户和项目交互图,若用户和项目存在交互,则它们之间有连接,反之,则没有。对于用户 u,通过在用户-项目交互图上进行一次传播(即用户-项目),获得用户的初始种子集 S_u^0:

$$S_u^0 = \{e \mid (i,e) \in E\}, i \in \{i \mid y_{ui} = 1\}。 \quad (1)$$

式中:E 为 KG 中的实体集;$y_{ui} = 1$ 指用户和项目之间有交互,否则 y_{ui} 值为 0。类似地,项目端也考虑协作信号,在交互图上进行二次传播(即项目-用户-项目),获得项目的初始种子集 S_i^0:

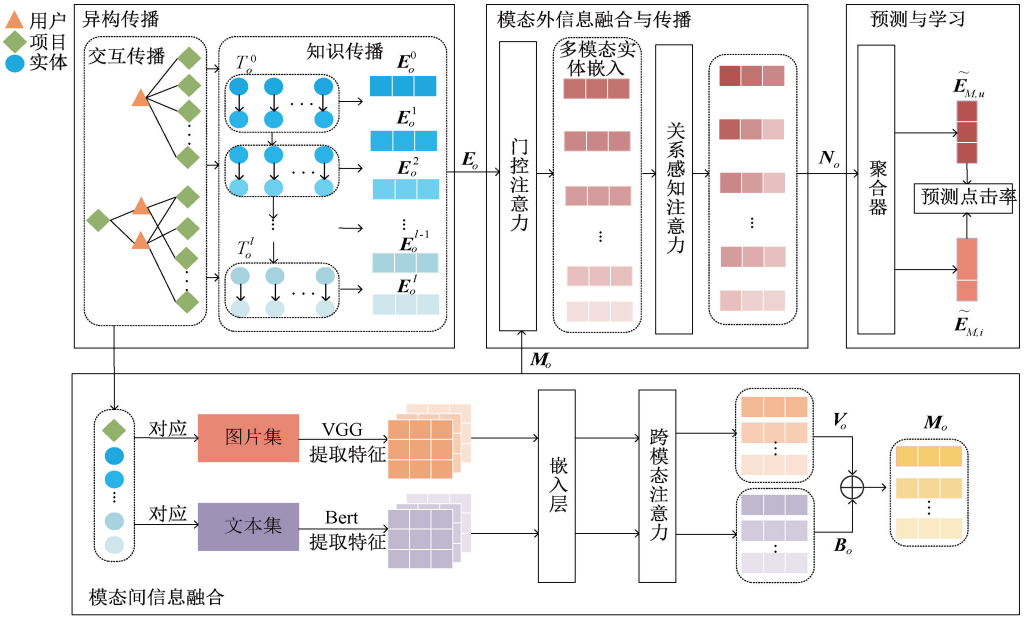

图 1 KRIM 模型结构图

Figure 1 KRIM model architecture diagram

$$I_i = \{i_u \mid \{u \mid y_{ui} = 1\}, y_{ui_u} = 1\}; \quad (2)$$

$$S_i^0 = \{e \mid (i_u, e) \in E\}, i_u \in I_i。 \quad (3)$$

式中:I_i 为项目-用户-项目所得到的项目集。初始种子集会通过知识传播获取高阶的实体集。

1.1.2 知识传播

知识图谱中富含大量的语义信息和结构信息,可以增强用户表示和项目表示。知识传播是通过用户初始种子集和项目初始种子集在图谱上沿着不同关系路径进行多跳传播来扩展用户实体集和项目实体集。获取每一跳的实体表示:

$$S_o^l = \{t \mid (h, r, t) \in G\}, h \in S_o^{l-1}。 \quad (4)$$

式中:l 表示与初始实体集之间存在 l 跳的距离,$l = 1, 2, \cdots, L$;下标符号 o 为用户 u 或者项目 i 的占位符。定义用户 u 和项目 i 的第 l 个三元组:

$$T_o^l = \{(h, r, t) \mid (h, r, t) \in G\}, \quad h \in S_o^{l-1}, l = 1, 2, \cdots, L。 \quad (5)$$

扩展后的用户实体集和项目实体集进行嵌入,得到用户和项目的多跳实体嵌入集,将送入模态外信息融合与传播模块进行下一步处理。

1.2 模态间信息融合

多模态数据由图片、文本、音频等不同类型模态数据组成,富含多模态数据特有的语义信息和用户潜在的个性偏好。模态间信息融合模块处理多模态信息,设计模态间注意力来融合各模态信息,挖掘多模态数据中的深层语义信息。多模态数据由图片和文本两个模态数据进行表示,也可扩展到不同模态。对获取的项目图片通过 VGG16[19] 进行预处理并经过嵌入层,得到项目图片嵌入特征。文本信息通过 BERT[20] 进行预处理并经过嵌入层,得到项目文本嵌入特征。

首先,将图片嵌入特征和文本嵌入特征映射到同一空间中:

$$\begin{cases} \boldsymbol{V}_i^m = \boldsymbol{W}_v \cdot \boldsymbol{V}_i; \\ \boldsymbol{B}_i^m = \boldsymbol{W}_r \cdot \boldsymbol{B}_i。 \end{cases} \quad (6)$$

式中:$\boldsymbol{V}_i^m$ 和 $\boldsymbol{B}_i^m$ 分别为图片和文本的映射矩阵;$\boldsymbol{W}_v$ 和 $\boldsymbol{W}_r$ 均为可以学习到的参数矩阵。

其次,计算不同模态之间交互信息的重要性:

$$a_{tt} = \mathrm{Softmax}(\boldsymbol{B}_i^m \cdot \boldsymbol{V}_i^{m\mathrm{T}})。 \quad (7)$$

式中:a_{tt} 为模态间注意力得分。采用 Softmax 函数进行归一化,计算加权后的图片嵌入特征 $\boldsymbol{V}_i$ 和文本嵌入特征 $\boldsymbol{B}_i$:

$$\boldsymbol{V}_i = a_{tt} \cdot \boldsymbol{V}_i^m, \boldsymbol{B}_i = a_{tt} \cdot \boldsymbol{B}_i^m。 \quad (8)$$

最后,二者相加得到多模态嵌入特征 $\boldsymbol{M}_i$,将送入模态外信息融合与传播模块进行下一步处理。

$$\boldsymbol{M}_i = \boldsymbol{V}_i + \boldsymbol{B}_i。 \quad (9)$$

1.3 模态外信息融合与传播

模态外信息融合与传播模块由模态外信息融合和多模态实体高阶传播组成。模态外信息融合主要针对实体与多模态的特征融合,多模态实体高阶传播聚合不同关系权重的高阶邻居信息。

1.3.1 模态外信息融合

设计门控注意力将多模态嵌入特征与实体嵌入进行融合,门控机制可以自适应地融合不同特征,关注重要信息,过滤噪声数据的干扰。门控信号保留有用信息,去除无用信息,计算如下:

$$g_i = \sigma(\boldsymbol{W}_g(\mathrm{Concat}(\boldsymbol{E}_i, \boldsymbol{M}_i)) + \boldsymbol{b}_g)。 \quad (10)$$

式中:g_i 为门控信号;$\boldsymbol{E}_i$ 和 $\boldsymbol{M}_i$ 分别为实体嵌入和其实体相关的多模态嵌入特征;σ 表示 Sigmoid 激活函数;$\boldsymbol{W}_g$ 和 $\boldsymbol{b}_g$ 分别为可训练的权重矩阵和偏差。使用门控信号来控制不同特征的权重,以便有选择地关注更重要的特征,注意力得分为

$$a_i^g = g_i \cdot \tanh(\boldsymbol{W}_a(\mathrm{Concat}(\boldsymbol{E}_i, \boldsymbol{M}_i)) + \boldsymbol{b}_a)。 \quad (11)$$

式中:$\boldsymbol{W}_a$ 和 $\boldsymbol{b}_a$ 分别为可训练的权重矩阵和偏差。

计算加权后的特征向量:

$$\begin{cases} \boldsymbol{E}'_i = a_i^g \cdot \boldsymbol{E}_i; \\ \boldsymbol{M}'_i = (1 - a_i^g) \cdot \boldsymbol{M}_i。 \end{cases} \quad (12)$$

式中:$\boldsymbol{E}'_i$ 和 $\boldsymbol{M}'_i$ 分别为加权后的实体嵌入和多模态嵌入特征。将加权后的特征进行相加,得到多模态实体嵌入 $\boldsymbol{E}_{M,i}$:

$$\boldsymbol{E}_{M,i} = \boldsymbol{E}'_i + \boldsymbol{M}'_i。 \quad (13)$$

多模态实体嵌入 $\boldsymbol{E}_{M,i}$ 可以通过多模态实体高阶传播进行处理获取高阶信息。

1.3.2 多模态实体高阶传播

聚合实体的高阶信息可以探究用户的深层兴趣偏好,挖掘用户和项目之间的潜在关系。考虑每个多模态尾实体在不同多模态头实体和关系中具有不同的含义,采用知识感知注意力[13] 来捕捉邻居的语义信息。设第 l 层邻居中,第 i 个三元组知识感知注意力 $\boldsymbol{Z}_n$ 嵌入,计算公式如下:

$$\boldsymbol{Z}_n = a(\boldsymbol{E}_h^n, \boldsymbol{R}^n) \cdot \boldsymbol{E}_t^n。 \quad (14)$$

式中:$\boldsymbol{E}_h^n$ 为第 n 个多模态头实体嵌入;$\boldsymbol{E}_t^n$ 为第 n 个多模态尾实体嵌入;$\boldsymbol{R}^n$ 为第 n 个多模态头实体和第 n 个多模态尾实体之间的关系;$a(\boldsymbol{E}_h^n, \boldsymbol{R}^n)$ 表示多模态头实体和关系产生的注意力权重,表达式为

$$a(\boldsymbol{E}_h^n, \boldsymbol{R}^n) = \sigma(\boldsymbol{W}_1 \mathrm{ReLU}(\boldsymbol{W}_2 \mathrm{ReLU}(\boldsymbol{W}_3 \cdot \mathrm{Concat}(\boldsymbol{E}_h^n, \boldsymbol{R}^n) + \boldsymbol{b}_3) + \boldsymbol{b}_2) + \boldsymbol{b}_1)。 \quad (15)$$

式中:σ 为 Sigmoid 激活函数;ReLU 为非线性激活函数;$\boldsymbol{W}_1, \boldsymbol{W}_2, \boldsymbol{W}_3$ 为可训练的权重矩阵;$\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$

为偏差。此外,采用 Softmax 函数对知识感知注意力权重进行归一化,以防止梯度爆炸问题,计算如下:

$$a(\boldsymbol{E}_h^n,\boldsymbol{R}^n)=\frac{\exp(a(\boldsymbol{E}_h^n,\boldsymbol{R}^n))}{\sum_{(h',r',t')\in T_i^l}\exp(a(\boldsymbol{E}_{h'}^n,\boldsymbol{R}^{n'}))}。\quad(16)$$

式中:T_i^l 为项目的 l 层邻居,获得项目的第 l 层邻居表示 $\boldsymbol{E}_{M,i}^l$。

$$\boldsymbol{E}_{M,i}^l=\sum_{n=1}^{|T_i^l|}\boldsymbol{Z}_n,l=1,2,\cdots,L。\quad(17)$$

获得项目 l 层的表示集:

$$\boldsymbol{N}_i=(\boldsymbol{E}_{M,i}^0,\boldsymbol{E}_{M,i}^1,\cdots,\boldsymbol{E}_{M,i}^l)。\quad(18)$$

用户端采用相同的计算方式,获取用户 l 层的表示集:

$$\boldsymbol{N}_u=(\boldsymbol{E}_{M,u}^0,\boldsymbol{E}_{M,u}^1,\cdots,\boldsymbol{E}_{M,u}^l)。\quad(19)$$

将用户 l 层的表示集和项目 l 层的表示集送入预测与学习模块进行下一步处理。

1.4 预测与学习

为了预测用户的个性化偏好和优化模型。在知识图谱上进行多次迭代传播后,得到用户 l 层的表示集 $\boldsymbol{N}_u$ 和项目 l 层的表示集 $\boldsymbol{N}_i$,采用 Sum 聚合器聚合每层的嵌入表示,计算如下:

$$\widetilde{\boldsymbol{E}}_{M,o}=\sigma(\boldsymbol{W}_s\cdot(\boldsymbol{E}_{M,o}^0+\boldsymbol{E}_{M,o}^1+\cdots+\boldsymbol{E}_{M,o}^l)+\boldsymbol{b}_s)。\quad(20)$$

式中:o 为占位符,表示用户或项目,得到最终的用户表示 $\widetilde{\boldsymbol{E}}_{M,u}$ 和项目表示 $\widetilde{\boldsymbol{E}}_{M,i}$。使用内积来预测用户对项目的偏好分数:

$$\hat{y}_{ui}=\sigma(\widetilde{\boldsymbol{E}}_{M,u}^{\mathrm{T}}\widetilde{\boldsymbol{E}}_{M,i})。\quad(21)$$

定义损失函数[21]进行模型优化,公式如下:

$$L=\sum_{(u,i,j)\in O}-\ln\sigma(\hat{y}_{ui}-\hat{y}_{uj})+\lambda\|\theta\|_2^2。\quad(22)$$

式中:$O=\{(u,i,j)\mid(u,i)\in O^+,(u,j)\in O^-\}$ 表示训练集;σ 为 Sigmoid 函数;θ 为模型参数集;λ 为学习率。

2 实验及结果分析

2.1 实验设计

为了评估 KRIM 方法的有效性,在 Movielens-1M 和 Book-Crossing 两个数据集上进行实验。Movielens-1M 数据集中包含用户与项目的交互信息、电影相关的知识图谱信息、图片信息和文本信息,其中图片信息是指电影的封面信息,文本信息是指电影的文本简介内容。Book-Crossing 数据集中包含用户与项目的交互信息、书籍相关的知识图谱信息、图片信息和文本信息,其中图片信息是指书籍的封面信息、文本信息是指书籍的作者、出版社等文本内容。具体的数据集统计信息如表 1 所示。

由表 1 可知,MovieLens-1M 数据集中用户和项目交互相对稠密,KG 实体丰富,但多模态信息相对较少。Book-Crossing 数据集中用户和项目交互相对稀疏,KG 实体相对较少,但多模态信息相对较多。实验选取用户交互稀疏程度不同、KG 中实体数量不同、多模态信息多少不同的两个数据集对本方法性能进行验证,可以体现方法在不同数据集上的有效性。

实验中训练集、验证集和测试集数据比例设置为 6∶2∶2,采用 Adam 优化器进行模型优化,使用 Xavier 初始化模型参数。

参数设置值均为在 Movielens-1M 和 Book-Crossing 两个数据集中实验性能最佳时的数值。其中,两个数据集的嵌入维度均为 64,正则化损失权重 $L2$ 均为 10^{-5},学习率 λ 均为 0.002。对于实体传播层数、用户的邻居数、项目的邻居数的参数设置,在两个数据集上有所不同,因 Movielen-1M 数据集中的实体更为丰富,实体传播层数设为 2 层能达到更好的性能,Book-Crossing 选取 3 层能达到最佳性能。用户的邻居数、项目的邻居数在 Movielens-1M 数据集上分别设置为 128 和 256,在 Book-Crossing 数据集上为 256 和 32,参数选取为多次实验中性能最佳的数值。

2.2 方法性能分析

为验证本文方法的有效性,与 RippletNet、KGCN、CKAN 等 9 种主流方法进行对比分析,采用 *AUC* 和 *ACC* 作为点击率(CTR)的性能评价指标。具体的性能对比结果如表 2 所示。

由表 2 可知,本文方法在两个数据集上都具有较优的性能。在 MovieLens-1M 和 Book-Crossing 数据集上,KRIM 方法的 *AUC* 分别为 0.936 6 和 0.763 7,与其他模型的平均值相比,增幅为 0.027 2 和 0.029 1;

表 1 数据集信息统计

Table 1 Statistics of datasets

数据集	用户数	项目数	交互数	交互密度	实体数	关系数	三元组数	图片数	文本数
MovieLens-1M	6 036	2 445	753 772	0.051 00	182 011	12	1 241 996	3 256	2 444
Book-Crossing	17 860	14 967	139 746	0.000 52	77 903	25	151 500	11 882	14 967

表 2　不同方法在数据集上的实验结果

Table 2　Experimental results of the method on the different datasets

方法	MovieLens-1M		Book-Crossing	
	AUC	*ACC*	*AUC*	*ACC*
COAT[22]	0.892 2	0.811 2	0.740 7	0.655 9
CKE[23]	0.898 0	0.825 0	0.713 0	0.646 0
KGCN[24]	0.900 1	0.823 0	0.734 1	0.693 2
RippletNet[10]	0.900 5	0.822 9	0.726 3	0.643 2
KGAT[11]	0.910 2	0.838 9	0.719 4	0.689 9
CKAN[12]	0.917 0	0.844 0	0.745 3	0.697 3
SKGCR[25]	—	—	0.733 0	—
KGCL[26]	0.928 0	0.853 0	0.747 0	0.700 0
LKGR[13]	0.929 0	0.854 0	0.753 0	0.702 0
KRIM	**0.936 6**	**0.862 3**	**0.763 7**	**0.708 9**

注:表中横线意味着相应的实验结果未显示。

与 LKGR 方法相比,增幅分别为 0.007 6 和 0.010 7。在 MovieLens-1M 和 Book-Crossing 数据集上,KRIM 方法的 *ACC* 分别为 0.862 3 和 0.708 9,与其他模型的平均值相比,增幅为 0.028 3 和 0.030 5;与 LKGR 方法相比,增幅分别为 0.008 3 和 0.006 9。结果表明,将多模态信息引入 KG 中,能够补充实体的语义信息,增强其表示,可以细粒度地捕捉用户潜在的兴趣偏好,提高推荐性能。而且 KRIM 采用的模态间注意力关注多模态数据的交互信息,可以挖掘各模态数据的深层语义信息,使用的门控注意力也可以降低融合与实体相关的多模态信息时的噪声数据干扰。仔细观察实验结果可以发现,本文方法在 Book-Crossing 数据集上性能提升相对更加明显。原因是 Book-Crossing 数据集相较 MovieLens-1M 数据集用户和项目交互更稀疏,KG 中的实体也相对较少,但实体相关的多模态信息多。本文方法通过多模态信息来丰富实体表示,增强实体的语义信息,可以缓解用户和项目交互少和 KG 中实体稀疏的问题,性能提升明显。

KRIM 模型的计算复杂度主要来自知识高阶传播、模态间注意力和门控注意力 3 个部分。知识高阶传播是指知识图谱中的实体根据图谱中的关系路径进行高阶知识传播的过程,其计算复杂度为 $O(2G_n dL)$,其中:L 为传播层数;G_n 为知识图谱的三元组样本数;d 为特征嵌入维度。模态间注意力是针对多模态信息之间的特征融合,其计算复杂度为 $O(PTd)$,其中:P 为图片特征样本数;T 为文本特征样本数;d 为特征嵌入维度。门控注意力是为了融合实体信息和多模态信息,其计算复杂度为 $O(MNd)$,其中:M 为融合后的多模态特征样本数;N 为与之对应的实体特征样本数;d 为特征嵌入维度。因此,KRIM 模型的总体计算复杂度为 $O(2G_n dL+PTd+MNd)$。基于上述分析,在相同的实验设置下,KRIM 虽然因为引入多模态信息导致计算成本有所增加,但仍在大多数推荐方法所允许的计算复杂度范围之内。

2.3　消融实验

为了证明引入多模态信息和多模态融合模块的有效性以及对于推荐性能的影响,进行了消融实验。消融实验结果如表 3 所示。w/o M 表示去除多模态信息,只保留 KG 信息;w/o MF 表示消去模态间信息融合模块,即在融合多模态信息时只使用拼接操作;w/o FP 表示消去模态外信息融合与传播模块,即只使用拼接操作融合多模态信息和实体信息,不进行高阶信息传播。

表 3　消融实验结果

Table 3　Ablation experiment results

方法	MovieLens-1M		Book-Crossing	
	AUC	*ACC*	*AUC*	*ACC*
w/o M	0.930 1	0.856 8	0.703 5	0.631 5
w/o MF	0.933 3	0.858 6	0.732 9	0.676 0
w/o FP	0.930 6	0.857 0	0.729 8	0.673 3
KRIM	0.936 6	0.862 3	0.763 7	0.708 9

从表 3 的实验结果可知,w/o M、w/o MF 和 w/o FP 在两个数据集上性能下降明显,证明了利用多模态信息来增强相关实体表示的必要性,也证明模态间信息融合模块中所采用的模态间注意力的有效性,模态外信息融合与传播模块中使用的门控注意力以及高阶传播的有效性。在 Book-Crossing 数据集上 w/o M、w/o MF 和 w/o FP 性能降低尤其明显,分析原因是 Book-Crossing 数据集相较于 MovieLens-1M 数据集,KG 中实体较少同时用户和项目的交互稀疏。w/o M 去除多模态信息,仅通过图谱中的实体信息来增强表示,实验性能下降,这表明了利用多模态信息可以有效增强稀疏 KG 中实体的表示和缓解用户交互少的问题。Book-Crossing 数据集中多模态信息多,w/o MF 性能下降,分析原因是只使用拼接操作是无法挖掘多模态信息之间的语义关联性和互补性的,没有探究用户的多模态潜在兴趣偏好,导致推荐性能下降。w/o FP 性能下降明显,说明去除门控注意力影响多模态信息与实体信息的有效融合,会将各模态数据中的噪声数据引入特征表示中,同时不进行高阶传播则无法聚合高阶的邻居信息,导致特征表示不充分,影响推荐的准确性。

综上,通过消融实验证明了添加多模态信息和

方法中所设计的融合模块对于模型的重要性,去除任意技术和模块都会导致推荐性能的下降。

2.4 融合信息的噪声分析实验

融合信息的噪声问题是指多模态信息和实体信息融合时容易将一些不必要的噪声数据引入融合后特征中。实际上,多模态信息与实体信息本身就含有噪声数据,在实验中所使用的 Movielen-1M 数据集和 Book-Crossing 数据集中,多模态的噪声数据包含图片噪声和文本噪声。其中,图片噪声是指图片中与主体无关的背景信息、图片模糊等,文本噪声是指在文本信息中会存在的拼写错误、错误表述等。这些噪声数据会影响特征表示,降低推荐的准确性。本文方法 KRIM 可以减少多模态信息与知识图谱实体信息融合时的噪声数据影响。

为验证不同方法在减少噪声数据干扰的有效性,进行融合信息的噪声分析实验。通过 KRIM 与未采用噪声处理机制的基线方法 MKGAT、MKGCN 以及采用不同噪声处理策略的 KRIM 变体方法 KRIM-MH、KRIM-SA、KRIM-CA 这 5 种方法进行对比。KRIM-MH 是采用多头注意力来融合多模态信息和实体信息;KRIM-SA 是采用自注意力来处理多模态信息和实体信息的融合问题;KRIM-CA 是使用跨模态注意力来融合多模态信息和实体信息。具体结果如表 4 所示。

表 4 噪声实验结果

Table 4 Noise experiment results

方法	MovieLens-1M		Book-Crossing	
	AUC	*ACC*	*AUC*	*ACC*
MKGAT	0.917 3	0.839 8	0.733 6	0.668 9
MKGCN	0.912 0	0.836 2	0.730 4	0.667 1
KRIM-MH	0.930 0	0.855 2	0.742 6	0.683 6
KRIM-SA	0.927 9	0.851 3	0.740 9	0.681 0
KRIM-CA	0.932 7	0.856 9	0.751 3	0.692 9
KRIM	0.936 6	0.862 3	0.763 7	0.708 9

由表 4 可知,KRIM 的性能结果最佳,证明本文方法可以有效减少多模态信息与知识图谱实体信息融合时的噪声数据影响。通过与未采用噪声处理机制的基线方法 MKGAT、MKGCN 对比,KRIM 的性能增加明显,证明了噪声处理机制的重要性。与 KRIM-MH、KRIM-SA、KRIM-CA 对比,KRIM 方法性能始终优于上述变体方法,证明在处理信息融合所产生的噪声问题时,本文方法所采用的门控注意力机制更有优势。

2.5 不同用户数和项目数实验

为了探索用户和项目在知识图谱中聚合邻居信息时不同的邻居数对方法性能的影响,在 MovieLens-1M 和 Book-Crossing 数据集上采用不同的用户邻居数和项目邻居数进行多次实验,在 MovieLens-1M 数据集中的实验结果如图 2 所示。在 Book-Crossing 数据集中的实验结果如图 3 所示。

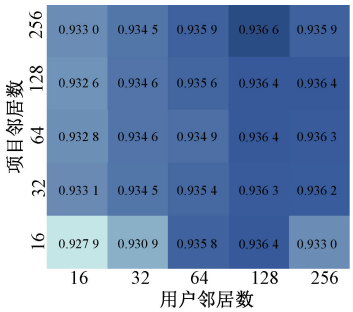

图 2 在 MovieLens-1M 中的 *AUC* 结果

Figure 2 *AUC* results in MovieLens-1M

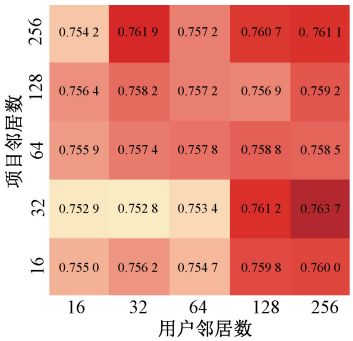

图 3 在 Book-Crossing 中的 *AUC* 结果

Figure 3 *AUC* results in Book-Crossing

分析图 2 可知,当用户邻居数取 128、项目邻居数取 256 时,*AUC* 获得最佳值 0.936 6。原因可能是 MovieLens-1M 数据集中多模态信息较少,所以较多邻居数可以获取更多的多模态信息来丰富表示。但当邻居数过多时会引入噪声,导致推荐性能反而下降。

由图 3 可以看出,当用户邻居数和项目邻居数分别设置为 256、32 时,*AUC* 将达到最佳值 0.763 7。分析原因是当邻居数越大时,模型能捕捉到更多的 KG 实体信息和多模态信息,这将丰富用户和项目的表示,同时提高推荐性能。但设置较大邻居数将会引入更多的噪声数据,从而导致性能下降。因此,在设置用户数和项目数时要考虑丰富的邻居信息和可能会同时产生的干扰数据。

2.6 聚合器性能分析

为了探索不同聚合器对方法性能的影响,采用 Sum、Pool 和 Concat 聚合器进行多次实验。聚合器性能对比实验结果如表 5 所示。

分析表 5 结果可知,Sum 聚合器始终优于 Concat 聚合器和 Pool 聚合器,原因在于 Sum 聚合将各阶邻居信息进行累加来得到最终的用户表示和项目

表 5　不同聚合器实验结果

Table 5　Experimental results of different aggregators

聚合器	MovieLens-1M		Book-Crossing	
	AUC	*ACC*	*AUC*	*ACC*
Concat	0.932 4	0.859 8	0.753 5	0.705 0
Pool	0.927 4	0.852 5	0.731 4	0.660 7
Sum	0.936 6	0.862 3	0.763 7	0.708 9

表示,邻居信息保留完整。Concat 聚合器虽然也将各阶邻居信息进行拼接来保留邻居信息,但同时可能会引入一些噪声干扰,而 Sum 聚合器对噪声数据不敏感。Pool 聚合器采用最大池化聚合的邻居信息导致严重缺少,性能不佳。综上,相较于 Concat 聚合器和 Pool 聚合器,Sum 聚合器能够更好地捕获各阶的信息,推荐性能最优。

3　结论

本文提出的 KRIM 方法在 KG 传播的基础上集成与 KG 中实体语义相关的多模态信息,并将其与对应的实体进行特征融合,丰富实体表示,进而增强用户和项目的表示。

从本文方法的实验结果可以发现,利用多模态信息来增强实体表示从而提升推荐性能是有效的,但如何更好地对多模态数据进行降噪处理,同时获取更多有用信息来丰富表示是需要解决的难题。此外,对多模态信息所隐含的用户偏好的挖掘还处于单一层面。针对上述问题,未来可以考虑分层提取多模态特征,挖掘多模态信息中包含的浅层特征和深层特征,并通过模态注意力等融合技术减少融合过程中所产生的大量噪声问题。

参考文献:

[1] GUO H, YANG C Y, ZHOU L Q, et al. A novel knowledge graph recommendation algorithm based on graph convolutional network[J]. Connection Science, 2024, 36(1): 2327441.

[2] LIU T Y, SHEN H J, CHANG L, et al. Iterative heterogeneous graph learning for knowledge graph-based recommendation[J]. Scientific Reports, 2023, 13(1): 6987.

[3] LI D Z, QU H B, WANG J Q. A survey on knowledge graph-based recommender systems[C]//2023 China Automation Congress (CAC). Piscataway: IEEE, 2023: 2925-2930.

[4] GAO C, ZHENG Y, LI N, et al. A survey of graph neural networks for recommender systems: challenges, methods, and directions[J]. ACM Transactions on Recommender Systems, 2023, 1(1): 1-51.

[5] WANG H W, ZHANG F Z, XIE X, et al. DKN[C]// Proceedings of the 2018 World Wide Web Conference. New York: ACM, 2018: 1835-1844.

[6] HU B 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. New York: ACM, 2018: 1531-1540.

[7] WANG F, LI Y S, ZHANG Y J, et al. KLGCN: knowledge graph-aware light graph convolutional network for recommender systems[J]. Expert Systems with Applications, 2022, 195: 116513.

[8] RUAN S Q, YANG C, LI D S. Knowledge-enhanced personalized hierarchical attention network for sequential recommendation[J]. World Wide Web, 2024, 27(1): 2.

[9] CHEN F K, YIN G S, DONG Y X, et al. KHGCN: knowledge-enhanced recommendation with hierarchical graph capsule network[J]. Entropy, 2023, 25(4): 697.

[10] TAO S H, QIU R H, CAO Y, et al. Intent with knowledge-aware multiview contrastive learning for recommendation[J]. Complex & Intelligent Systems, 2024, 10(1): 1349-1363.

[11] WANG H W, ZHANG F Z, WANG J L, 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. New York: ACM, 2018: 417-426.

[12] 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. New York: ACM, 2019: 950-958.

[13] WANG Z, LIN G Y, TAN H B, et al. CKAN[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 219-228.

[14] CHEN Y K, YANG M L, ZHANG Y X, et al. Modeling scale-free graphs with hyperbolic geometry for knowledge-aware recommendation[C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2022: 94-102.

[15] 王海荣, 徐玺, 王彤, 等. 多模态命名实体识别方法研究进展[J]. 郑州大学学报(工学版), 2024, 45(2): 60-71.
WANG H R, XU X, WANG T, et al. Research progress of multimodal named entity recognition[J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(2): 60-71.

[16] BAI H Y, WU L, HOU M, et al. Multimodality invariant learning for multimedia-based new item recommendation[C]//Proceedings of the 47th International ACM SIGIR

Conference on Research and Development in Information Retrieval. New York: ACM, 2024: 677-686.
[17] CUI X H, QU X L, LI D M, et al. MKGCN: multi-modal knowledge graph convolutional network for music recommender systems[J]. Electronics, 2023, 12(12): 2688.
[18] SUN R, CAO X Z, ZHAO Y, et al. Multi-modal knowledge graphs for recommender systems[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 1405-1414.
[19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10)[2024-06-11]. https://arxiv.org/abs/1409.1556v6.
[20] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2019-05-24)[2024-06-11]. https://arxiv.org/abs/1810.04805v2.
[21] MA T, HUANG L T, LU Q Q, et al. KR-GCN: knowledge-aware reasoning with graph convolution network for explainable recommendation[J]. ACM Transactions on Information Systems, 2023, 41(1): 1-27.
[22] DAI Q Y, WU X M, FAN L, et al. Personalized knowledge-aware recommendation with collaborative and attentive graph convolutional networks[J]. Pattern Recognition, 2022, 128: 108628.
[23] 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. New York: ACM, 2016: 353-362.
[24] WANG H W, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems[C]//The World Wide Web Conference. New York: ACM, 2019: 3307-3313.
[25] LIU X K, YANG B, XU J Y. SKGCR: self-supervision enhanced knowledge-aware graph collaborative recommendation[J]. Applied Intelligence, 2023, 53(17): 19872-19891.
[26] YANG Y H, 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. New York: ACM, 2022: 1434-1443.

Knowledge-aware Recommendation Method Integrating Multi-modal Information

WANG Hairong[1,2], WANG Yimeng[1], ZHOU Beijing[1], YI Zhihang[1]

(1. College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China; 2. The Key Laboratory of Images & Graphics Intelligent Processing of State Ethnic Affairs Commission, North Minzu University, Yinchuan 750021, China)

Abstract: It is found that multi-modal information such as images and text possesses semantic complementarity, which could effectively enhance the representation of entities in knowledge graphs, thereby improving the accuracy and interpretability of recommendations. A knowledge-aware recommendation method that could integrate multimodal information was proposed by analyzing the characteristics of semantically related multimodal data in recommendation systems. On the basis of knowledge graph propagation, multi-modal information that was semantically related to entities in the graph was integrated, and feature fusion was performed with the corresponding entities to enrich entity representation, aiming to explore users' potential interest preferences. In this method, the dependency and interactivity between multimodal information was considered, intermodal attention was adopted to focus on important information of each modality, and semantically associated multimodal embedding features were obtained. Through gated attention, the multi-modal embedding features corresponding to entities were fused with entity representations, further enriching the multi-modal semantic information of entities, thereby enhancing the representation of users and items. In order to verify the effectiveness of the method, experiments were conducted on MovieLens-1M and Book-Crossing data sets, and comparative analysis was conducted with 9 methods including RippletNet, KGAT, CKAN, LKGR, COAT, CKE, KGCN, SKGCR and KGCL. The experimental results showed that it was better than the other two indicators in *AUC* and *ACC*. On the MovieLens-1M and Book-Crossing datasets, the *AUC* of the proposed method were 0.936 6 and 0.763 7, respectively, with an increase of 0.027 2 and 0.029 1 compared to the average values of other models. The *ACC* values of the proposed methods were 0.862 3 and 0.708 9, respectively, with an increase of 0.028 3 and 0.030 5 compared to the average values of other models.

Keywords: knowledge graph; recommendation system; multi-modal information; feature fusionl; embedding propagation