DOI:10.3969/j.issn.1003-5060.2026.05.008

基于张量分解的多模态知识图谱嵌入

胡雨琪[1], 杨依忠[1], 张晴宇[2]
(1.合肥工业大学 微电子学院,安徽 合肥 230601; 2.合肥工业大学 计算机与信息学院,安徽 合肥 230601)

摘 要:由于不同模态信息之间存在异构性,简单的多模态信息融合方式无法有效捕捉深层语义信息,难以辅助多模态知识在低维连续空间中的张量嵌入。为了探究视觉模态信息优化表示学习的有效方法,文章提出一种基于张量分解的多模态知识表示学习模型,旨在解决多源知识融合和多模态知识嵌入的问题。该模型首先利用多模态数据筛选模块对视觉信息进行基础编码,并计算图像的公共相似度过滤潜在的信息干扰;然后在多模态数据融合模块中设置权重将不同模态信息进行融合,从而得到初始化的多模态知识编码;最后在多模态知识嵌入模块中通过张量分解原理,将多模态知识编码嵌入到低维连续张量空间;此外还设计了基于多预测任务的联合损失函数,以提高模型训练效率和知识嵌入效果。链接预测实验结果表明,该模型在多个指标上均超越了经典基线模型,说明通过合理的数据过滤、编码与嵌入方法,多模态信息能够显著提高知识表示效果。

关键词:多模态信息;知识图谱;知识表示;信息融合;知识挖掘
中图分类号:TP301.6 **文献标志码**:A **文章编号**:1003-5060(2026)05-0629-08

Multimodal knowledge graph embedding based on tensor decomposition

HU Yuqi[1], YANG Yizhong[1], ZHANG Qingyu[2]
(1. School of Microelectronics, Hefei University of Technology, Hefei 230601, China; 2. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China)

Abstract: Due to the heterogeneity between different modal information, simple multimodal information fusion methods cannot effectively capture the deep semantic information, which makes it difficult to assist the tensor embedding of multimodal knowledge in low-dimensional continuous space. In order to explore an effective method for optimal representation learning of visual modal information, this paper proposes a multimodal knowledge representation learning model based on tensor decomposition, aiming to solve the problems of multi-source knowledge fusion and multimodal knowledge embedding. The model first uses a multimodal data filtering module to perform basic encoding on the visual information, and calculates the common similarity of images to filter potential information interference. Then, the weights are set in the multimodal data fusion module to fuse the information from different modalities to obtain the initialized multimodal knowledge encoding. Finally, the multimodal knowledge encoding is embedded into a low-dimensional continuous tensor space through the tensor decomposition principle in the multimodal knowledge embedding module. The paper also designs a joint loss function based on the multi-prediction task to improve the model training efficiency and knowledge embedding effect. The results of the link prediction experiments show that the model exceeds the classical baseline model in several indexes, indicating that the multimodal information can significantly improve the effect of knowledge representation through reasonable data filtering, encoding and embedding methods.

收稿日期:2024-01-05;**修回日期**:2024-02-27
基金项目:安徽省自然科学基金资助项目(2208085MF177)
作者简介:胡雨琪(1996—),女,安徽怀宁人,合肥工业大学硕士生;
杨依忠(1975—),男,安徽和县人,博士,合肥工业大学副教授,硕士生导师,通信作者,E-mail:zireael77@163.com.

Key words: multimodal information; knowledge graph; knowledge representation; information fusion; knowledge mining

0 引 言

知识图谱作为知识工程的重要分支,是一种结构化的语义知识库,它采用(头节点,关系,尾节点)的三元组形式描述客观世界的知识和概念,例如 DBpedia[1]、Freebase[2] 和 WordNet[3] 等,知识图谱在生活和工业领域已经得到广泛应用。知识表示是知识图谱构建的核心环节,旨在解决数据稀疏问题。受到 Word2Vec 模型[4] 的启发,在低维稠密张量空间中进行知识嵌入可以表示实体与关系,从而缓解数据稀疏问题。同时,通过计算实体与关系在空间中的距离,可以揭示它们之间的潜在语义关联。多模态知识图谱在传统知识图谱的基础上融入了实体的视觉、音频和文本等数据。文献[5-6]证实,利用多模态知识图谱可以显著提升相关下游领域的应用效果,如为智能问答[7] 或推荐系统[8] 等任务提供丰富的背景故事或先验知识。

文献[9]表明,如果数据融合的方法过于简单,那么多模态信息比仅使用单模态信息的知识表示效果会更差。为解决多模态信息的语义挖掘与知识嵌入问题,本文提出一个多模态知识表示模型。该模型包含 3 个主要模块:① 多模态数据筛选模块,它利用图像公共相似性来筛选最佳视觉信息;② 多模态编码融合模块,它用于对齐不同模态知识进行融合;③ 多模态嵌入模块,它基于张量分解原理将多模态知识嵌入到低维稠密张量空间。最终,该模型整合不同模态之间的互补信息,实现多模态知识图谱的知识表示。

综上所述,本文贡献主要有以下 3 点:

1) 提出一种多模态信息筛选方法,该方法利用图像公共相似度的过滤机制来消除潜在的信息噪声,从而提高多模态信息的编码效果。

2) 提出一种多模态编码融合方法,该方法使用映射矩阵将不同模态数据投影到同一张量空间以解决模态间的异构性,并利用动态权重进行信息融合,获得高质量的多模态编码。

3) 提出一种多模态编码的张量嵌入方法,该方法使模型不仅能够针对多样化关系进行建模,还能够让多模态知识在嵌入的张量空间中得到更好表达。

1 相关工作

1.1 单模态知识图谱嵌入

单模态表示学习是最经典的表示学习方法,也是多模态表示学习的基础。文献[10]提出的 TransE 模型将关系视为低维向量空间中头实体到尾实体的翻译操作;为了编码更多样的关系,文献[11]提出的 TransH 模型将关系视为超平面上的转换操作,每个实体在不同超平面中有不同表示;文献[12]提出的 TransR 模型利用映射矩阵在不同关系空间中进行转换;文献[13]提出的 TorusE 模型将嵌入空间定义在圆环面,以克服 TransE 的局限性。语义模型将实体与关系的深层语义嵌入张量空间;文献[14]的 RESCAL 模型用满秩矩阵表示关系,实现实体与关系的交互;文献[15]的 DisMult 模型简化关系矩阵为对角矩阵,降低了 RESCAL 模型的复杂度;文献[16]的 ComplEx 模型扩展实体表示为复数空间,能同时处理对称与非对称关系;文献[17]的 CP 模型改进传统张量分解方法来约束张量在低维连续空间的嵌入。还有一些研究工作将知识图谱嵌入视为分类问题,如文献[18]的 ConvE 模型利用卷积层和全连接层捕获实体与关系的交互信息,文献[19]的 ConvKB 模型使用胶囊网络进行知识嵌入表达。但这些模型在处理知识图谱实体的层次性和关系的多样性时存在局限,且可解释性不足。

1.2 多模态知识图谱嵌入

目前,多数研究采用多模态信息来增强知识表示学习的效果。文献[20]提出的 IKRL 模型首次提出通过三元事实和图像来学习知识嵌入;文献[21]将多模态自动编码器与 TransE 模型相结合,提出一种名为 TransAE 的表示学习方法;文献[6]提出一种新的方法,该方法将不同模态分开编码,并利用多模态信息和结构化信息的评分函数来计算实体间关系的可能性;文献[22]引入注意力机制,以深入研究不同模态间的融合比例,使重要模态在融合嵌入中发挥更大作用,并学习基于图像的知识嵌入方法。

2 多模态数据融合

2.1 多模态链接预测任务

链路不完整性是知识图谱中亟待解决的关键

问题之一，链接预测任务旨在利用知识图谱中已知的事实，探索实体间的潜在关系，以解决链路不完整性问题。结构化三元组中头实体、关系、尾实体可以表示为 $\boldsymbol{G}=(h,r,t)$，链接预测任务的目标是在某一节点缺失时，通过其余 2 种元素推断出需要补充的实体对象。评分函数用于衡量一个三元组的置信度，即其正确性的概率，表示为 $F_{(h,r,t)}\in[0,1]$，评分函数 $F_{(h,r,t)}$ 可以为每个候选三元组计算一个置信度分数，该分数表示三元组在知识图谱中存在的可能性。

2.2　多模态知识表示结构

多模态知识表示模型包含 3 个主要部分，即编码部分、融合部分和嵌入部分。编码模块负责为每个实体的多模态信息生成对应编码。融合部分负责整合这些编码得到多模态信息的初始张量。嵌入部分利用低秩张量分解原理，将融合后的初始张量嵌入到向量空间中。模型的整体框架结构如图 1 所示。模型首先使用不同的编码器对结构化数据、视觉模态和文本模态进行编码，其中结构化数据进一步分为实体节点嵌入和关系嵌入；然后根据不同模态的融合权重整合这些编码，因为只有实体节点有对应的多模态数据，而关系仅存在于结构化文本数据中，所以关系嵌入不参与数据融合；最后使用融合后的跨模态编码，通过矩阵分解原理和多重损失函数，在低维稠密空间中完成嵌入。

假定最终复原预测张量为 $\boldsymbol{R}^{E\times F\times G}$，$E$ 代表头实体数量，F 代表关系数量，G 代表尾实体数量，那么预测张量 $\boldsymbol{R}$ 中，长、宽、高分别为 h、r、t，对应的点就是 (h,r,t) 这个三元组，该点的数值 p 表示这个三元组 (h,r,t) 对应事实为真的概率 $P_{(h,r,t)}$，即 $P_{(h,r,t)}=p$。通过这种方式，可以利用预测张量 $\boldsymbol{R}$ 来估算数据集中任意实体与关系组合的三元组概率，进而完成知识补全等任务。

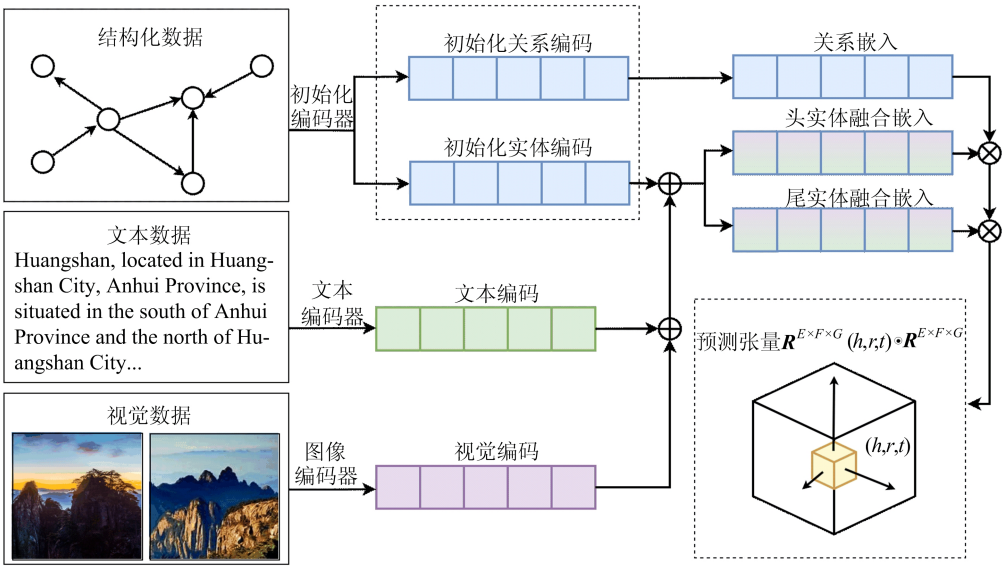

图 1　模型的整体框架结构

2.3　实体图像编码与筛选

在多模态知识图谱中，实体的视觉信息可能存在不相关且易引起歧义的图片，这大大影响知识的准确表示。为了避免这种问题，需要从所有相关图像中挑选出与实体最匹配的图像。整体的筛选过程如图 2 所示。

首先计算每张候选图像与查询图像之间的相似度，这类似于 Word2vec[4] 词嵌入方法，通过图像编码器来计算每张图像的语义信息；然后计算两幅图像编码的余弦相似度，得到相似度分数，相似度分数总和最高的图像被视为该实体的最佳匹配图像。图像编码器用于获得图像的初始编码，由于数据融合过程中需要图像的密集编码而非离散类别，因此要移除预训练图像编码器的最后一层分类全连接层。

对于不同的预训练模型，采用不同的网络隐藏层作为图像的基本嵌入，以确保编码包含图像的深层特征和语义信息。模型采用3 种图像编码器，即 VGG16[23]、Resnet50[24]、VIT[25]。

VGG16 由多个卷积层和池化层交替堆叠而成，并在每个卷积层和池化层后均使用 ReLU 激活函数，采用其最后一个隐藏层的输出。Res-

net50 针对网络特征的退化现象引入残差连接，采用其 Stage4 的输出。VIT 是基于 Transfomer 模型的改进，采用其倒数第 2 个隐藏层输出的非 CLS 部分作为图像编码。

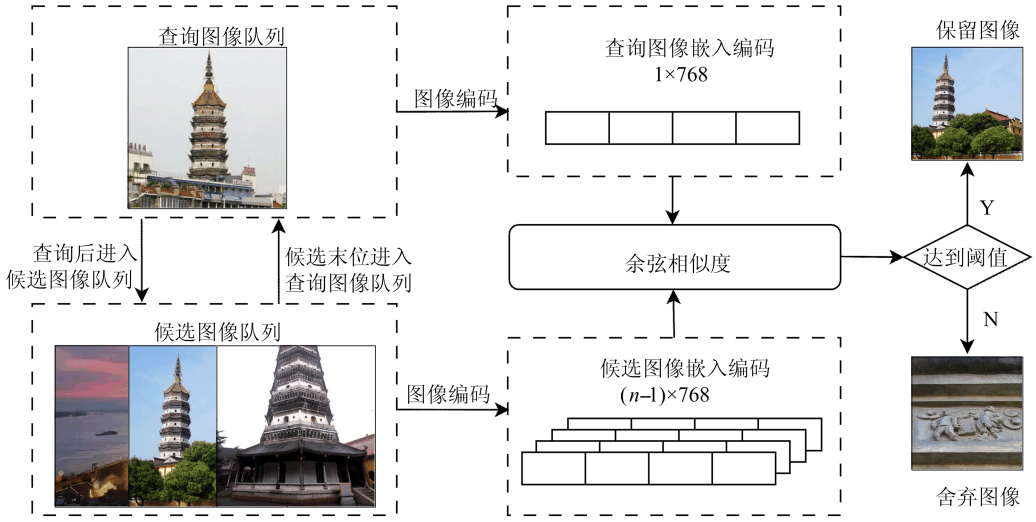

图 2 通过图像相似度获取最佳图像

通过计算查询嵌入与所有候选嵌入队列间的相似度得分，构建图像间的相似度矩阵。相似度矩阵中每一行表示一个查询嵌入与每个候选队列嵌入的余弦相似度。计算完所有查询队列图像后，统计每张查询图像的相似性得分总和，以此作为其公共相似度。图像计算公共相似度的公式为：

$$X_i = \left\| \sum_{j=1}^{n} \cos(F_{(i)}, F_{(j)}) \right\|, \quad i,j = 1,2,\cdots,n \tag{1}$$

其中：X_i 为实体第 i 张相关图像的公共相似度；n 为与该实体相关的视觉查询图像的总数；$F_{(i)}$ 为实体第 i 张图像通过图像编码器后得到的输出编码；$\cos(x,y)$ 为 x、y 这 2 个编码的余弦相似度。在相似度矩阵构建完成后，选择得分最高的图像作为最佳图像，最佳图像定义公式为：

$$S_t = \underset{i=1,2,\cdots,n}{\operatorname{argmax}}(X_1, X_2, \cdots, X_i) \tag{2}$$

其中：S_t 为某一实体节点 t 对应所有子图像中的最佳图像；n 为该实体对应的视觉查询图像总数量；X_i 为实体第 i 张相关图片的公共相似度。将每张图像的相似度总得分进行排序，相似度总得分最高的图像即为该实体的最佳代表性图像。

3 基于张量分解的知识嵌入

3.1 多模态编码融合

在数据融合前，需将各类数据映射至统一维度空间，图像需从初始编码结果映射至与张量初始化一致的维度。假设图像编码器的初始输出为 $\boldsymbol{E}_{i1}$，图像编码器映射后的输出为 $\boldsymbol{E}_{i2}$，文本编码器的初始输出为 $\boldsymbol{E}_{t1}$，文本编码器映射后的输出为 $\boldsymbol{E}_{t2}$，$\boldsymbol{E}_{i2}$ 与 $\boldsymbol{E}_{t2}$ 的计算公式为：

$$\boldsymbol{E}_{i2} = \text{sigmoid}(F_i(\boldsymbol{E}_{i1}) + \boldsymbol{b}_i) \tag{3}$$

$$\boldsymbol{E}_{t2} = \text{sigmoid}(F_t(\boldsymbol{E}_{t1}) + \boldsymbol{b}_t) \tag{4}$$

其中：$F_i(x)$ 代表图像编码器；$F_t(x)$ 代表文本编码器；sigmoid 为激活函数；$\boldsymbol{b}_i$ 为图像编码的偏置项；$\boldsymbol{b}_t$ 为文本编码的偏置项。最终 $\boldsymbol{E}_{i2}$、$\boldsymbol{E}_{t2}$ 与 $\boldsymbol{E}_s$ 统一到同一维度 d_s。在文本模态和图像模态的编码完成后，设定相应权重以均衡各模态在融合中的比例，模态融合的公式为：

$$\boldsymbol{E}_f = \alpha F_s(\boldsymbol{E}_s) + \beta \boldsymbol{E}_{i2} + \delta \boldsymbol{E}_{t2}, \quad f = 1,2,\cdots,N_e, \alpha+\beta+\delta = 1 \tag{5}$$

其中：$\boldsymbol{E}_f$ 为对任意实体 f 的初始融合张量；N_e 为实体的总数量；$\boldsymbol{E}_s$ 为结构性数据的初始化编码，编码维度为 d_s；F_s 为对结构性数据的映射函数；α、β、δ 为用来平衡不同模态的权重占比的超参数。

3.2 多模态编码嵌入

为了在低维稠密空间中更有效地表达嵌入张量，并能建模一对多和多对一的关系，模型利用 Canonical 分解方法将融合后的多模态张量进行嵌入，其核心目的是恢复知识图谱的三阶概率矩阵，也称为预测矩阵。预测矩阵中的任意元素都代表 2 个实体节点与某一关系构成的三元组的正确概率。利用该矩阵可以执行知识图谱相关的下游任务。假设三阶预测矩阵是 $\boldsymbol{Y} \in \mathbf{R}^{E \times F \times G}$，用列向量表示预测矩阵的公式为：

$$\boldsymbol{Y} \approx \sum_{r=1}^{R} \boldsymbol{a}_r \otimes \boldsymbol{b}_r \otimes \boldsymbol{c}_r,$$
$$\boldsymbol{a}_r \in \mathbf{R}^E, \boldsymbol{b}_r \in \mathbf{R}^F, \boldsymbol{c}_r \in \mathbf{R}^G \tag{6}$$

其中:$\boldsymbol{Y}$ 为三阶预测张量;$\otimes$代表张量外积;$\boldsymbol{a}_r$、$\boldsymbol{b}_r$、$\boldsymbol{c}_r$ 分别为因子矩阵中的列向量。其目的是利用 3 个因子矩阵来恢复预测矩阵 $\boldsymbol{Y}$,因子矩阵分别对应于头实体、关系和尾实体在低维连续空间中的嵌入。将经过多模态数据融合得到的初始张量值作为因子矩阵的输入进行模型训练。对于预测矩阵中任意子元素 y_{ijk},该三元组(i,j,k)为真的概率的公式为:

$$y_{ijk} \approx \sum_{r=1}^{R} e_{ir} f_{jr} g_{kr},$$
$$i = 1,2,\cdots,m; j = 1,2,\cdots,n; k = 1,2,\cdots,t \tag{7}$$

其中:m、n、t 分别为头实体、关系、尾实体的数量;e_{ir}、f_{jr}、g_{kr}分别为因子矩阵中列向量 $\boldsymbol{a}_r$、$\boldsymbol{b}_r$、$\boldsymbol{c}_r$ 中某一个元素。

3.3　多预测任务的联合损失函数

为提升张量空间中的嵌入效果,模型引入联合损失函数,以提高训练效率。联合损失函数旨在处理链接预测任务,它有 2 种形式,即通过头节点和关系预测尾节点或通过头节点和尾节点预测关系。

将这 2 种损失求和得到基于 2 种预测任务的联合损失。负例三元组是通过将正确三元组中的尾实体替换为其他实体而得到的错误三元组。不同于传统方法随机选择部分实体进行替换,本文在训练中使用所有候选实体来替换正确三元组的尾实体,因此负例三元组的数量与实体数量相等,使得嵌入效果得到稳定提升。通过计算正确三元组的概率与负例三元组的损失,能够提升知识在低维稠密张量空间中的嵌入效果。计算尾节点预测损失 L_{e} 的公式为:

$$L_{\mathrm{e}} = -\frac{1}{n_{\mathrm{e}}} \sum_{i=1}^{n_{\mathrm{e}}} [y_{\mathrm{e}}^{(i)} \ln p_{\mathrm{e}}^{(i)} + (1 - y_{\mathrm{e}}^{(i)}) \ln(1 - p_{\mathrm{e}}^{(i)})] \tag{8}$$

其中:n_{e} 为实体节点数量;y_{e} 为真实尾实体的标签;p_{e} 为预测尾实体的概率。

计算关系的预测损失 L_{r} 的公式为:

$$L_{\mathrm{r}} = -\frac{1}{n_{\mathrm{r}}} \sum_{i=1}^{n_{\mathrm{r}}} [y_{\mathrm{r}}^{(i)} \ln p_{\mathrm{r}}^{(i)} + (1 - y_{\mathrm{r}}^{(i)}) \ln(1 - p_{\mathrm{r}}^{(i)})] \tag{9}$$

其中:n_{r} 为关系数量;y_{r} 为真实关系结果的标签;p_{r} 为预测关系的概率。

将式(8)与式(9)加权求和,得到联合损失函数 L_{a} 的公式为:

$$L_{\mathrm{a}} = \beta L_{\mathrm{r}} + (1 - \beta) L_{\mathrm{e}} \tag{10}$$

其中,超参数 β 是根据数据集中关系数量与实体数量的比值计算得出。

联合损失函数可以让模型同时通过 2 种预测损失进行计算,从而加速空间张量嵌入的拟合过程。

4　实验结果及分析

4.1　实验数据集

实验采用多模态知识图谱的基准数据集。FB15K-IMG 数据集是 Freebase[2] 数据集的一个子集,WN18-IMG 是 Wordnet[3] 数据集的一个子集,每个三元组包含头实体、关系、尾实体及对应图像。OpenBG-IMG[26] 数据集是电子商务领域的多模态知识图谱,该数据集中的每个三元组也包含头实体、关系、尾实体及对应图像。

FB15K-IMG、WN18-IMG 和 OpenBG-IMG 数据集具体参数见表 1 所列。

表 1　FB15K-IMG、WN18-IMG 和 OpenBG-IMG 数据集参数

数据集	实体数量	关系数量	训练集大小	验证集	测试集
FB15K-IMG	14 951	1 345	483 142	50 000	59 071
WN18-IMG	40 943	18	141 442	5 000	5 000
OpenBG-IMG	27 910	136	230 087	5 000	14 675

4.2　实验任务与基线模型

链接预测是知识图谱补全中的项任务,旨在预测缺失三元组中的特定实体,例如利用学习得到的张量嵌入来预测缺失三元组中的尾实体。实验采用 Hits@k 作为评估指标,Hits@k 表示在 k 个预测张量中出现正确实体的可能性,Hits@k 越高代表性能越好。

实验选择多模态嵌入模型 TransAE 和语义匹配模型 TorusE 作为基线模型,并且对比其他翻译模型(如 TransH 等)以及神经网络模型(如 ConvE 等)的效果。尽管其他多模态知识嵌入工作在多模态知识表征学习方面取得了一定成绩,但它们主要使用较为小众或未公开的多模态数据集。本实验采用了广泛使用的 FB15K-IMG、

WN18-IMG 和 OpenBG-IMG 数据集,许多现有的多模态知识表示研究工作都基于这 3 个数据集进行测试,具有代表性。

4.3 实验内容

模型在 FB15K-IMG、WN18-IMG 和 OpenBG-IMG 数据集上的实验结果见表 2 所列。实验使用 Hits@1、Hits@3、Hits@10 作为评估模型预测准确性的指标,这些指标分别表示在前 1、前 3和前 10 个预测结果中找到正确答案的比例。与多个基线模型相比,本文模型在 3 个数据集上均取得了最佳成绩或接近最佳的成绩。

在 FB15K-IMG 数据集上,本文模型相较于 TorusE 在 Hits@1、Hits@3,Hits@10 上分别提升了 3.8%、5.4%、7.5%,这表明引入多模态数据有助于知识图谱在空间中获得更好的嵌入效果。

在 WN18-IMG 数据集上,本文模型与 ConvE 和 TorusE 模型的性能基本相当,差距在 0.1%的范围波动,这可能是由于 WN18-IMG 的关系数量较少且语义环境简单,导致模型能迅速拟合嵌入关系,使得大多数模型都能取得较好效果。因此,在 WN18-IMG 数据集上,多模态数据的提升作用相对有限。

在 OpenBG-IMG 数据集上,本文模型取得了最优成绩,相较于 TransAE 在 Hits@1、Hits@3、Hits@10 上分别提升了 17.5%、18.3%、10.2%,说明本文模型在 OpenBG-IMG 的张量空间实现了较好的知识嵌入,同时强调了视觉信息融合在多模态知识表示中的重要性。TransAE 采用基础的 TransE 模型进行嵌入,而本文模型则采用更先进的张量分解嵌入方式,进一步证明嵌入方式在多模态表示学习中的重要性。

表 2 链接预测任务结果

模型	FB15K-IMG			WN18-IMG			OpenBG-IMG		
	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10
文献[10]TransE			0.471	0.040	0.745	0.923	0.150	0.387	0.647
文献[11]TransH			0.664				0.129	0.525	0.743
文献[15]DistMult	0.546	0.733	0.824	0.335	0.876	0.947	0.060	0.157	0.279
文献[16]ComplEx	0.599	0.759	0.840	0.936	0.945	0.947	0.143	0.244	0.371
文献[18]ConvE	0.558	0.723	0.831	0.935	0.946	**0.956**			
文献[13]TorusE	0.674	0.771	0.832	**0.943**	0.950	0.954			
文献[21]TransAE			0.645				0.274	0.489	0.715
本文	**0.712**	**0.825**	**0.907**	0.941	**0.951**	0.955	**0.449**	**0.672**	**0.817**

4.4 消融实验

本文模型的消融实验结果见表 3 所列,该实验涉及基础嵌入、图像编码和联合损失 3 个模块。基础嵌入模块使用文本数据生成张量嵌入;图像编码模块在基础嵌入上融入了清洗后的图像数据;联合损失模块则决定是否采用多预测任务损失。

通过多次实验得到引入不同模块下的平均实验指标结果。从表 3 可以看出,引入图像编码模块显著提升了知识表示的嵌入效果,在 3 个数据集上分别有 6.1%±0.6%、3.4%±0.3%、4.8%±0.5%的提升。这表明图像数据对知识嵌入具有巨大的潜力和价值,并且在复杂数据集上的表现更好,即在 FB15K-IMG 和 Openbg-IMG 数据集上的提升效果比在 WN18-IMG 上更为显著,这表明引入多模态数据能够为知识嵌入带来更多维度的特征,有助于知识嵌入时得到更好的约束和表达。

引入联合损失模块时,模型性能在 3 个数据集上的提升均较为有限,3 个数据集上分别提升 1.3%±0.2%、0.5%±0.1%、1.3%±0.1%,这表明引入联合损失函数,特别是关系损失确实可以提高关系嵌入在低维连续空间中的张量表达,但同时关系数量也成为影响联合损失函数方法有效性的关键因素,例如在 WN18-IMG 数据集上,关系数量占比远远小于实体数量的占比,因此在该数据集上提升更为有限。

当两个模块共同作用时,本文模型性能达到最优,相较于原版的基础嵌入,最终在 3 个数据集上分别提升了 8.1%、4.3%、6.5%的最佳结果,这表明两个模块的协同作用对模型性能提升具有相互促进的效果,并且对于复杂数据集时引入知识的多模态信息的整体优化效果会更加明显,说明实体的多模态信息丰富了实体与关系之间的内在联系,使多模态知识与三元组关系的嵌入更加精准。

表 3　本文模型的消融实验结果

模块			Hits@10		
基础嵌入	图像编码	联合损失	FB15K-IMG 数据集	WN18-IMG 数据集	OpenBG-IMG 数据集
√			0.826	0.912	0.752
√	√		0.887±0.006	0.946±0.003	0.800±0.005
√		√	0.839±0.002	0.917±0.001	0.765±0.001
√	√	√	0.907	0.955	0.817

4.5　嵌入维度与嵌入效果

有关向量嵌入维度与 Hits@10 的变化关系如图 3 所示。

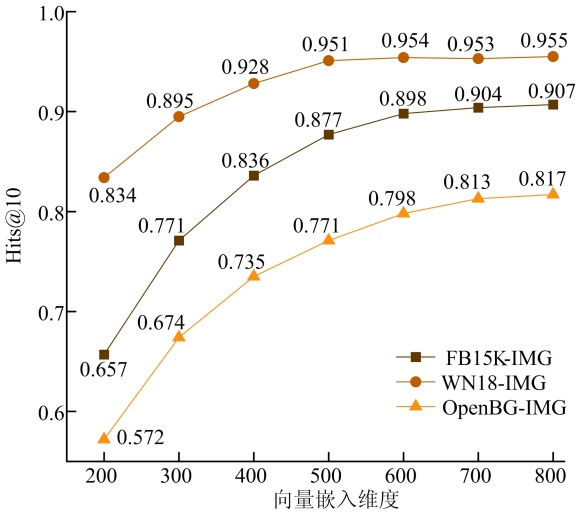

图 3　模型的维度与 Hits@10 指标的变化关系

从图 3 可以看出，Hits@10 指标表明预测结果的准确性与张量嵌入维度呈正相关。在 WN18-IMG 数据集中，当维度达到约 400 时，Hits@10 指标接近 95%，这表明模型在该维度下已充分表达，增加维度不会进一步提升嵌入效果；在 FB15k-IMG 数据集中，Hits@10 增长幅度随维度增加而稳步提升，当维度接近 750 时，Hits@10 的增长幅度逐渐放缓；同样地，OpenBG-IMG 在约 650 维时 Hits@10 达到平稳状态。3 个数据集的表现差异表明本文模型完全表达的维度与数据集复杂性正相关。

对于复杂数据集，完全表达所需的张量嵌入维度通常更高，例如 FB15K-IMG 和 OpenBG-IMG 数据集中的实体关系更复杂，需要更高维度的嵌入空间。相较之下，WN18-IMG 数据集关系更简单，传统单模态方法即可获得良好嵌入效果，即使加入多模态数据也只需较小维度即可达到最佳性能。

5　结　　论

本文提出一种基于张量分解的多模态知识表示学习模型，旨在解决当前多模态知识图谱中的数据异构性融合和多模态表示学习等挑战。该模型通过多模态数据筛选、多模态编码融合及多模态编码嵌入等步骤，实现了实体张量的嵌入。链接预测实验结果显示，该模型在 3 个基准测试集上均取得良好成绩，进一步证明通过合理的编码、筛选和融合方式处理多模态数据，能够显著提升知识图谱表示学习的效果。高质量的张量嵌入对于人们理解知识、挖掘知识以及执行相关下游任务具有重要贡献。

[参　考　文　献]

[1] LEHMANN J, ISELE R, JAKOB M, et al. Dbpedia: a large-scale, multilingual knowledge base extracted from wikipedia [J]. Semantic Web, 2015, 6(2): 167-195.

[2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//The 2008 ACM SIGMOD International Conference on Management of Data. [S. l.]: ACM, 2008: 1247-1250.

[3] MILLER G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.

[4] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[C]//The 1st International Conference on Learning Representations. [S. l.]: IMLS and NeurIPS Foundation, 2013: 1-12.

[5] MOUSSELLY H, BOTSCHEN T, GUREVYCH I, et al. A multimodal translation-based approach for knowledge graph representation learning[C]//The 7th Joint Conference on Lexical and Computational Semantics. [S. l.]: IEEE, 2018: 225-234.

[6] XIE R, HEINRICH S, LIU Z, et al. Integrating image-based and knowledge-based representation learning [J]. IEEE Transactions on Cognitive and Developmental Systems, 2019, 12(2): 169-178.

[7] ZHANG M, DAI R, DONG M, et al. DRLK: dynamic hierarchical reasoning with language model and knowledge

graph for question answering[C]//The 2022 Conference on Empirical Methods in Natural Language Processing. [S. l.]: ACL, 2022: 5123-5133.

[8] XIE R, LIU Z, LUAN H, et al. Image-embodied knowledge representation learning[C]//The 26th International Joint Conference on Artificial Intelligence. [S. l.]: AAAI, 2016: 1-7.

[9] ELLIOTT D. Adversarial evaluation of multimodal machine translation[C]//The 2018 Conference on Empirical Methods in Natural Language Processing. [S. l.]: ACL, 2018: 2974-2978.

[10] BORDES A, USUNIER N, GARCIA A, et al. Translating embeddings for modeling multi-relational data[C]//The 26th International Conference on Neural Information Processing Systems. [S. l.]: NIPS, 2013: 2787-2795.

[11] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]//The AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2014: 1112-1119.

[12] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//The AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2015: 345-354.

[13] EBISU T, ICHISE R. Toruse: knowledge graph embedding on a lie group[C]//The AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2018: 1955-1961.

[14] NICKEL M, TRESP V, KRIEGEL H. A three-way model for collective learning on multi-relational data[C]//The International Conference on Machine Learning. [S. l.]: IMLS, 2011: 809-816.

[15] YANG B, YIH W, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//The International Conference on Learning Representations. [S. l.]: IMLS, 2015: 1-12.

[16] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]//The International Conference on Machine Learning. [S. l.]: IMLS, 2016: 2071-2080.

[17] LACROIX T, USUNIER N, OBOZINSKI G. Canonical tensor decomposition for knowledge base completion[C]//The International Conference on Machine Learning. [S. l.]: IMLS, 2018: 2863-2872.

[18] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[C]//The AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2018: 1-14.

[19] NGUYEN D, NGUYEN T, NGUYEN D, et al. A novel embedding model for knowledge base completion based on convolutional neural network[C]//The 2018 Conference of the North American Chapter of the Association for Computational Linguistics. [S. l.]: ACL, 2018: 327-333.

[20] XIE R, LIU Z, LUAN H, et al. Image-embodied knowledge representation learning[C]//The 26th International Joint Conference on Artificial Intelligence. [S. l.]: AAAI, 2016: 3140-3146.

[21] WANG Z, LI L, LI Q, et al. Multimodal data enhanced representation learning for knowledge graphs[C]//The 2019 International Joint Conference on Neural Networks. [S. l.]: IEEE and INNS, 2019: 411-419.

[22] XU D, XU T, WU S, et al. Relation-enhanced negative sampling for multimodal knowledge graph completion [C]//The 30th ACM International Conference on Multimedia. [S. l.]: ACM, 2022: 3857-3866.

[23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//The International Conference on Learning Representations. [S. l.]: IMLS, 2014: 730-734.

[24] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//The IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2016: 770-778.

[25] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale[C]//The International Conference on Learning Representations. [S. l.]: IMLS, 2021: 1-22.

[26] DENG S, WANG C, LI Z, et al. Construction and applications of billion-scale pre-trained multimodal business knowledge graph[C]//The IEEE 39th International Conference on Data Engineering. [S. l.]: IEEE, 2023: 2988-3002.

(责任编辑 胡亚敏)