软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(4):1590−1603 [doi: 10.13328/j.cnki.jos.007213] [CSTR: 32375.14.jos.007213]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

融合任务知识的多模态知识图谱补全*

陈 强, 张 栋, 李寿山, 周国栋

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

通信作者: 张栋, E-mail: dzhang@suda.edu.cn

摘 要: 知识图谱补全任务旨在根据已有的事实三元组 (头实体、关系、尾实体) 来挖掘知识图谱中缺失的事实三元组. 现有的研究工作主要致力于利用知识图谱中的结构信息来进行知识图谱补全任务. 然而, 这些工作忽略了知识图谱中蕴含的其他模态的信息也可能对知识图谱补全有帮助. 并且, 由于基于特定任务的知识通常没有被注入通用的预训练模型, 因而如何在抽取模态信息的过程中融合任务的相关知识变得至关重要. 此外, 因为不同模态特征对于知识图谱补全的贡献不一样, 所以如何有效地保留有用的多模态信息也是一大挑战. 为了解决上述问题, 提出一种融合任务知识的多模态知识图谱补全方法. 利用在当前任务上微调过的多模态编码器, 来获取不同模态下的实体向量表示. 并且, 通过一个基于循环神经网络的模态融合过滤模块, 去除与任务无关的多模态特征. 最后, 利用同构图网络表征并更新所有特征, 从而有效地完成多模态知识图谱补全任务. 实验结果表明, 所提出的方法能有效地抽取不同模态的信息, 并且能够通过进一步的多模态过滤融合来增强实体的表征能力, 进而提高多模态知识图谱补全任务的性能.

关键词: 知识图谱补全; 多模态; 知识融合; 多模态融合

中图法分类号: TP18

中文引用格式: 陈强, 张栋, 李寿山, 周国栋. 融合任务知识的多模态知识图谱补全. 软件学报, 2025, 36(4): 1590–1603. http://www.jos.org.cn/1000-9825/7213.htm

英文引用格式: Chen Q, Zhang D, Li SS, Zhou GD. Task Knowledge Fusion for Multimodal Knowledge Graph Completion. Ruan Jian Xue Bao/Journal of Software, 2025, 36(4): 1590–1603 (in Chinese). http://www.jos.org.cn/1000-9825/7213.htm

Task Knowledge Fusion for Multimodal Knowledge Graph Completion

CHEN Qiang, ZHANG Dong, LI Shou-Shan, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: The task of completing knowledge graphs aims to reveal the missing fact triples within the knowledge graph based on existing fact triples (head entity, relation, tail entity). Existing research primarily focuses on utilizing the structural information within the knowledge graph. However, these efforts overlook that other modal information contained within the knowledge graph may also be helpful for knowledge graph completion. In addition, since task-specific knowledge is typically not integrated into general pre-training models, the process of incorporating task-related knowledge into modal information extraction becomes crucial. Moreover, given that different modal features contribute uniquely to knowledge graph completion, effectively preserving useful multimodal information poses a significant challenge. To address these issues, this study proposes a multimodal knowledge graph completion method that incorporates task knowledge. It utilizes a fine-tuned multimodal encoder tailored to the current task to acquire entity vector representations across different modalities. Subsequently, a modal fusion-filtering module based on recurrent neural networks is utilized to eliminate task-independent multimodal features. Finally, the study utilizes a simple isomorphic graph network to represent and update all features, thus effectively accomplishing multimodal knowledge graph completion. Experimental results demonstrate the effectiveness of our approach in extracting information from different modalities. Furthermore, it shows that our method enhances entity representation capability through additional multimodal filtering and fusion, consequently improving the performance of multimodal knowledge graph completion tasks.

* 基金项目: 国家自然科学基金 (62206193, 62076176, 62076175)
收稿时间: 2023-08-25; 修改时间: 2023-11-03; 采用时间: 2024-04-16; jos 在线出版时间: 2024-07-03
CNKI 网络首发时间: 2024-07-05

Key words: knowledge graph completion (KGC); multimodal; knowledge fusion; multimodal fusion

知识图谱, 如 FreeBase[1]和 WordNet[2], 已被广泛应用到人工智能领域中, 包括智能问答[3]、信息抽取[4]、推荐系统[5]等, 因而其重要性不言而喻. 众所周知, 知识图谱中结构化的知识通常表示为事实三元组 (头实体、 关系、尾实体). 但是, 现存的知识图谱通常是稀疏的, 并且大多数是不完整的 (两个实体之间缺少对应的关系). 因此, 在知识图谱研究领域, 知识图谱补全 (knowledge graph completion, KGC) 是一项非常重要的工作. KGC 任务是利用现有知识图谱中的事实三元组, 将实体和关系的表示映射到低维向量表示空间中, 通过实体向量和关系向量来评估每个预测出的三元组的合理性.

然而, 以往大部分的研究[6−8]都专注于使用知识图谱的结构信息进行建模, 通过知识图谱中的有效事实三元组来不断更新随机初始化的实体向量和关系向量. 实际上, 这些方法都忽视了知识图谱中与实体相关的语义信息, 例如实体对应的文本描述和图像信息. 如图 1 所示, 单纯地通过已有的三元组信息很难判断实体“Joe Biden”和“Donald John Trump”之间的关系, 但是根据图片中两者类似的背景、年龄、装扮, 以及文本描述中重复的相关词汇“46th”“45th”“president of the United States”, 我们可以很容易预测出两者存在“竞争”关系. 受此启发, 我们可以利用图像和描述信息来建模实体的多模态向量表示, 从而进一步提升多模态知识图谱补全的完整度和正确度.

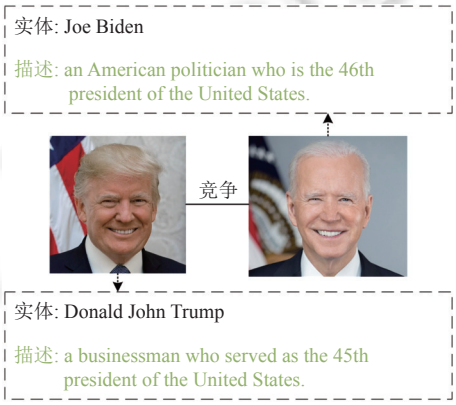


图 1 实体的多模态信息表示

虽然已有一些研究[9−11]开始考虑将文本模态信息或者图像模态信息应用到知识图谱补全任务中, 但是他们依然存在一些缺陷. 一方面, 这些方法在预抽取文本和图像特征的过程中使用的是两个通用且独立的文本和图像编码器. 在预抽取特征过程中, 通用的特征编码器无法自动适应知识图谱补全任务的特征特点, 并且无法很好地综合多模态特征, 进而导致各模态之间的特征无法根据任务进行动态调整. 另一方面, 由于知识图谱中实体的数量很大, 在匹配相关图片的过程中, 很容易引入过多不相关信息, 造成辅助知识图谱补全的信息质量下降, 也影响到知识图谱补全任务的发展.

因此, 为了解决上述挑战, 本文提出了一种融合任务知识的多模态知识图谱补全方法. 具体而言, 首先本文使用一个基于多模态知识图谱补全 (multimodal knowledge graph completion, MMKGC) 任务的多模态编码器来获取实体的文本向量表示和图像向量表示. 通过这种方式, 与知识图谱补全任务相关的信息已经被融合进文本和图像的向量表示中. 其次为了进一步增强结构向量、文本向量和图像向量的表征能力, 本文使用一个基于循环神经网络的多模态融合过滤模块, 保留与任务相关的多模态特征信息.

在此基础上, 为了更好地使用图卷积网络[12]来聚合邻居节点和边的信息, 本文使用多种不引入额外参数的聚合算子来聚合邻居节点和边, 从而将异构图网络转换为同构图网络, 简化图网络结构. 接着使用图卷积算法更新节点的多模态融合向量表示, 从而将知识图谱三元组之间的结构信息融入到实体的向量表示当中.

实验结果表明, 本文提出的融合任务知识的多模态知识图谱补全方法在性能上要显著优于最先进的基线方法, 包括一些基于结构特征的知识图谱补全方法和一些最新的基于多模态特征的知识图谱补全方法. 其中, 在指

标 *MRR* (mean reciprocal rank) 上, 与基于结构特征的知识图谱补全方法中表现最优的方法 ConvE 相比, 我们方法的性能在两个数据集上分别提升了 3.6% 和 1.0%; 在基于多模态特征的知识图谱补全方法中, 与在 FB15k-237 数据集上表现最优的方法 MKGformer 和在 WN9 数据集上表现最优的方法 VisualBERT 相比, 我们的方法的性能分别提升了 1.7% 和 32.6%.

综上所述, 本文的主要贡献总结如下.

(1) 探索了一种融合任务知识的多模态编码方式, 来适应多模态知识图谱补全任务中多模态向量表示.

(2) 提出一种基于循环神经网络的模态融合过滤方法, 来提纯对任务有效的不同模态的信息.

(3) 本文在两个公开的多模态知识图谱数据集上进行了实验. 实验结果表明本文提出的方法在多个指标上都大幅优于基准方法.

本文第 1 节对知识图谱补全的相关工作进行概述, 并系统地介绍与之对应的计算方法. 第 2 节介绍本文构建的融合任务知识的多模态知识图谱补全模型. 第 3 节通过主要的对比实验, 验证所提出的模型相比以前的研究工作展现出明显的优势. 第 4 节给出进一步的分析与讨论, 来说明我们所提出方法的重要性和有效性. 第 5 节针对我们的贡献, 总结全文.

1 相关工作

与本文的研究最相关的工作是基于结构信息的知识图谱向量表示和基于模态信息的知识图谱向量表示, 以及图神经网络表示学习, 具体内容如下.

1.1 基于结构信息的知识图谱向量表示

近期, 研究者们提出了各种各样的知识图谱向量表示方法. 这些方法通过随机初始化实体和关系的结构向量, 利用知识图谱中事实三元组来得到实体和关系的最终向量表示. 根据实体和关系的交互方式, 这些方法大致可以分为 3 种: 1) 基于距离[6,13]的方法. 此类方法通过比较实体和关系的距离来衡量事实三元组存在的合理性. 例如, TransE[6]基于向量空间存在的平移不变性, 将实体和关系映射到同一向量空间中, 采用头实体向量和关系向量相加的方式来为事实三元组打分. 2) 基于语义匹配[7,8]的方法. 此类方法通过匹配实体和关系的潜在语义关系来衡量事实三元组存在的合理性. 例如, ComplEx[8]在复数空间中向量化实体和关系, 对反对称关系进行建模. 3) 基于卷积神经网络[14,15]的方法. 此类方法通过卷积神经网络引入较少的参数来学习实体和关系深层次的特征, 对复杂关系进行建模. 例如, ConvE[14]首次使用二维卷积来完成知识图谱补全任务.

但是, 上述方法仅使用知识图谱的结构信息来得到实体和关系的向量表示, 几乎都忽略了知识图谱中与实体相关的其他模态信息对实体和关系向量表示的影响.

1.2 基于模态信息的知识图谱向量表示

对于多模态知识图谱, 知识图谱中的其他模态信息, 包括以实体和关系描述为代表的文本信息和以实体图像为代表的图像信息, 直观来说, 也需要被纳入考虑. 例如, KG-BERT[9]将三元组结构信息与文本描述信息视为一个序列, 将知识图谱补全任务转化为一个序列分类任务. MKGformer[16]在 KG-BERT 的基础上增加了实体的图像信息和文本图像融合模块来进一步挖掘不同模态之间的有效信息, 并且使用掩码语言模型来预测缺失实体的类别. RSME[10]则使用 3 个门控模块来去除无用的图像信息. 而 MANS[17]则通过探索不同的负样本采样方法来增强向量表示的鲁棒性.

但是, 这就面临两个问题: 一方面, 这些方法将知识图谱中的事实三元组独立为一个个样本, 并没有考虑到知识图谱中事实三元组之间的网络结构, 在训练过程中并没有利用知识图谱中隐藏的结构信息. 另一方面, 它们大多数使用通用型预训练模型来抽取文本和图像特征, 并利用一个线性网络将不同模态的特征映射到与结构信息相同的向量表示空间, 而这样就缺少微调这一关键步骤, 忽略了与知识图谱补全任务相关知识的作用, 导致预抽取的多模态特征可能并不能很好适应多模态知识图谱补全任务.

1.3 图卷积神经网络表示学习

随着深度学习技术的发展, 图卷积神经网络的关系型建模能力引起了越来越多的研究者的注意, 在许多领域

都得到了广泛的应用, 比如视觉问答[18]、对话中情感检测[19]和文本表示学习[20]. 但是, 上述任务都是在同构图网络上应用图卷积神经网络, 他们只需要考虑节点的不同, 不需要考虑边的类别. 而知识图作为一种多关系的异构图结构, 包括不同的实体节点和不同类型的边. 传统的图卷积神经网络无法有效聚合异构图中边的关系, 因此Schlichtkrull 等人[21]提出了关系图卷积神经网络, 通过枚举每一个实体和关系的组合来表示知识图中可能存在的(节点, 边) 的类型, 从而将异构图网络转换为同构图网络, 如图 2 所示. 图 2(a) 和图 2(b) 的差异在于边的类型的个数, 同构图网络的边类型个数为 1, 异构图网络的边类型个数大于 1. 但这么做会导致随着关系个数的增加, 参数量也因而急剧上升, 这就限制了其在大规模知识图上的应用. Vashishth 等人[22]放弃了枚举实体和关系组合的方法, 采用聚合算子的方式将边的信息融合到实体当中, 也成功地将原来的多种关系类型转换为一种关系类型, 实现了异构图向同构图的转化.

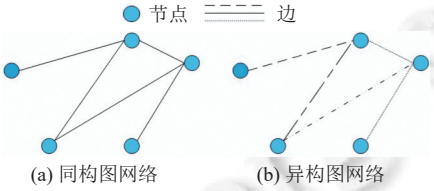


图 2 同构图网络和异构图网络

受到以上工作的启发, 为了更好地融合各个模态之间的信息, 并且利用知识图谱这一天然的图网络结构, 本文提出了一种基于图卷积网络的融合任务知识的多模态知识图谱补全模型, 从融合任务知识编码和利用图网络结构信息两个方面来增强多模态实体关系的向量表示.

2 方 法

本节将详细介绍本文提出的基于图网络结构的融合任务知识的多模态知识图谱补全方法, 模型整体框架如图 3 所示. 模型首先使用一个多模态编码器进行文本和图像的多模态特征提取. 并且, 结合循环网络结构保留对任务有用的多模态特征信息. 然后, 使用图卷积神经网络对知识图谱进行实体和关系信息的传播和聚合, 得到最终的实体和关系的向量表示. 接着, 使用被广泛使用的解码器来对给定实体和关系的三元组进行特征融合. 最后, 将最终特征输入分类器预测新的三元组, 从而补全原来的图谱.

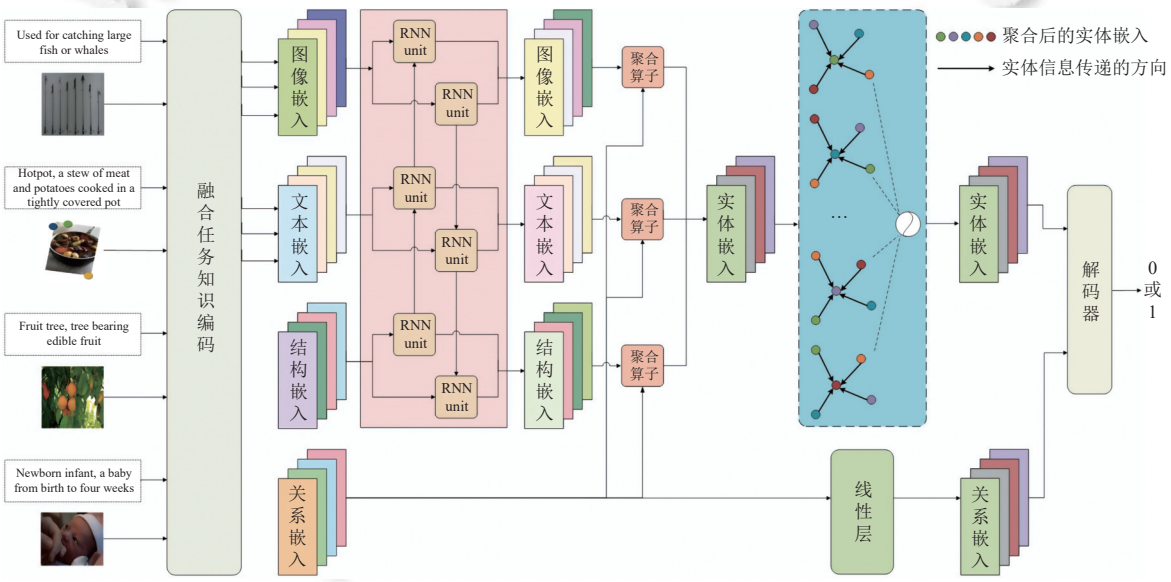


图 3 本文所提出的模型的结构示意图

2.1 符号和任务定义

在形式上, 一个多模态知识图谱可以表示为 $\mathcal{G}=\{\mathcal{E},\mathcal{R},\mathcal{T},\mathcal{V}\}$, 其中 $\mathcal{E}=\{e_1,e_2,\ldots,e_N\}$ 代表知识图谱中的实体集合, $\mathcal{R}=\{r_1,r_2,\ldots,r_M\}$ 代表知识图谱中的关系集合, $\mathcal{T}=\{x_1,x_2,\ldots,x_N\}$ 代表与实体相关联的描述文本集合, $\mathcal{V}=\{v_1,v_2,\ldots,v_N\}$ 表示与实体相关联的图像集合. 其中每个实体 e 拥有 1 个描述文本和 u ($\geqslant 0$) 张图像. 同时, 本文使用 $\mathcal{D}=\{(p,k,q)|p,q\in\mathcal{E},k\in\mathcal{R}\}$ 表示知识图谱 $\mathcal{G}$ 中事实三元组集合, N 表示实体的个数, M 表示关系的个数.

基于此, 给定一组查询: 头实体 p'和某个关系 k' ($query=\{p',k',?\}$), 我们的多模态知识图谱补全任务就是将所有可能的尾实体排序, 找到最合适的尾实体, 来匹配给定的头实体和关系. 本文就是为了学习一个多模态模型来达到此目的:

$$y=\mathcal{M}_{\Theta}(query,\mathcal{D}) \tag{1}$$

其中, $\mathcal{M}$ 表示我们计划学习出的多模态模型, Θ 代表所要学习的所有参数, y 表示所有可能满足要求的实体的概率.

另外, 本文分别使用 h_i^s, h_i^t, h_i^v 来表示实体 $e_i\in\mathcal{E}$ 的结构向量表示、描述文本向量表示和图像向量表示. 使用 r_j^s 来表示关系 $r_j\in\mathcal{R}$ 的结构向量表示.

2.2 融合任务知识的多模态编码

首先, 因为知识图谱中的实体数量和硬件设备的限制, 本文提前抽取所有实体的文本向量表示和图像向量表示. 同时, 为了将 MMKGC 的相关知识融入到多模态实体向量表示中, 本文在一个多模态编码器的基础上使用 MMKGC 任务对其进行微调. 然后使用微调后的多模态编码器 (multi-modal encoder, *MME*) 来抽取实体的文本向量表示和图像向量表示.

$$h_i^t,h_i^v=MME(x_i,v_i),\ h_i^t,h_i^v\in\mathbb{R}^{d_1} \tag{2}$$

其中, h_i^t 表示 i 个实体的文本描述向量表示. h_i^v 表示第 i 个实体的图像向量表示. d_1 表示向量维度, 本文设置为 768. *MME* 使用的是 MKGformer[16], 它是一种基于 Transformer[23]的多模态视觉-文本表示模型, 该模型在 CLIP[24]模型的基础上在最后几层 Transformer 中进行模态间的交互融合, 该模型能高效地抽取视觉和文本特征并应用于各种自然语言处理任务, 如: 多模态命名实体识别[25]和多模态关系抽取[26]等.

接着, 我们使用两个线性变换分别将实体 $e_i\in\mathcal{E}$ 的文本向量表示 h_i^t 和图像向量表示 h_i^v 映射到与结构向量表示 h_i^s 相同的向量空间.

$$\hat{h}_i^t=W^t h_i^t+b_1,\ \hat{h}_i^t\in\mathbb{R}^d \tag{3}$$

$$\hat{h}_i^v=W^v h_i^v+b_2,\ \hat{h}_i^v\in\mathbb{R}^d \tag{4}$$

其中, W^t, $W^v\in\mathbb{R}^{d\times d_1}$ 是可训练的权重矩阵, b_1, $b_2\in\mathbb{R}^d$ 是可训练的偏置. d 表示统一的向量空间维度, 本文设置为 200.

最终, 得到所有实体的 3 种模态下的向量表示 H^t,H^v,H^s 以及关系的结构向量表示 H^r.

$$H^t=\{\hat{h}_1^t,\hat{h}_2^t,\ldots,\hat{h}_N^t\},\ \hat{h}_i^t\in\mathbb{R}^d \tag{5}$$

$$H^v=\{\hat{h}_1^v,\hat{h}_2^v,\ldots,\hat{h}_N^v\},\ \hat{h}_i^v\in\mathbb{R}^d \tag{6}$$

$$H^s=\{h_1^s,h_2^s,\ldots,h_N^s\},\ h_i^s\in\mathbb{R}^d \tag{7}$$

$$H^r=\{h_1^r,h_2^r,\ldots,h_M^r\},\ h_i^r\in\mathbb{R}^d \tag{8}$$

其中, N 表示知识图谱中实体的个数, M 表示知识图谱中关系的个数.

2.3 多模态信息融合过滤

为了更好地获得实体在多个模态下全局向量以增强实体的表征能力. 同时, 根据 Wang 等人[10]的实验, 并不是所有的多模态特征都对任务有效. 并且, 模态两两交互的顺序对于模态融合效果也有一定的影响[27]. 因此, 本文使用基于序列的循环神经网络 (recurrent neural network, RNN) 结构, 利用其中的门控机制来过滤不同模态向量表示中的噪声, 同时保留适合当前任务的多模态信息.

$$\hat{H}^t,\hat{H}^v,\hat{H}^s=RNN(H^t,H^v,H^s) \tag{9}$$

其中, *RNN* 在本文中指的是长短期记忆网络 (long short term memory, LSTM)[28]和门控循环单元 (gated recurrent unit, GRU)[29]. 每个模态的向量表示作为一个时间步的输入, 也就是把 3 个模态看成一个序列, 每两个模态都可以在一个时间步中充分交互. 值得注意的是, 这么做既可以满足双模态两两交互, 也可以完成三模态同时交互, 这也是我们采用循环融合结构的动机之一.

2.4 基于聚合算子的实体关系融合

在经过多模态信息融合之后, 为了简化后续操作, 我们将边类型不同的异构知识图[30]转换为边类型相同的同构知识图, 即通过将知识图中不同类型的节点 (实体) 和所连接的边 (关系) 进行聚合. 同时, 为了降低模型对硬件设备的要求, 根据 Vashishth 等人[22]的实验, 使用 3 种不引入额外参数的聚合算子: 乘法 (multiplication, Mult)、减法 (subtraction, Sub)、循环相关 (circular correlation, Corr)[31], 用 $\otimes$ 代替, 每次实验选择一种算子.

$$z_{pk}^{t} = \hat{h}_{p}^{t} \otimes h_{k}^{r} \tag{10}$$

$$z_{pk}^{v} = \hat{h}_{p}^{v} \otimes h_{k}^{r} \tag{11}$$

$$z_{pk}^{s} = \hat{h}_{p}^{s} \otimes h_{k}^{r} \tag{12}$$

其中, $p \in [1,\ldots,N]$ 表示事实三元组中已知实体向量的索引, $k \in [1,\ldots,M]$ 表示事实三元组中关系向量的索引.

最后, 当前三元组中已知的节点和关系融合后的向量表示 z_{pk} 如公式 (13) 所示. 并最终得到所有事实三元组中已知节点和关系融合后的实体向量表示集合 Z.

$$z_{pk} = \frac{z_{pk}^{t} + z_{pk}^{v} + z_{pk}^{s}}{3} \tag{13}$$

$$Z = \{z_{pk} | \ (p,k,q) \in \mathcal{D}\} \tag{14}$$

2.5 实体和关系向量的更新

对于融合节点和关系的实体向量表示 Z, 我们使用同构图卷积网络来聚合邻居节点的信息, 并获得最终的实体特征表示 H^e. 同时使用一个线性层来更新关系向量 H^r.

$$H^e = GCN(\mathcal{D}, Z) \tag{15}$$

$$\hat{H}^r = W^r H^r + b \tag{16}$$

其中, $\mathcal{D}$ 表示知识图谱 $\mathcal{G}$ 中事实三元组集合. $W^r \in \mathbb{R}^{d\times d}$ 是一个可训练的参数矩阵, $b \in \mathbb{R}^d$ 为可训练的偏置. *GCN* 表示图卷积神经网络 (graph convolutional network).

2.6 解码过程

在通过编码器获得节点和关系的向量表示 H^e 和 $\hat{H}^r$ 之后, 将 H^e 和 $\hat{H}^r$ 输入到解码器中来对当前一批次事实三元组 B 中所有可能存在的三元组进行评分.

$$\hat{y} = Decoder(H^e, \hat{h}^r, B) \tag{17}$$

其中, 本文中的 *Decoder* 表示目前在知识图谱补全任务中被广泛使用的 3 个解码器, 包括 DistMult[7]、ComplEx[8] 和 ConvE[14], 具体计算方式依次如公式 (18)–公式 (20) 所示:

$$\hat{y}_j = H_p^e * \hat{H}_k^r * H_q^e \tag{18}$$

其中, $*$ 表示乘法操作. j 表示第 j 个的可能存在的三元组. p,k,q 分别代表一个三元组中的头实体, 关系和尾实体.

$$\hat{y}_j = \left(Re\left(\hat{H}_k^r\right) * Re\left(H_p^e\right) - Im\left(\hat{H}_k^r\right) * Im\left(H_p^e\right)\right) * Re\left(H_q^e\right)^{\mathrm{T}} + \left(Re\left(\hat{H}_k^r\right) * Im\left(H_p^e\right) + Im\left(\hat{H}_k^r\right) * Re\left(H_p^e\right)\right) * Im\left(H_q^e\right)^{\mathrm{T}} \tag{19}$$

其中, $Re(\cdot)$ 函数取出复数的实部, $Im(\cdot)$ 函数取出复数的虚部. $*$ 表示乘法操作. 在本文中, H^e, H^r 的向量维度为 200, 因此在使用 ComplEx 编码器的过程中, H^e, H^r 的 [0, 100] 维度表示复数实部, [101, 200] 维度表示复数虚部.

$$\hat{y}_j = \sigma\left(vec\left(\sigma\left(\left[\overline{\hat{H}_k^r}, \overline{H_p^e}\right] \odot \omega\right)\right) W\right)^{\mathrm{T}} H_q^e \tag{20}$$

其中, σ 表示 ReLU 激活函数, $vec(\cdot)$ 函数将张量转化为向量, $\overline{\hat{H}_k^r}, \overline{H_p^e}$ 分别表示 $\hat{H}_k^r, H_p^e$ 的 2D 重塑张量, $\odot$ 表示卷积

操作, ω 为一组滤波器, W 为一个可训练的权重矩阵.

2.7 训练过程

本文将存在于知识图谱中的三元组的标签值设置为 1, 将不存在于知识图谱中的三元组的标签值设置为 0, 使用的目标函数是二分类交叉熵函数 (binary cross entropy, BCE).

$$\mathcal{L}=-\sum_{d}^{|\mathcal{D}|}\sum_{i}^{N}(y_i^d\log(\hat{y}_i^d)+(1-y_i^d)\log(1-\hat{y}_i^d)) \tag{21}$$

其中, $\mathcal{D}$ 表示知识图谱 $\mathcal{G}$ 中事实三元组集合, N 表示整个数据集中的实体个数, $y_i^d\in\{0,1\}$ 表示第 d 个样本中第 i 个三元组的标签值, $\hat{y}_i^d$ 表示第 d 个样本的第 i 个三元组的预测得分.

3 实验结果

3.1 数据集

本文选取了两个公开的数据集 FB15k-237[32]和 WN9[33]. 其中 FB15k-237 数据集来自于 FreeBase[1], 共包含 14 541 个实体和 237 个关系, 该数据集的关系主要有对称关系、非对称关系和组合关系. WN9 数据集来自于 WordNet[2], 共包含 6 555 个实体和 9 个关系, 该数据集的关系主要有对称关系、非对称关系和反转关系. 在实验中, 如果一个实体只有文本信息而没有图像信息, 本文使用随机向量来初始化图像信息. 数据集类别等具体信息如表 1 所示.

表 1 数据集的统计学信息

数据集	实体	关系	图像	训练集	验证集	测试集
FB15k-237	14 541	237	14 297	272 115	17 535	20 466
WN9	6 555	9	6 547	11 741	1 337	1 319

3.2 参数设置

在实验中, 模型采用 Adam[34]优化器, 初始学习率设置为 0.001. 另外, 为了防止模型在训练过程中出现过拟合现象, 模型使用 Dropout[35], 其他一些重要神经网络参数如表 2 所示.

表 2 实验参数信息

参数	值
Batch size	128
Dropout	0.3
RNN网络层数	1
RNN网络输出维度	200
图卷积网络层数	1
图卷积网络输出维度	200

3.3 评价指标

在测试阶段, 给定一个实体和一个关系, 如果和预测出的实体组成的三元组存在于知识图谱当中, 则预测正确, 反之则预测错误. 本文中的实验性能通过 *Hit@K*[36]和 *MRR*[37]两种指标进行评估. 前者表示预测正确实体的得分在所有实体的预测得分中的降序排名前 K 的比例. 后者表示所有预测正确实体排名的倒数的均值. 具体计算方式如公式 (22), 公式 (23) 所示:

$$Hit@K=\frac{1}{|S|}\sum_{i=1}^{|S|}f(rank_i\leqslant K) \tag{22}$$

$$MRR=\frac{1}{|S|}\sum_{i=1}^{|S|}\frac{1}{rank_i} \tag{23}$$

其中, $|S|$ 表示数据集中事实三元组的个数, $rank_i$ 表示第 i 个事实三元组的预测排名. $f(\cdot)$ 表示如果条件为真则为

1, 否则为 0. $K=1,3,10$.

3.4 基准方法

为了验证基于融合任务知识的多模态知识图谱补全方法的有效性, 本文使用了以下几种最新的基准方法与之进行比较.

(1) TransE[6]. 该方法将实体向量和关系向量近似于一种向量运算关系 $p+k\approx q$, 其中 p 为头实体向量, k 为关系向量, q 为尾实体向量.

(2) DistMult[7]. 该方法使用对角矩阵来表示关系矩阵, 通过头实体向量、关系向量和尾实体向量三者的内积来计算三元组的得分.

(3) ComplEx[8]. 该方法首次将复数空间引入到知识图谱补全任务中, 将 DistMult 模型拓展到复数空间, 能够同时解决对称和非对称关系.

(4) ConvE[14]. 该方法使用二维卷积网络在实体向量和关系向量上预测知识图谱中缺失的三元组.

(5) RGCN[21]. 该方法将图卷积网络应用到异构图中, 提出了一个异构图中多关系融合的方法.

(6) KG-BERT[9]. 该方法于 2019 年提出了一个使用大规模预训练语言模型来融合实体上下文信息的知识图谱补全模型, 该模型将实体和关系拼接为序列, 将知识图谱补全任务转换为序列分类任务.

(7) RSME[10]. 该方法于 2021 年提出了一个通过门控机制自动过滤对任务无关的视觉信息的多模态知识图谱补全模型, 该模型可以保留有效的视觉信息, 从而增强编码能力.

(8) VisualBERT[38]. 该方法于 2019 年提出了一个单流的多模态表征模型, 依靠注意力机制将输入文本和图像对齐.

(9) ViLBERT[39]. 该方法于 2019 年提出了一个双流的多模态表征模型, 使用共同注意力 Transformer 层完成图像和文本的交互.

(10) MKGformer[16]. 该方法于 2022 年提出了一个通用的多模态表征模型, 在 CLIP[24]模型的基础上增加了跨模态交互模块, 促进了文本信息和图像信息地充分融合. 然后利用掩码语言模型来对缺失实体的类型进行预测.

(11) MANS[17]. 该方法于 2023 年提出了一个模态已知的负采样模型来对齐结构信息和模态信息, 进而学习对多模态知识图谱补全有价值的向量信息.

需要注意的是, 上述基准实验中的前 5 种方法都是基于结构向量独立完成知识图谱补全任务, 因此无法使用文本模态信息和图像模态信息.

第 2 组剩余 6 种方法则是在结构信息的基础上增加了文本模态信息或者图像模态信息. 其中, KG-BERT 方法只使用了文本模态信息, 其他 5 种方法则同时使用了文本模态信息和图像模态信息.

3.5 主实验分析

如表 3 所示, 在 FB15k-237 数据集和 WN9 数据集上, 本文比较了知识图谱补全领域 11 个基准方法的性能. 从表 3, 我们可以得出如下结论.

(1) 对于基准模型, 在数据集 FB15k-237 上, 基于结构信息的方法和基于模态信息的方法的性能不分伯仲: 就 *Hit*@1 和 *Hit*@3 而言, 基于结构信息的方法相对有优势, 特别是 TransE 模型. 而对于指标 *Hit*@10 和 *MRR*, 基于模态信息的方法相对更好, 特别是 MKGformer 展现出绝对的优势. 另外, 在 WN9 数据集上, 基于结构信息的方法性能明显强于基于模态信息的方法. 这主要因为该数据集规模较小, 比较依赖图谱中的结构信息. 这同时表明如何有效地利用多模态信息来提升知识图谱补全性能变得至关重要, 与我们的研究动机完全一致.

(2) 整体来讲, 本文所提出的融合任务知识的多模态知识图谱补全模型均优于 11 个基准方法, 这表明了本文所提出方法的优越性. 具体而言,

1) 与结构向量信息的基准实验相比, 本文提出的方法相较于 5 个基准实验性能均有一定的提升. 同时, 相较于使用 ConvE 的方法, 我们所提出方法的 *MRR* 指标在两个数据集上分别提高了 3.6% 和 1.0%. 而这主要归结于我们的方法能够充分利用模态信息来获得较好的实体向量表示.

2) 与基于模态信息的基准实验相比, 在 FB15k-237 数据集上, 本文提出的方法相较于表现最好的基准方法

MKGformer 在 *MRR* 指标上提升 1.7%, 在 *Hit*@10 指标上提升 3.3%. 而在 WN9 数据集上, 本文提出的方法相较于表现最好的基准方法 MANS-T 在 *MRR* 指标上有 33.3% 的大幅度提升, 在 *Hit*@10 指标上也提升了 3.5%. 而这主要归结于我们的方法能够有效利用多模态特征编码器中隐藏的先验任务知识和知识图谱中三元组之间的结构信息.

表 3 本文方法与其他基准方法的对比实验结果

信息类型	方法	FB15k-237				WN9			
		Hit@1	*Hit*@3	*Hit*@10	*MRR*	*Hit*@1	*Hit*@3	*Hit*@10	*MRR*
结构信息	TransE[6]	0.198*	0.376*	0.441*	—	0.861	0.904	0.920	0.886
	DistMult[7]	0.205	0.310	0.442	0.284	0.538	0.875	0.900	0.708
	ComplEx[8]	0.209	0.312	0.442	0.287	0.900	0.904	0.911	0.903
	ConvE[14]	0.233	0.354	0.499	0.321	0.900	0.906	0.912	0.904
	RGCN[21]	0.153*	0.258*	0.414*	0.248*	0.864	0.906	0.914	0.886
模态信息	KG-BERT[9]	—	—	0.420*	—	0.136	0.285	0.560	0.262
	RSME[10]	0.242*	0.344*	0.467*	—	—	—	—	—
	VisualBERT[38]	0.243	0.356	0.497	0.327	0.484	0.651	0.773	0.588
	ViLBERT[39]	0.233*	0.335*	0.457*	—	—	—	—	—
	MKGformer[16]	0.256	0.369	0.506	0.340	0.426	0.644	0.828	0.562
	MANS-S[17]	0.151	0.283	0.448	0.249	0.208	0.786	0.875	0.503
	MANS-T[17]	0.174	0.297	0.446	0.265	0.348	0.804	0.891	0.581
	MANS-H[17]	0.184	0.310	0.460	0.276	0.236	0.831	0.899	0.534
	MANS-A[17]	0.184	0.311	0.463	0.277	0.216	0.824	0.906	0.523
	$\text{Ours}_{\text{w/GRU}}$	0.265	0.389	0.538	0.356	**0.908**	**0.915**	**0.926**	**0.914**
	$\text{Ours}_{\text{w/LSTM}}$	**0.266**	**0.389**	**0.539**	**0.357**	0.906	0.915	0.926	0.913

注：*表示此项数据来自于MKGformer[16]

综上所述, 本文提出的方法能够更好地利用多模态信息增强实体的表征能力, 有效提升模型在多模态知识图谱补全任务中的表现.

4 分析与讨论

为了验证方法中不同模块的有效性, 本节主要进行 7 个消融实验, 来研究以下几个问题.

(1) 相较于使用不经过微调的通用文本和图像编码器所抽取的描述文本和图像特征, 使用融合任务知识的编码器进行特征抽取能否获得更好的预测效果? (第 4.1 节)

(2) 模态过滤融合模块的使用与否是否会造成模型性能的差异? (第 4.2 节)

(3) 考虑到 RNN 本身是一种基于序列的结构, 3 种模态过滤融合的顺序是否对模型性能产生影响? (第 4.3 节)

(4) 模态过滤融合模块的深度是否会对性能产生影响? (第 4.4 节)

(5) 不同的聚合算子会对模型性能产生怎样的影响? (第 4.5 节)

(6) 不同的解码器会对模型性能产生怎样的影响? (第 4.6 节)

(7) 图卷积神经网络的深度是否对性能产生影响? (第 4.7 节)

4.1 任务知识融合模块的影响

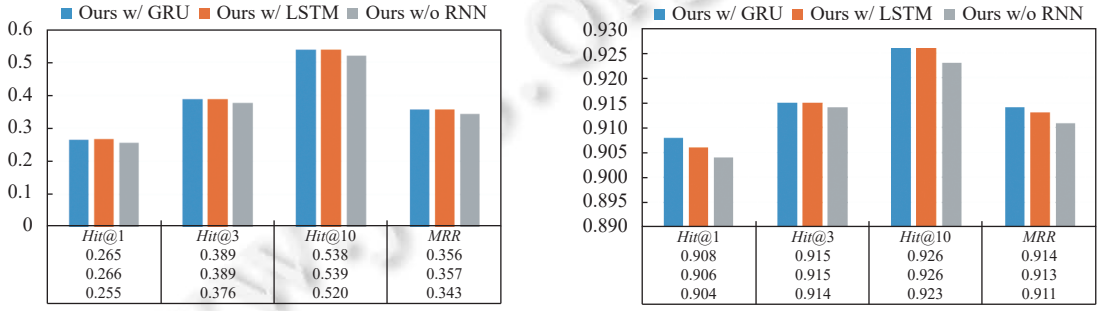
为了验证特定任务知识融合模块的有效性, 本文使用 VGG16[40]、ResNet50[41]、ResNet101[41]、ResNet152[41]和 Vision Transformer (ViT)[42]这 5 个图像编码器抽取实体相关联的图像特征. 使用 BERT[43]文本编码器抽取实体相关联的文本特征. 结果如表 4 所示, 我们可以观察到使用通用的图像编码器和文本编码器会造成模型明显的性能损失. 这主要是因为通用的模态信息编码器产生的特征编码不能很好地适应多模态知识图谱补全任务的特点, 无法利用任务知识来增强模态信息的表征能力, 进而表明本文提出的任务知识融合模块的必要性.

表 4 任务知识融合模块对于实验结果的影响

图像编码器	FB15k-237				WN9			
	LSTM		GRU		LSTM		GRU	
	Hit@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*
ResNet50	0.527	0.348	0.531	0.349	0.921	0.909	0.921	0.908
ResNet101	0.534	0.352	0.530	0.350	0.918	0.907	0.921	0.907
ResNet151	0.530	0.351	0.529	0.349	0.919	0.907	0.919	0.907
VGG16	0.524	0.348	0.532	0.353	0.918	0.905	0.922	0.905
ViT	0.534	0.351	0.531	0.351	0.917	0.906	0.918	0.908
Ours	**0.539**	**0.357**	**0.538**	**0.356**	**0.926**	**0.913**	**0.926**	**0.914**

4.2 多模态融合过滤模块的影响

为了验证多模态融合过滤模块的有效性, 我们移除了模态融合过滤模块, 直接将 3 种模态的实体向量进行求和后求均值的操作, 通过实验得到在不同聚合算子下的性能, 结果如图 4 所示. 可以发现, 使用了多模态融合过滤之后, 各个指标下的性能均有所提升. 我们认为这得益于本文提出的模态融合过滤模块可以充分过滤掉不同模态中与任务无关的信息, 融合和保留对当前任务有效的信息.



(a) FB15k-237 数据集 (b) WN9 数据集

图 4 多模态融合过滤模块对实验的影响

4.3 模态序列顺序的影响

根据模态两两交互的顺序对于模态融合效果可能存在影响. 我们采用基于序列结构的 RNN 网络, 因而我们改变多模态融合过滤模块中不同模态的输入顺序, 来进一步验证顺序对性能的影响. 结果如表 5 所示, 从表 5 中可以发现, 采用 S-I-T 顺序可以带来最佳性能, 但是实际不同顺序带来的差异较小, 所以我们没有进一步深入分析. 在实际应用中, 通常选择其中一种就可以达到相对较好的性能.

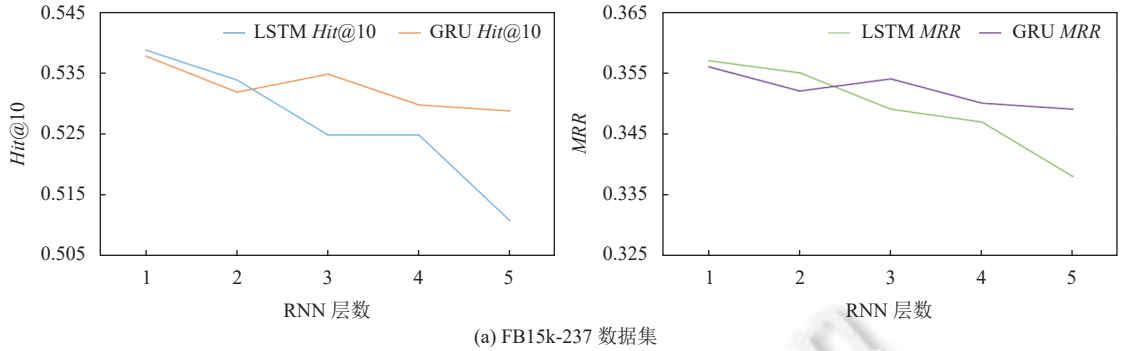
表 5 模态顺序对于实验结果的影响.

模态顺序	FB15k-237				WN9			
	LSTM		GRU		LSTM		GRU	
	Hit@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*
S-I-T	**0.539**	**0.357**	**0.538**	**0.356**	**0.926**	**0.913**	**0.926**	**0.914**
S-T-I	0.537	0.355	0.530	0.352	0.920	0.909	0.921	0.909
T-S-I	0.535	0.356	0.536	0.356	0.923	0.913	0.922	0.911

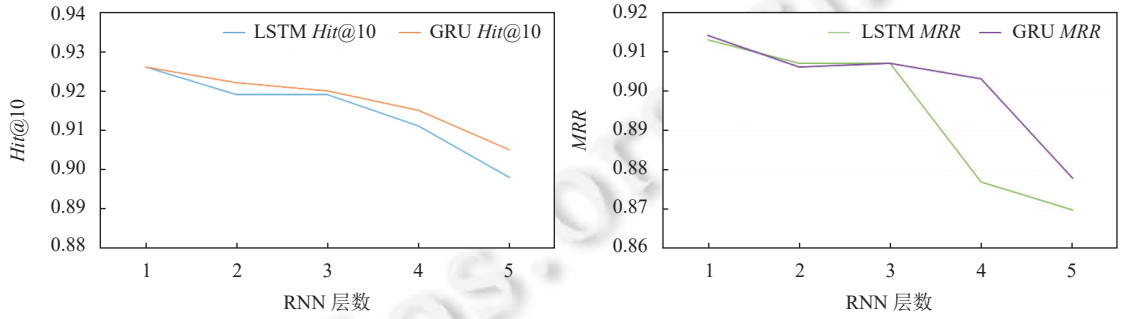
注: S表示实体的结构向量表示, I表示图像向量表示, T表示文本描述向量表示

4.4 RNN 层数的影响

本文通过改变 RNN 的层数来观察模态融合过滤模块中 RNN 层数变化对于性能的影响, 实验结果如图 5 所示. 从结果可以看出, 随着 RNN 层数增加, LSTM 和 GRU 模型的两个评价指标均出现降低的现象. 我们认为随着 RNN 的层数的增加, 不同的模态的实体向量表示逐渐趋于平滑, 实体之间的差异性降低, 甚至会引入噪声, 从而导致模型性能的下降. 所以, 采用 1 层 RNN 网络对于多模态知识图谱补全任务就已经足够.



(a) FB15k-237 数据集



(b) WN9 数据集

图 5 RNN 层数对于实验结果的影响

4.5 聚合算子的影响

考虑到知识图谱包含的三元组数量巨大, 为了降低对显存的要求, 本文使用 3 种常规的不引入额外参数的聚合方法: 减法 (Sub), 乘法 (Mult), 共现 (Corr), 实验结果表 6 所示. 根据实验结果可知, 在 FB15k-237 数据集上使用 ConvE 解码器的情况下, 就指标 *MRR* 而言, 采用 LSTM 结构融合过滤, 使用乘法算子相较于减法算子和共现算子, 性能分别提升了 0.4% 和 0.5%; 采用 GRU 模态融合结构, 使用乘法算子相较于减法算子和共现算子, 性能分别提升了 0.3% 和 0.2%. 在 WN9 数据集上, 采用 LSTM 和 GRU 结构, 乘法算子同样都是明显优于减法和共现算子. 这些结果表明乘法聚合算子对于本文提出的方法是一个不错的选择.

表 6 聚合算子和解码器对于实验结果的影响

数据集	解码器	Sub				Mult				Corr			
		LSTM		GRU		LSTM		GRU		LSTM		GRU	
		Hit@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*	*Hit*@10	*MRR*
FB15k-237	D	0.519	0.340	0.515	0.336	0.517	0.337	0.517	0.338	0.521	0.341	0.519	0.354
	C	0.517	0.340	0.519	0.339	0.521	0.339	0.521	0.340	0.519	0.339	0.519	0.339
	CE	**0.534**	**0.353**	**0.538**	**0.353**	**0.539**	**0.357**	**0.538**	**0.356**	**0.530**	**0.352**	**0.534**	**0.354**
WN9	D	0.915	0.770	0.917	0.771	0.920	0.759	0.912	0.771	0.914	0.766	0.915	0.760
	C	**0.925**	0.909	**0.925**	0.906	0.923	0.911	0.922	0.909	0.903	0.910	**0.926**	**0.910**
	CE	0.920	**0.910**	0.924	**0.913**	**0.926**	**0.913**	**0.926**	**0.914**	**0.923**	**0.910**	0.920	0.909

注: D、C、CE依次表示DistMult, ComplEx, ConvE

4.6 解码器的影响

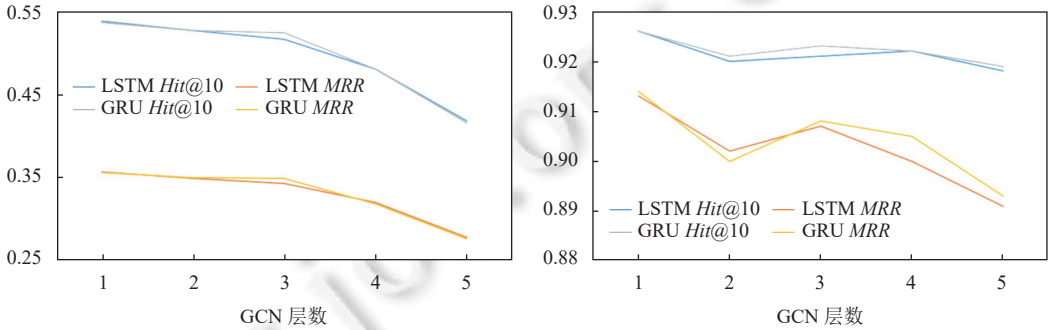
为了验证本文提出的编码器结构在不同解码器 (DistMult, ComplEx, ConvE) 下的性能表现, 我们将编码器与不同的解码器进行拼接, 实验结果如表 6 所示. 根据实验结果, 以乘法聚合算子和 LSTM 模态过滤融合为例, 就 *MRR* 指标而言, 相比于 DistMult 和 ComplEx, ConvE 解码器在 FB15k-237 数据集上的性能分别高了 2.0% 和

1.8%. 在 WN9 数据集上的性能分别提升了 15.4% 和 0.2%.

值得注意的是, 1) 对于在两个数据集上性能提升的差异, 我们认为 WN9 数据集相比于 FB15k-237 数据集在难易程度上相对容易, 使用简单的方法就可以得到很好的效果, 因此在 WN9 数据集上提升幅度较小. 2) ConvE 在 3 个解码器中表现最优, 我们认为这是由于 ConvE 解码器中使用卷积操作对实体和关系向量进行了更复杂的融合所导致的.

4.7 图卷积网络层数的影响

本文通过改变 GCN 层数来观察实体聚合深度对于性能的影响, 实验结果如图 6 所示. 由结果可知, 随着 GCN 层数增加, 两个评价指标在两个数据集上均出现降低的现象, 其中在 FB15k-237 数据集上性能出现了显著下降. 一方面, 我们认为随着 GCN 层数的增加, 同一个实体确实能聚合更多邻居节点 (一阶邻居、二阶邻居、…、五阶邻居) 的信息, 但是不同节点之间邻居的重合度也会不断升高, 导致节点的可判别性降低, 从而造成模型性能的下降. 另一方面, FB15k-237 数据集相比于 WN9 数据集, 剔除了反向关系, 对特征的要求更高, 所以在本实验中在 FB15k-237 数据集上的性能降低幅度高于在 WN9 数据集上的性能降低幅度.



(a) FB15k-237 数据集 (b) WN9 数据集

图 6 GCN 层数对于实验结果的影响

5 总 结

本文提出了一种融合任务知识的多模态知识图谱补全方法, 用于更好地从不同模态信息中获取与任务相关的有效信息, 来增强知识图谱中节点的表征能力. 具体而言, 该方法首先通过融合该领域内其他方法来提取节点的多模态特征表示, 其次通过多模态融合过滤网络来进一步取出冗余信息, 筛选有效的多模态特征信息, 然后使用多种聚合算子来将关系和节点进行聚合, 进而将异构图网络转换为同构图网络以便更有效利用图卷积网络来完成多模态知识图谱补全任务. 实验结果表明, 本文提出的基于融合任务知识的多模态知识图谱补全方法能够有效地利用其他方法的特点以及更好地融合过滤多个模态的全局信息来提高多模态知识图谱补全方法的性能.

References:

[1] Bollacker K, Cook R, Tufts P. FreeBase: A shared database of structured general human knowledge. In: Proc. of the 22nd National Conf. on Artificial Intelligence. Vancouver: AAAI Press, 2007. 1962–1963.

[2] Miller GA. WordNet: A lexical database for English. Communications of the ACM, 1995, 38(11): 39–41. [doi: 10.1145/219717.219748]

[3] Chen Y, Wu LF, Zaki MJ. Bidirectional attentive memory networks for question answering over knowledge bases. arXiv:1903.02188, 2019.

[4] Chen YH, Tan CY, Chen WL, Jia YH, He ZQ. Chinese knowledge graph complementation with multiple embeddings. Journal of Chinese Information Processing, 2023, 37(1): 54–63 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2023.01.005]

[5] Zhang FZ, Yuan NJ, Lian DF, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 353–362. [doi: 10.1145/2939672.2939673]

[6] Bordes A, Usunier N, Garcia-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795.

[7] Yang BS, Yih WT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. arXiv:1412.6575, 2015.

[8] Trouillon T, Welbl J, Riedel S, Gaussier É, Bouchard G. Complex embeddings for simple link prediction. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: JMLR.org, 2016. 2071–2080.

[9] Yao L, Mao CS, Luo Y. KG-BERT: BERT for knowledge graph completion. arXiv:1909.03193, 2019.

[10] Wang M, Wang S, Yang H, Zhang Z, Chen X, Qi GL. Is visual context really helpful for knowledge graph? A representation learning perspective. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 2735–2743. [doi: 10.1145/3474085.3475470]

[11] Zhang NY, Xie X, Chen X, Deng SM, Ye HB, Chen HJ. Knowledge collaborative fine-tuning for low-resource knowledge graph completion. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3531–3545 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6628.htm [doi: 10.13328/j.cnki.jos.006628]

[12] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2017.

[13] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph embedding by translating on hyperplanes. In: Proc. of the 28th AAAI Conf. on Artificial Intelligence. Québec City: AAAI, 2014. 1112–1119. [doi: 10.1609/aaai.v28i1.8870]

[14] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 1811–1818. [doi: 10.1609/aaai.v32i1.11573]

[15] Nguyen DQ, Nguyen TD, Nguyen DQ, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 2 (Short Papers). New Orleans: Association for Computational Linguistics, 2018. 327–333. [doi: 10.18653/v1/N18-2053]

[16] Chen X, Zhang NY, Li L, Deng SM, Tan CQ, Xu CL, Huang F, Si L, Chen HJ. Hybrid Transformer with multi-level fusion for multimodal knowledge graph completion. In: Proc. of the 45th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Madrid: ACM, 2022. 904–915. [doi: 10.1145/3477495.3531992]

[17] Zhang YC, Chen MY, Zhang W. Modality-aware negative sampling for multi-modal knowledge graph embedding. In: Proc. of the 2023 Int'l Joint Conf. on Neural Networks (IJCNN). Gold Coast: IEEE, 2023. 1–8. [doi: 10.1109/IJCNN54540.2023.10191314]

[18] Liang WX, Jiang YH, Liu ZX. GraghVQA: Language-guided graph neural networks for graph-based visual question answering. arXiv:2104.10283, 2021.

[19] Ghosal D, Majumder N, Poria S, Chhaya N, Gelbukh A. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In: Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019. 154–164. [doi: 10.18653/v1/D19-1015]

[20] Ying CX, Cai TL, Luo SJ, Zheng SX, Ke GL, He D, Shen YM, Liu TY. Do Transformers really perform bad for graph representation? arXiv:2106.05234, 2021.

[21] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling relational data with graph convolutional networks. In: Proc. of the 15th Int'l Conf. on the Semantic Web. Heraklion: Springer, 2018. 593–607. [doi: 10.1007/978-3-319-93417-4_38]

[22] Vashishth S, Sanyal S, Nitin V, Talukdar P. Composition-based multi-relational graph convolutional networks. arXiv:1911.03082, 2020.

[23] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

[24] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. Learning transferable visual models from natural language supervision. In: Proc. of the 38th Int'l Conf. on Machine Learning. 2021. 8748–8763.

[25] Zhang D, Wei SZ, Li SS, Wu HQ, Zhu QM, Zhou GD. Multi-modal graph fusion for named entity recognition with targeted visual guidance. In: Proc. of the 35th AAAI Conf. on Artificial Intelligence. AAAI, 2021. 14347–14355. [doi: 10.1609/aaai.v35i16.17687]

[26] Zheng CM, Feng JH, Fu Z, Cai Y, Li Q, Wang T. Multimodal relation extraction with efficient graph alignment. In: Proc. of the 29th ACM Int'l Conf. on Multimedia. ACM, 2021. 5298–5306. [doi: 10.1145/3474085.3476968]

[27] Sun H, Wang HY, Liu JQ, Chen YW, Lin LF. CubeMLP: An MLP-based model for multimodal sentiment analysis and depression estimation. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 3722–3729. [doi: 10.1145/3503161.3548025]

[28] Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. Neural Computation, 2000, 12(10): 2451–2471. [doi: 10.1162/089976600300015015]

[29] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv: 1412.3555, 2014.

[30] Zhang CX, Song DJ, Huang C, Swami A, Chawla NV. Heterogeneous graph neural network. In: Proc. of the 25th ACM SIGKDD Int'l

Conf. on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 793–803. [doi: 10.1145/3292500.3330961]

[31] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Phoenix: AAAI, 2016. 1955–1961. [doi: 10.1609/aaai.v30i1.10314]

[32] Toutanova K, Chen DQ, Pantel P, Poon H, Choudhury P, Gamon M. Representing text for joint embedding of text and knowledge bases. In: Proc. of the 2015 Conf. on Empirical Methods in Natural Language Processing. Lisbon: Association for Computational Linguistics, 2015. 1499–1509. [doi: 10.18653/v1/D15-1174]

[33] Xie RB, Liu ZY, Luan HB, Sun MS. Image-embodied knowledge representation learning. arXiv:1609.07028, 2017.

[34] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2017.

[35] Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. arXiv:1207.0580, 2012.

[36] Chen ZR, Wang X, Wang CX, Zhang SW, Yan HY. Towards time-aware knowledge hypergraph link prediction. Ruan Jian Xue Bao/Journal of Software, 2023, 34(10): 4533–4547 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6888.htm [doi: 10.13328/j.cnki.jos.006888]

[37] Pang J, Liu XQ, Gu Y, Wang X, Zhao YH, Zhang XL, Yu G. Knowledge hypergraph link prediction based on multi-granular attention network. Ruan Jian Xue Bao/Journal of Software, 2023, 34(3): 1259–1276 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6788.htm [doi: 10.13328/j.cnki.jos.006788]

[38] Li LH, Yatskar M, Yin D, Hsieh CJ, Chang KW. VisualBERT: A simple and performant baseline for vision and language. arXiv:1908.03557, 2019.

[39] Lu JS, Batra D, Parikh D, Lee S. ViLBERT: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 13–23.

[40] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2015.

[41] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. In: Proc. of the 2016 IEEE Conf. on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: 10.1109/CVPR.2016.90]

[42] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai XH, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929, 2021.

[43] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional Transformers for language understanding. arXiv:1810.04805, 2019.

附中文参考文献:

[4] 陈跃鹤, 谈川源, 陈文亮, 贾永辉, 何正球. 结合多重嵌入表示的中文知识图谱补全. 中文信息学报, 2023, 37(1): 54–63. [doi: 10.3969/j.issn.1003-0077.2023.01.005]

[11] 张宁豫, 谢辛, 陈想, 邓淑敏, 叶宏彬, 陈华钧. 基于知识协同微调的低资源知识图谱补全方法. 软件学报, 2022, 33(10): 3531–3545. http://www.jos.org.cn/1000-9825/6628.htm [doi: 10.13328/j.cnki.jos.006628]

[36] 陈子睿, 王鑫, 王晨旭, 张少伟, 闫浩宇. 面向时间感知的知识超图链接预测. 软件学报, 2023, 34(10): 4533–4547. http://www.jos.org.cn/1000-9825/6888.htm [doi: 10.13328/j.cnki.jos.006888]

[37] 庞俊, 刘小琪, 谷峪, 王鑫, 赵宇海, 张晓龙, 于戈. 基于多粒度注意力网络的知识超图链接预测. 软件学报, 2023, 34(3): 1259–1276. http://www.jos.org.cn/1000-9825/6788.htm [doi: 10.13328/j.cnki.jos.006788]

陈强(1999－), 男, 硕士生, 主要研究领域为自然语言处理.

李寿山(1980－), 男, 博士, 教授, CCF 专业会员, 主要研究领域为自然语言处理.

张栋(1991－), 男, 博士, 副教授, CCF 专业会员, 主要研究领域为自然语言处理.

周国栋(1967－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为自然语言处理.