

关系敏感型多子图图神经网络的多模态实体对齐^①



金佳惠^{1,2}, 李治江^{1,2}, 刘谊章^{1,2}

¹(武汉大学 信息管理学院, 武汉 430072)

²(武汉大学 图像传播与印刷包装研究中心, 武汉 430072)

通信作者: 李治江, E-mail: lizhijiang@whu.edu.cn

摘要: 作为融合多源异构知识图谱的主要手段, 实体对齐一般首先编码实体等图结构信息, 而后通过计算实体间相似度来获取对齐实体. 然而, 现存的多模态对齐方法往往直接引入预训练方法表达模态特征, 忽略了模态间的融合以及模态特征与图结构间的融合. 因此, 本文提出一种关系敏感型的多子图图神经网络 (RAMS) 方法. 通过多子图图神经网络编码方法对模态信息与图结构进行结合并获得实体表征, 通过跨域相似度计算得到对齐结果. 广泛且多角度的实验证明了本文所提出的模型在准确率、效率、鲁棒性方面均超过了基线模型.

关键词: 多模态实体对齐; 图神经网络; 知识图谱; 机器学习; 深度学习

引用格式: 金佳惠, 李治江, 刘谊章. 关系敏感型多子图图神经网络的多模态实体对齐. 计算机系统应用, 2024, 33(3): 245-254. <http://www.c-s-a.org.cn/1003-3254/9437.html>

Multimodal Entity Alignment Based on Relation-aware Multi-subgraph Graph Neural Network

JIN Jia-Hui^{1,2}, LI Zhi-Jiang^{1,2}, LIU Yi-Zhang^{1,2}

¹(School of Information Management, Wuhan University, Wuhan 430072, China)

²(Research Center of Graphic Communic, Printing and Packaging, Wuhan University, Wuhan 430072, China)

Abstract: Multi-modal entity alignment (MMEA) is a crucial technique for integrating multi-source heterogeneous multi-modal knowledge graphs (MMKGs). This integration is typically achieved by encoding graph structure and calculating the plausibility of multi-modal representation between entities. However, existing MMEA methods tend to directly employ pre-trained models and overlook the fusion between modalities as well as the fusion between modal features and graph structures. To address these limitations, this study proposes a novel approach called relation-aware multi-subgraph graph neural network (RAMS) for obtaining multi-modal representation in the context of entity alignment. RAMS utilizes a multi-subgraph graph neural network for fusing modality information and graph structure to derive entity representation. The alignment results are subsequently obtained through cross-domain similarity calculation. Extensive experiments demonstrate that RAMS outperforms baseline models in terms of accuracy, efficiency, and robustness.

Key words: multimodal entity alignment; graph neural network (GNN); knowledge graph; machine learning; deep learning

知识图谱, 最初由谷歌公司于 2012 年提出^[1], 旨在使用实体、概念及其关系形成的巨大语义网络描述真实世界. 目前, 已经有许多大型知识图谱被构建, 如由谷歌、维基百科等公司开源的 DBpedia^[2], YAGO^[3]等通用知识图谱, 以及特定领域如医疗^[4]、科学^[5]领域知

识图谱. 然而, 不同知识图谱之间的质量和覆盖领域存在差异, 为了增强知识图谱完整性和知识覆盖度, 并融合多个异构图谱之间的知识, 实体对齐方法 (entity alignment, EA) 被提出^[6,7], 用于整合来自不同知识图谱的信息, 以补充和扩展现有的知识图谱.

① 基金项目: 国家重点研发计划 (2021YFB3900903)

收稿时间: 2023-09-07; 修改时间: 2023-10-20; 采用时间: 2023-11-03; csa 在线出版时间: 2024-01-18

CNKI 网络首发时间: 2024-01-22

传统的实体对齐方法基于特征工程,即通过实体的文字信息(如实体名称、实体描述等)来判断实体是否一致.然而,由于重名、别名、跨语种等多种因素,基于特征的方式缺乏鲁棒性和可扩展性.基于深度学习的方法一般分为基于翻译的方法和基于图神经网络(graph neural network, GNN)的方法^[8],它们将不同图谱采用表示学习方法嵌入到低维向量空间^[9],并通过两个实体向量之间的距离来确定它们是否匹配.基于这种思想, MTransE、BootEA^[10]、MuGNN^[11]等被提出用于解决实体对齐问题.随着多模态知识图谱得到广泛应用,多模态实体对齐任务(multi-modal entity alignment, MMEA)也被进一步提出,如图1所示.

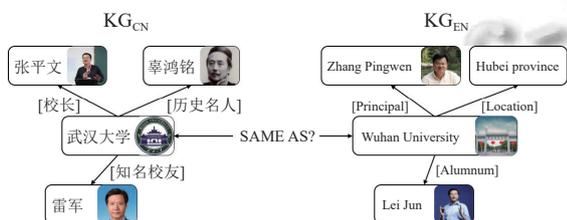


图1 多模态知识图谱中的实体对齐任务

以 DBP15k (ZH-EN) 为例,统计数据集内各度数区间的节点数量,如图2所示,大部分节点的度数在 0-20 之间,只有少数节点包含丰富的边信息.这些度数较低的中心节点往往特征表达能力较差,因而属于实体对齐任务中较难被对齐的节点,其对齐效果显著低于高度数区间^[12].

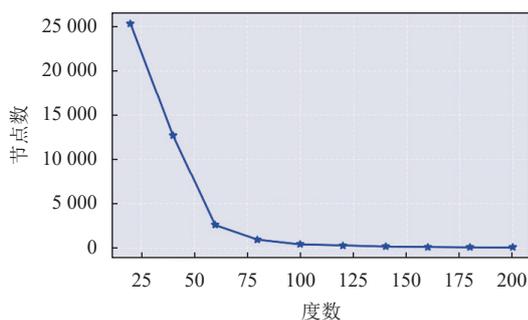


图2 DBP15k (ZH-EN) 数据集节点度数统计图

另一方面,在引入了图像信息的 DBP15k 数据集中,除了约 60%–85% 的实体外,仍然存在很多实体缺乏图像表征,如表1所示.在多模态实体对齐中,这部分实体的图像表征往往被直接地以正态分布中采样的随机向量来表示,由于缺少对邻居实体图像信息的引入而缺乏准确性^[13].

表1 DBP15k 的图像覆盖情况统计表^[13]

数量	FR-EN		JA-EN		ZH-EN	
	FR	EN	JA	EN	ZH	EN
图像数量	14 174	13 858	12 739	13 741	15 912	14 125
实体数量	19 661	19 993	19 814	19 780	19 388	19 572

尽管大多数模型已经在一些公开数据集如 DBP-15k^[14]、DWY100K^[10]、SRPRS^[15]上实现了 80% 以上的准确率,但许多方法只是直接地使用了预训练模型编码多模态表征,使邻域较少或图像模态特征缺失的节点获得的是图结构无关或弱关联的特征.

基于以上分析,本文提出一种端到端的多模态实体对齐方法(relation-aware multi-subgraph graph neural network, RAMS),通过图神经网络实现模态信息与图结构信息的融合,通过多子图实现不同模态间的信息交互,并通过基于关系注意力的子图采样平衡节点特征表达.在开源数据集上的实验表明,本文所提出的方法在效果、运行效率、鲁棒性、长尾节点等多个方面均具有优异表现.

1 实体对齐相关工作

1.1 基于深度学习的实体对齐方法

现有的知识图谱实体对齐方法通常采用先编码后对齐的思想.首先,模型学习图谱内实体的表征,然后通过映射矩阵或实体交换等策略对齐两个图谱的语义空间,并最终通过余弦相似度、曼哈顿距离、欧几里得距离等来判断实体之间的相似性.

基于深度学习的实体对齐方法可以分为 TransE 系列方法和 GNN 系列方法.基于 TransE 的图谱编码方法通常将关系视作从头实体到尾实体的平移,即 $h+r \approx t$. MTransE^[6]是 TransE 的一种变体,它用于编码不同语言的图谱,并对齐这两个语义空间. COTSAE^[16]引入了属性对比联合学习模块来编码实体的语义信息. JAPE^[14]引入了负样本学习,并联合学习属性预测任务和对齐任务.

TransE 系列方法只在单个三元组上进行训练,缺乏对全局信息的利用能力,而这一点在基于 GNN 的图谱编码方法中得到了改善. GCN-Align^[17]是第 1 个将 GNN 应用于实体对齐任务的方法,然而由于普通图卷积网络在建模中的缺陷,该方法无法有效地利用图谱中丰富的关系信息. MuGNN^[11]认为图谱间的结构差异会影响对齐效果,因此首先通过规则补全来调和异构

图谱之间的结构差异,然后通过多通道 GNN 对图谱进行编码. NAEA^[18]和 MRAEA^[19]根据实体之间的关系类型赋予实体不同的权重系数,使模型能够定义有用的邻居节点用于消息传递.此外,还有一些 GNN 方法^[11,20,21],如 MuGNN 等在变换矩阵的设计中采用了反直觉的约束,即强迫矩阵为单位或对角,这样可以保证在线性转换后异构图谱之间的距离分布不会发生改变.

1.2 基于深度学习的多模态实体对齐方法

与常规实体对齐任务相比,多模态实体对齐涉及额外的步骤,即多模态特征融合.在多模态嵌入阶段,实体具有多种模态的特征,这样可以产生相对应的实体向量表示.然后,使用收敛或加权级联等方法将这些表示进行组合,生成实体的多模态融合特征.利用得到的实体表示来计算要对齐的实体之间的相似性得分,从而预测实体的对齐结果.

PoE (product of experts)^[22]最早被提出,通过结合多个专家模型的预测来获得实体对齐的置信度.为了表征实体不同模态的信息, MMEA^[23]正式提出了多模态实体对齐任务,并采用 TransE、VGG16 和 RBF 模型分别表征图结构、图像、属性特征,通过强制各模态特征与融合特征相似来对齐实体. HMEA^[24]采用 GCN 和 DenseNet,并在双曲空间完成了 MMEA 任务. EVA^[13]提出了一种无监督的方法来生成视觉中枢,并使用加权级联的方法组合实体的多模态特征. MCLEA^[25]使用类似的嵌入器编码实体,并设计了模态间和模态内的损失来实现模态融合和对齐.此外, MEAformer^[26]引入了 Transformer 结构,以增强模态融合的效果.

上述方法^[13,23-25]通过直接地采用预训练的图像模型,如 VGG16、DenseNet、ResNet 等,较快地实现了

实体多模态信息的表征.然而,这些方法大多更关注于单一模态的表达,通过预训练和前馈层获得的模态特征往往直接通过强制嵌入相似或加权级联的方法得到实体的最终表征,而忽略了多模态特征与图的结构及其邻域间的相关性,这限制了实体对齐任务的发展.

尽管 MSNEA^[27]在向多模态表征中引入图结构信息方面做出了尝试,即将实体的多模态信息与 TransE 型的图结构表征方法相结合,通过令实体在每个模态下的表征满足 $h+r \approx t$ 的约束,实现模态信息与图结构信息的融合.但 TransE 在复杂关系处理方面具有局限性,这限制了模型的表现.

2 多子图信息融合的多模态实体对齐方法

规定一个多模态知识图谱 $G = \{E, R, I, T^R, T^V\}$, 包括实体集 E , 关系集 R , 图像集 I , 及关系三元组集 T^R , 实体图像对 T^V . 每个三元组 $t^R = (e_i, r_{ij}, e_j) \in T^R$ 表示头实体 e_i 能够通过关系 $r_{ij} \in R$ 连接至 e_j . 每个实体图像对 $t^V = (e_i, v_i) \in T^V$ 表示图像 v_i 能够作为实体 e_i 的描述.

给定两个图谱 G_1 和 G_2 , 有已知的对齐种子实体 $S = \{(e_1, e_2) \in E_1 \times E_2 | e_1 \leftrightarrow e_2\}$. 实体对齐的任务就是通过 S 、 G_1 和 G_2 找到未在 S 中的对齐种子实体.

我们所提出的方法的框架如图 3 所示. 对输入的图谱,通过采样生成子图,并使用多视图聚合编码的方式将各模态信息输入多子图正交图神经网络. 具体而言,采样后的子图经邻域聚合、关系聚合和图像聚合,得到各节点的表征. 这些表征经过正交映射图神经网络编码后,进行多跳拼接形成节点向量表示. 编码后的关系被用于计算关系注意力,并支持子图采样过程,最后进行联合学习预测对齐节点.

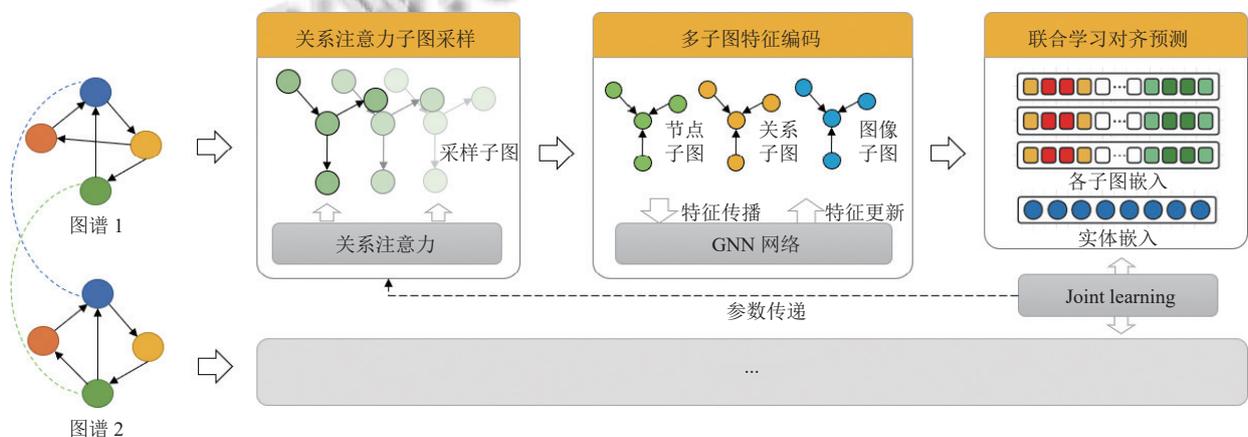


图 3 基于多子图图神经网络的多模态实体对齐方法

2.1 多子图增强的节点特征编码方法

多子图增强的特征编码网络如图4所示。

(1) 邻域聚合编码

对输入的图谱 $G=\{E, R, I, T^R, T^I\}$, 从三元组 (e_i, r_{ij}, e_j) 中保留实体对 (e_i, e_j) , 则节点的初始邻域特征按式(1)进行计算:

$$\mathbf{h}_{e_i}^{(0)} = \left[\frac{1}{|N_{e_i}^e| + 1} \sum_{e_j \in N_{e_i}^e \cup \{e_i\}} \mathbf{h}_{e_j} \right] \quad (1)$$

其中, $e_i \in E$ 表示当前节点, $N_{e_i}^e$ 表示节点 e_i 的所有邻居节点, $\mathbf{h}_{e_i}^{(0)}$ 表示汇聚了邻域信息的节点特征向量. \mathbf{h}_{e_j} 为节点 e_j 参数化的特征, 将在迭代中不断更新。

将 $\mathbf{h}_{e_i}^{(0)}$ 拼接得到邻域聚合的初始节点特征矩阵

$\mathbf{H}^{(0)}$, 用于输入至关系敏感型正交 GNN 网络中. 每层 GNN 网络更新过程如式(2)所示:

$$\mathbf{H}^{(l)} = GNN(\mathbf{H}^{(l-1)}) \quad (2)$$

其中, $\mathbf{H}^{(l-1)}$ 代表输入第 l 层 GNN 的节点特征矩阵; $\mathbf{H}^{(l)}$ 代表第 l 层 GNN 输出的节点特征矩阵 (同时作为第 $l+1$ 层 GNN 的输入)。

为了得到全局特征表示, 对多层 GNN 的输出特征进行拼接, 多跳拼接的邻域聚合节点特征表示为 $\mathbf{h}_{e_i}^{\text{entity}}$, 如式(3)所示:

$$\mathbf{h}_{e_i}^{\text{entity}} = [\mathbf{h}_{e_i}^{(0)} \parallel \mathbf{h}_{e_i}^{(1)} \parallel \dots \parallel \mathbf{h}_{e_i}^{(l)}] \quad (3)$$

其中, $\mathbf{h}_{e_i}^{(l)}$ 代表邻域子图下, 第 l 层 GNN 输出的 e_i 特征, $\mathbf{h}_{e_i}^{(0)}$ 为初始的邻域表征。

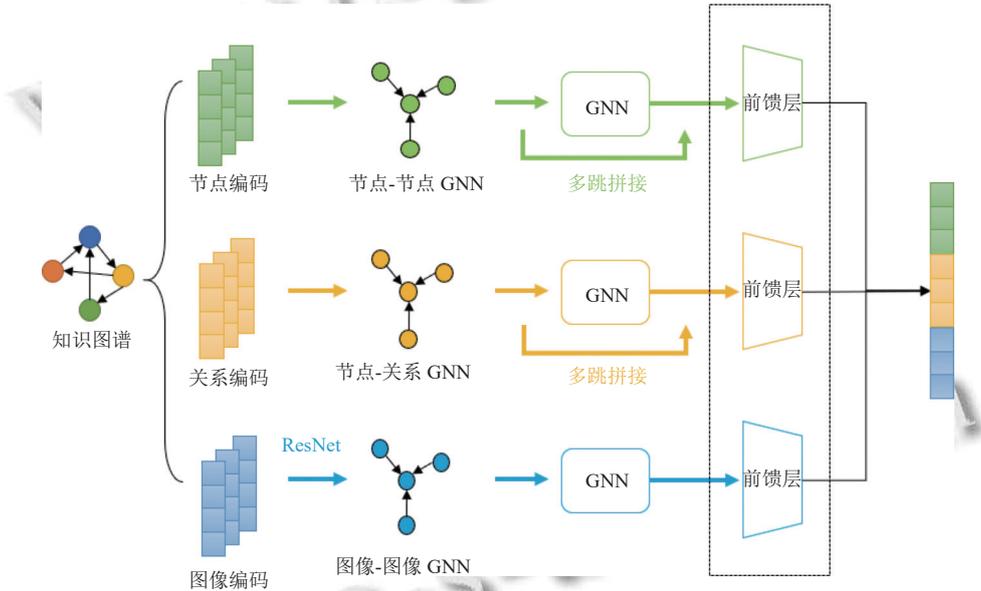


图4 多子图增强的节点特征编码方法结构图

(2) 关系聚合编码

与邻域聚合编码相类似, 从三元组 (e_i, r_{ij}, e_j) 中保留实体-关系对 (e_i, r_{ij}) 和 (e_j, r_{ij}) , 模仿空间域图神经网络传递方式, 将边信息也以聚合方式传递给关联节点, 如式(4)所示:

$$\mathbf{h}_{e_i}^{(0)} = \left[\frac{1}{|N_{e_i}^r|} \sum_{r_j \in N_{e_i}^r} \mathbf{h}_{r_j} \right] \quad (4)$$

其中, $e_i \in E$ 表示当前节点, $r_j \in N_{e_i}^r$ 表示节点 e_i 的相邻边, $\mathbf{h}_{e_i}^{(0)}$ 表示汇聚了邻边信息的节点特征向量. 边特征 $\mathbf{h}_{r_j} \in \mathbf{H}^R$ 为边 r_j 的参数化表示, 将在迭代中不断更新。

同上, 将边子图信息输入正交 GNN 网络, 并多跳链接生成关系聚合节点特征 $\mathbf{h}_{e_i}^{\text{relation}}$ 。

(3) 图像聚合编码

本文选取在 ImageNet^[28] 的目标识别任务上预训练的 ResNet-152 作为图像特征提取器, 取该模型最后一层全连接层及 Softmax 层前的输出特征作为图像表征, 如式(5)所示:

$$h_{e_i}^V = W_V \cdot \text{ResNet}(v_i) + b_V \quad (5)$$

其中, v_i 代表输入图像, $h_{e_i}^V$ 代表图像特征输出, W_V 、 b_V 为可训练参数. 对于没有图像的实体, 本文分配一个从正态分布中采样的随机向量, 由其他图像的平均值和

标准偏差进行参数化.

其次,同“节点聚合特征编码”步骤相同,从三元组 (e_i, r_{ij}, e_j) 中保留实体对 (e_i, e_j) , 并构建 (v_i, v_j) 对应二元组关系, 进行如式 (6) 所示的图像聚合节点特征:

$$\mathbf{h}_{e_i}^{(0)} = \left[\frac{1}{|N_{e_i}^v|} \sum_{v_j \in N_{e_i}^v} \mathbf{h}_{v_j} \right] \quad (6)$$

其中, e_i 表示当前节点, \mathbf{h}_{v_j} 代表节点 e_j 的图像特征, $N_{e_i}^v$ 表示 e_i 节点的所有相邻图像节点.

同上, 将视觉子图信息输入正交 GNN 网络生成最终图像特征传递节点表示. 由于图像特征与图结构特征不同, 两跳以上的节点聚合可能会引入过多的噪声与不相似图像, 因此, 图像聚合编码仅使用一层 GNN, 并以该层输出作为图像聚合的节点特征 $\mathbf{h}_{e_i}^{\text{visual}}$.

节点 e_i 经每个子图信息聚合后所得特征分别为 $\mathbf{h}_{e_i}^{\text{entity}}$, $\mathbf{h}_{e_i}^{\text{relation}}$, $\mathbf{h}_{e_i}^{\text{visual}}$. 3 类特征经可训练权重平衡得到用于对齐的实体表征 f_{e_i} , 如式 (7) 所示:

$$f_{e_i} = \left[w_1 \cdot \mathbf{h}_{e_i}^{\text{entity}} \parallel w_2 \cdot \mathbf{h}_{e_i}^{\text{relation}} \parallel w_3 \cdot \mathbf{h}_{e_i}^{\text{visual}} \right] \quad (7)$$

其中, 可训练权重 w_1 、 w_2 、 w_3 在运算前进行了 Softmax, 以保证权重和为 1.

2.2 基于关系注意力的知识图谱子图采样方法

受 GraphSAGE^[29] 启发, 本文采用一种小批量 mini-batch 的子图采样方式, 将知识图谱训练节点切分为多个子块进行训练. 更进一步, 由于三元组包含信息的质量不一, 对当前节点的影响程度也不一, 本文采用基于关系注意力计算的子图采样方法取代了平均随机采样, 实现非均匀概率采样.

对中心节点 e_i 进行采样时, 三元组 (e_i, r_k, e_j) 的路径采样概率记为 $P((e_i, r_k, e_j)|e_i)$, 令其等同于路径的注意力值 α_{ijk} . 对每个中心节点, 其采样期望为各路径权重与被采样概率相乘的加权和 (也可以用作衡量模型的选择能力), 如式 (8):

$$E[\alpha_{e_i}] = \sum_{e_j \in N_{e_i}^e} \sum_{r_k \in R_{ij}} P((e_i, r_k, e_j)|e_i) \alpha_{ijk} = \sum_{e_j \in N_{e_i}^e} \sum_{r_k \in R_{ij}} \alpha_{ijk}^2 \quad (8)$$

如果使用固定的采样次数 t 或简单地设定 t 与节点的度成反比, 这会导致部分节点与路径被过度采样或忽视, 进一步造成信息的冗余与不足. 因此, 我们期望对任一中心节点 e_i , 其注意力期望值之和在实验结束后应保持一定的稳定性, 那么在每个训练周期内采

样次数 t_{e_i} 应符合如下公式:

$$t_{e_i} = \left\lceil \frac{\tau}{E[\alpha_{e_i}]} \right\rceil \quad (9)$$

其中, τ 为控制采样率的超参数, $\lceil x \rceil$ 为向上取整操作.

基于期望加权计算的子图采样方法通过使采样次数与节点注意力期望成反比, 平衡了模型对各节点的学习效果, 并减少了由于对部分中心节点多次采样所造成的信息冗余.

2.3 联合学习损失函数

2.3.1 多子图特征融合损失函数

受 NCA^[30] 文本-图像匹配方法启发, 本文在融合实体的邻域、结构、图像信息时, 采用了一个与 NCA 损失类似的函数, 以增加对样本的区分能力.

新的多模态融合损失函数使用了局部和全局的统计数据衡量样本的重要程度, 为解决嵌入时的枢纽度问题, 采用软加权方案. 如式 (10) 所示:

$$L_{\text{NCA}} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{\alpha} \log \left(1 + \sum_{m \neq i} e^{\alpha S_{mi}} \right) + \frac{1}{\alpha} \log \left(1 + \sum_{n \neq i} e^{\alpha S_{in}} \right) - \log(1 + \beta S_{ii}) \right) \quad (10)$$

其中, α 、 β 为超参数, 本文中选取 α 为 15, β 为 10, N 为每个 batch 中的枢轴数. S_{xy} 代表其下标两个向量的余弦相似度, 如 S_{ii} 为实体 i 两种模态表征的相似度, S_{mi} 与 S_{in} 为实体 i 的某个模态表征与另一实体的另一模态表征. 对实体 i 的两种模态表征, 预期该实体的每个表征与其他实体在另一模态的表征尽可能有差异, 同一实体的两种模态表征尽可能相似.

2.3.2 无负样本对比损失函数

两图谱间的对齐采用无负样本对比损失函数^[12]. 具体而言, 图谱经过 RAMS 前向传播得到图谱特征 f , 同时记录一个无梯度的副本, 记为 f^{init} . 在编码后, 双方互相向对方的特征 f^{init} 进行学习, 计算式 (11) 所示损失:

$$L_{\text{NFC}} = \sum_{(e_i, e_j) \in S} (-\text{sim}(f_{e_i}, f_{e_j}^{\text{init}}) - \text{sim}(f_{e_j}, f_{e_i}^{\text{init}})) \quad (11)$$

实体对齐的置信度采用式 (11) 的形式进行计算:

$$\text{score}(e_i, e_j) = \text{sim}(f_{e_i}, f_{e_j}^{\text{init}}) + \text{sim}(f_{e_j}, f_{e_i}^{\text{init}}) \quad (12)$$

2.3.3 联合学习损失函数

联合学习损失函数如下:

$$L_{\text{Joint}} = \sum_i^n L_{\text{NCA}} + L_{\text{NFC}} \quad (13)$$

其中, $\sum_i^n L_{\text{NCA}}$ 代表所有模态的多模态融合损失相加结果, L_{NFC} 代表无负样本损失. 由于无负样本损失在编码过程中已经将所有模态进行融合, 因此无需进行模态间的求和计算.

2.3.4 跨域相似度计算

通过两个实体间的相似度直接地判断对齐可能性是较为粗暴的方法. 因为最近邻关系是不对称的: y 是 x 的最近邻 K 个实体并不意味着 x 是 y 的最近邻 K 个实体.

受 CSLS^[31] 启发, 我们令两个实体间的相似度受它们与其他实体间的相似度影响. 具体地, 考虑一个二分图, 分别为源图谱 KG_S 和目标图谱 KG_T . $N_T(x_i)$ 表示目标图谱中与源图谱实体 x_i 相似度最高的 K 个实体集合. 类似地, $N_S(y_j)$ 表示源图谱中与目标图谱实体 y_j 相似度最高的 K 个实体集合. 对于源图谱中的实体 x_i , 其与目标图谱中 K 个近邻间的平均相似度计算如式 (13):

$$avg_T(x_i) = \frac{1}{K} \sum_{y_j \in N_T(x_i)} \cos(\mathbf{h}_{x_i}, \mathbf{h}_{y_j}) \quad (14)$$

其中, \cos 计算的是向量余弦相似度. 如果实体 x_i 和另一个图谱中的 K 近邻实体都较为相似, 那么 $avg_T(x_i)$ 就会具有较高的值, 并会因此影响它与待对齐实体间的相似程度. 源图谱实体 x_i 与目标图谱实体 y_j 的跨域相似度计算如式 (14):

$$\text{sim}(x_i, y_j) = 2\cos(x_i, y_j) - avg_T(x_i) - avg_S(y_j) \quad (15)$$

此外, 本文采用了双向迭代策略来构造高质量的半监督对齐对. 根据 RAMS 跨域相似度计算获得的新的对齐节点, 当且仅当实体 e_i 和 e_j 是彼此最近的邻居时, (e_i, e_j) 将被视为新的半监督对齐实体.

3 实验分析

3.1 评价指标及实验设置

遵循现有的实体对齐方法, 本文采用 Hits@ k ($k=1, 10$) 和平均倒数排名 (MRR) 作为评价指标. 在预测阶段, 对于每个源实体, 目标实体将按其与源实体的距离分数进行升序排列. Hits@ k 反映前 k 个预测对齐实体中正确对齐实体的比例. MRR 表示正确对齐实体排名的倒数的平均值. Hits@ k 和 MRR 分数越高表明性能越好.

对于所有数据集, 本文都设置了相同的默认配置: 嵌入维度 $d=300$, GNN 层数 $l=2$, 采样率 $\tau=1$, 相似性阈值 $\epsilon=0.05$, 扩散系数 $\alpha=15$, $\beta=10$, 批量大小为 512, dropout 率设定为 30%, 采用 RMSprop 对模型进行优化, 学习率设置为 0.005.

3.2 多模态实体对齐实验结果与分析

为验证模型有效性, 本文选择近几年表现出色的公开模型进行对比. 为了保证模型之间比较的公平性, 将模型分为两组, 基础组 (Basic) 和数据增强组 (iterative learning, IL), 其中基础组仅使用了数据集原本的信息且并未使用任何数据增强方法, 数据增强组通过无监督或半监督方法获取了额外信息进行训练. 实验结果表 2 所示.

表 2 DBP15k 数据集上实体对齐方法对比结果 (%)

组别	对齐方法	DBP (ZH-EN)			DBP (JA-EN)			DBP (FR-EN)		
		Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
基础组	MTransE	30.8	61.4	36.4	27.9	57.5	34.9	22.4	55.6	33.5
	JAPE	41.2	74.5	49.0	36.3	68.5	47.6	32.4	66.7	43.0
	GCN-Align	43.4	76.2	55.0	42.7	76.2	54.0	41.1	77.2	53.0
	MuGNN	49.4	84.4	61.1	50.1	85.7	62.1	49.5	87.0	62.1
	RSN	50.8	74.5	59.1	50.7	73.7	59.0	51.6	76.8	60.5
	RAMS (Basic)	71.0	88.5	78.8	70.7	88.6	78.6	71.7	89.0	79.4
数据增强组	BootEA	62.9	84.7	70.3	62.2	85.3	70.1	65.3	87.4	73.1
	NAEA	65.0	86.7	72.0	64.1	87.2	71.8	67.3	89.4	75.2
	TransEdge	73.5	91.9	80.1	71.9	93.2	79.5	71.0	94.1	79.6
	EVA	75.7	93.0	82.7	75.8	93.4	82.6	78.1	94.8	84.9
	MRAEA	76.1	90.7	81.4	76.2	91.3	81.7	79.3	94.2	84.7
	RAMS (IL)	82.2	93.0	87.1	82.0	93.1	87.0	84.8	94.7	89.4

从实验结果中可以看出:

1) 考虑不使用迭代的模型, RAMS (Basic) 明显优

于以往的神经网络模型, 并在 Hits@1 指标上都超越了次优模型 10%–20% 的水平. 相比于 GCN-Align、MuGNN

等 GNN 模型, RAMS 融合了多模态特征, 在未经迭代的情况下具有很好的冷启动特征优势, 说明了子图特征融合的有效性。

2) 考虑使用迭代模型的情况下, RAMS (IL) 同样优于所有基线模型, 在 Hits@1 上超越了次优模型 4%–5% 的水平。相比于 BootEA、NAEA 等使用了负采样方式的模型, RAMS 采用对比无负采样损失进行训练, 提升了模型效率也同时保证了模型的准确率; 相比于 EVA、MRAEA 等多模态实体对齐模型, 本文方法采用了多模态子图方式编码, 效果均优于普通单独编码融合方法。

3) 对比所有方法, 使用迭代对齐策略的模型整体优于上半部分模型的表现, 本文方法中使用了迭代数据增强策略后, 整体 Hits@1 水平也上升了将近 10%, 验证了在实体对齐任务中使用迭代对齐等数据增强方式对模型的提升较大。在 MRAEA 与本文方法中, 都使用了双向迭代策略或无监督数据增强策略, 可以看出模型准确率有进一步的提升。

3.3 模型鲁棒性实验结果与分析

本文从 3 个角度测试实验参数对模型稳定性的影响。

(1) 预对齐种子比率

预对齐种子占比指训练集种子在所有标注实体对中的占比。实验采用 10%、20%、30%、40% 这 4 个占比进行实验。预对齐种子占比实验对比结果如图 5 所示。

本文选取 RREA、MuGNN、MRAEA 这 3 种表现较为出色的基线模型作为 RAMS 的对比。从图 5 中可以看出, 本文模型在各种预对齐种子占比率的表现都十分优秀, 在仅有 10% 对齐种子的情况下, 模型表现也能够达到 73%–75% 的 Hits@1 值, 这相比于 RREA、MuGNN、MRAEA 这 3 种模型在 40% 对齐种子的情况还要出色, 验证了本文即使在少量样本的情况下也可以获得出色的表现, 在数据输入方面的鲁棒性较高。

(2) 子图采样大小

子图采样大小指每个批次输入模型的子图采样节点数。从图 6 中可以看出, 在 128–1024 的采样大小情况下, RAMS (IL) 的 Hits@1 水平波动较小, 验证了模型在不同采样率情况下的鲁棒性。而模型运行时间则随着采样数量增加而大幅度减小, 一方面由于随着采

样子图增长, 训练轮次随之减小, 模型处理时间下降, 另一方面, 子图采样增加了模型并行性, 在显存的负担能力以内能够增强运行效率。

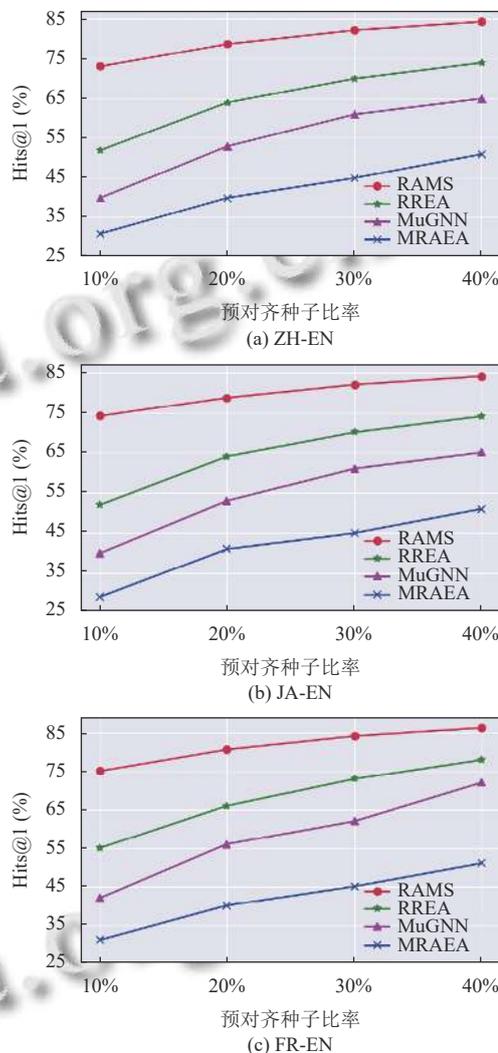


图 5 以 DBP15k 为例, 预对齐种子占比率对实验结果的影响

(3) 网络层数大小

网络层数指邻域、结构子图中 GNN 神经网络层数。图 7 展示了随着网络层数加深, Hits@1 水平波动也相对较小, 只有在一层时由于消息传递不足, 使得模型表现较弱, 验证模型在图神经网络层数方面的鲁棒性。此外, 随着层数增加, 由于模型复杂度上升, 模型运行时间也随之增加, 因此本文仅选用两层深度网络, 在较少的时间内, 保障模型的准确率。

3.4 模型效率实验结果与分析

本文进一步验证了 RAMS 在运行效率上的优势, 与各基线方法的对比结果如表 3 所示。

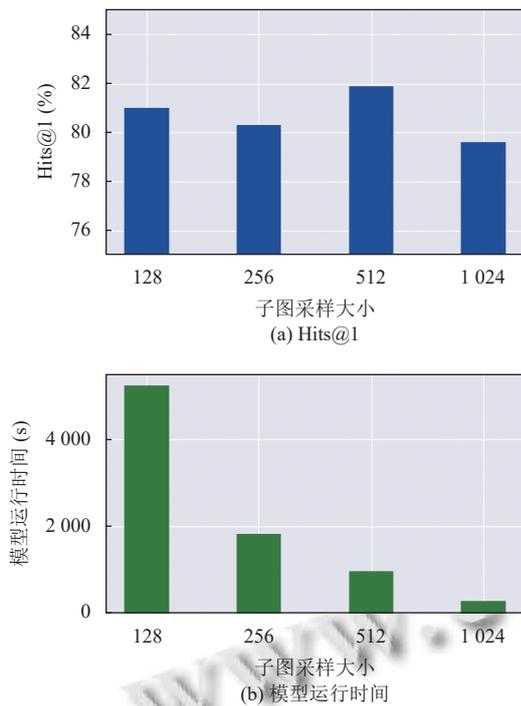


图6 以 DBP15k (ZH-EN) 为例, 不同子图采样大小对 Hits@1 及运行时间的影响

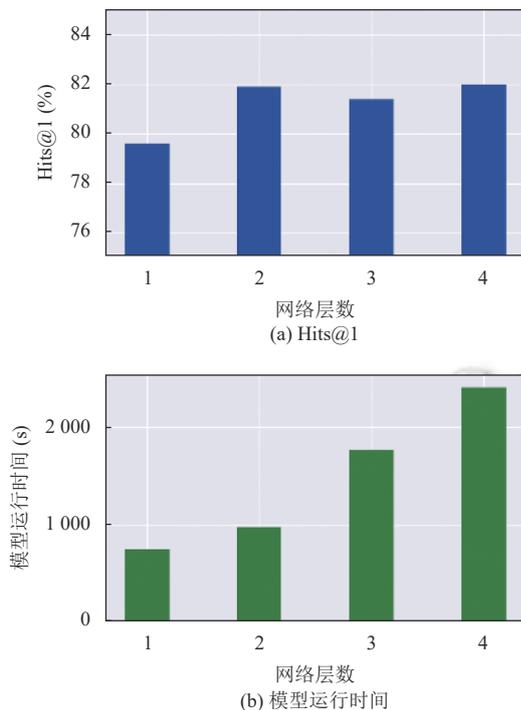


图7 以 DBP15k (ZH-EN) 为例, 不同网络层数对 Hits@1 及运行时间的影响

从未使用迭代策略的基础组看, 本文方法仅次于 GCN-Align, 且时间差距不大. 但由于 GCN-Align 仅使用了图谱结构信息, 且仅由双层基础 GCN 编码, 而本

文考虑多子图多模态信息, 在同一量级的时间内, 本文方法的准确率远超前者. 对比其他复杂模型, 本文方法在保证准确率的同时, 时间花费上大幅减小, 验证了本文方法对模型效率的提升.

表3 DBP15k 数据集上模型效率实验结果

组别	对齐方法	Time (s)	Hits@1 (%)
基础组	MTransE	6467	30.8
	GCN-Align	103	43.4
	MuGNN	3156	49.4
	RSN	7539	50.8
	RAMS (Basic)	140	71.0
数据增强组	BootEA	4661	62.9
	NAEA	19115	65.0
	TransEdge	3629	73.5
	MRAEA	3894	76.1
	RAMS (IL)	986	82.2

从使用了迭代策略的数据增强组来看, RAMS 同样在时间上大幅超越所有基线模型, 对比 NAEA、TransEdge、MRAEA 等使用了多模态信息的模型, 本文方法在低时间复杂度的情况下获得了较高的实体对齐准确率表现.

3.5 长尾节点实验结果与分析

将 DBP15k (ZH-EN) 中所有节点按照节点度数进行划分, 大部分节点的度数在 0-25 之间, 仅有少数节点包含丰富边信息. 这些邻居节点较少的中心节点往往特征表达能力一般较差. 分别计算所划分的各度区间内 Hits@1 值, 实验结果如图 8 所示.

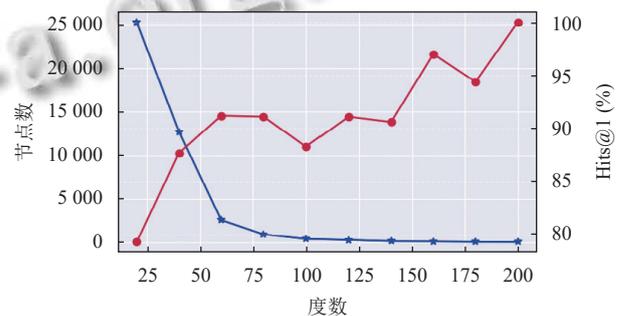


图8 以 DBP15k (ZH-EN) 为例, 各度数区间节点数及 Hits@1 结果图

可以看出, 模型在各区间内都具有良好的表现. 在度数小于 20 的稀疏节点区间, 本文模型的表现能达到 79.3%, 高于表 2 中部分方法的最终实验结果. 而在信息丰富的区间, 如度数高于 120 的节点区间, 模型的 Hits@1 值超过了 90%, 甚至在 180-200 的区间里达到了接近完美的水平. 这与图神经网络消息传递机制的

规律相符: 邻域信息越丰富, 当前节点所能获得的信息越充足, 而领域相似信息越多, 对应实体的对齐置信度越高。

为了进一步比较本文在长尾节点上的表现, 本文选取多模态基线模型 EVA 与仅使用结构信息的基线模型 GCN 进行比较, 实验结果如图 9 所示。

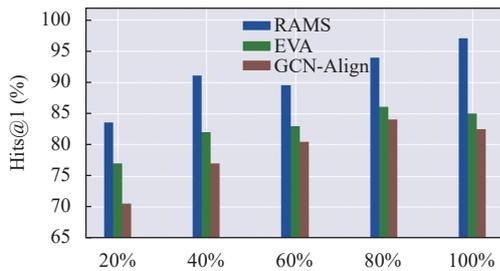


图 9 以 DBP15k (ZH-EN) 为例, 模型分度数区间实验对比结果

将节点按度数从小到大排列, 划分为 5 个区间, 统计各区间内模型的表现。仅使用结构信息的 GCN 模型在各度数区间都表现不佳, 随度数增加模型效果有所改善。相比 GCN, 加入了图像等多模态信息的 EVA 模型在效果上有较大提升, 这验证了多模态信息能够起到特征补充的作用。最后, 由于本文模型额外在多模态特征编码阶段引入了关系敏感型多子图神经网络, 使模态信息与图结构信息充分融合, 因此相比基线模型在各度数区间都获得了进一步提升, 验证了本文模型的有效性。

4 结论与展望

针对模态特征缺失问题, 本文提出了一种多模态多视图的关系敏感型正交图神经网络。该网络通过多视图特征编码和模态融合损失来补足模态间的信息差异, 并借助期望控制的采样方法和矩阵约束方法增强模型效果。广泛的对比实验表明, 本文提出的模型超越了所有的基线模型, 并验证了本文模型在鲁棒性、运行效率、长尾节点准确率等多个维度保持优势。

在未来的工作中, 我们计划在子图编码模块引入差异化的网络结构, 以进一步改进多子图消息传递机制, 从而平衡模态间信息差异并增强模态表征能力。同时, 考虑到现实中的知识图谱可能更加稀疏, 并且图像等额外信息的分布更加不均匀, 我们还将针对长尾节点和稀疏图进行研究与优化。

参考文献

- Ji SX, Pan SR, Cambria E, *et al.* A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(2): 494–514. [doi: [10.1109/TNNLS.2021.3070843](https://doi.org/10.1109/TNNLS.2021.3070843)]
- Auer S, Bizer C, Kobilarov G, *et al.* DBpedia: A nucleus for a Web of open data. *Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference on the Semantic Web*. Busan: Springer, 2007. 722–735.
- Suchanek FM, Kasneci G, Weikum G. YAGO: A core of semantic knowledge. *Proceedings of the 16th International Conference on World Wide Web*. Banff: ACM, 2007. 697–706.
- Rotmensch M, Halpern Y, Tlimat A, *et al.* Learning a health knowledge graph from electronic medical records. *Scientific Reports*, 2017, 7(1): 5994. [doi: [10.1038/s41598-017-05778-z](https://doi.org/10.1038/s41598-017-05778-z)]
- Tang J, Zhang J, Yao LM, *et al.* ArnetMiner: Extraction and mining of academic social networks. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Las Vegas: ACM, 2008. 990–998.
- Chen MH, Tian YT, Yang MH, *et al.* Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne: AAAI Press, 2017. 1511–1517.
- Hao YC, Zhang YZ, He SZ, *et al.* A joint embedding method for entity alignment of knowledge bases. *Proceedings of the 1st China Conference on Knowledge Graph and Semantic Computing: Semantic, Knowledge, and Linked Big Data*. Beijing: Springer, 2016. 3–14.
- Zhao X, Zeng WX, Tang JY, *et al.* An experimental study of state-of-the-art entity alignment approaches. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(6): 2610–2625.
- 张天成, 田雪, 孙相会, 等. 知识图谱嵌入技术研究综述. *软件学报*, 2023, 34(1): 277–311. [doi: [10.13328/j.cnki.jos.006429](https://doi.org/10.13328/j.cnki.jos.006429)]
- Sun ZQ, Hu W, Zhang QH, *et al.* Bootstrapping entity alignment with knowledge graph embedding. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm: AAAI Press, 2018. 4396–4402.
- Cao YX, Liu ZY, Li CJ, *et al.* Multi-channel graph neural network for entity alignment. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence: Association for Computational Linguistics, 2019. 1452–1461.

- 12 Mao X, Wang WT, Wu YB, *et al.* Are negative samples necessary in entity alignment? An approach with high performance, scalability and robustness. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. ACM, 2021. 1263–1273.
- 13 Liu FY, Chen MH, Roth D, *et al.* Visual pivoting for (unsupervised) entity alignment. Proceedings of the 35th AAAI Conference on Artificial Intelligence, the 33rd Conference on Innovative Applications of Artificial Intelligence, and the 11th Symposium on Educational Advances in Artificial Intelligence. AAAI Press, 2021. 4257–4266.
- 14 Sun ZQ, Hu W, Li CK. Cross-lingual entity alignment via joint attribute-preserving embedding. Proceedings of the 16th International Semantic Web Conference on the Semantic Web. Vienna: Springer, 2017. 628–644.
- 15 Guo LB, Sun ZQ, Hu W. Learning to exploit long-term relational dependencies in knowledge graphs. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 2505–2514.
- 16 Yang K, Liu SQ, Zhao JF, *et al.* COTSAE: Co-training of structure and attribute embeddings for entity alignment. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020. 3025–3032.
- 17 Wang ZC, Lv QS, Lan XH, *et al.* Cross-lingual knowledge graph alignment via graph convolutional networks. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 349–357.
- 18 Zhu QN, Zhou XF, Wu J, *et al.* Neighborhood-aware attentional representation for multilingual knowledge graphs. Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press, 2019. 1943–1949.
- 19 Mao X, Wang WT, Xu HM, *et al.* MRAEA: An efficient and robust entity alignment approach for cross-lingual knowledge graph. Proceedings of the 13th International Conference on Web Search and Data Mining. Houston: ACM, 2020. 420–428.
- 20 Li CJ, Cao YX, Hou L, *et al.* Semi-supervised entity alignment via joint knowledge embedding model and cross-graph model. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 2723–2732.
- 21 Yang HW, Zou YY, Shi P, *et al.* Aligning cross-lingual entities with multi-aspect information. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong: Association for Computational Linguistics, 2019. 4431–4441.
- 22 Liu Y, Li H, Garcia-Duran A, *et al.* MMKG: Multi-modal knowledge graphs. Proceedings of the 16th International Conference on the Semantic Web. Portorož: Springer, 2019. 459–474.
- 23 Chen LY, Li Z, Wang YJ, *et al.* MMEA: Entity alignment for multi-modal knowledge graph. Proceedings of the 13th International Conference on Knowledge Science, Engineering and Management. Hangzhou: Springer, 2020. 134–147.
- 24 Guo H, Tang JY, Zeng WX, *et al.* Multi-modal entity alignment in hyperbolic space. Neurocomputing, 2021, 461: 598–607. [doi: [10.1016/j.neucom.2021.03.132](https://doi.org/10.1016/j.neucom.2021.03.132)]
- 25 Lin ZX, Zhang ZH, Wang M, *et al.* Multi-modal contrastive representation learning for entity alignment. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju: International Committee on Computational Linguistics, 2022. 2572–2584.
- 26 Chen Z, Chen JY, Zhang W, *et al.* MEAformer: Multi-modal entity alignment Transformer for meta modality hybrid. Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023. 3317–3327.
- 27 Chen LY, Li Z, Xu T, *et al.* Multi-modal Siamese network for entity alignment. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2022. 118–126.
- 28 Deng J, Dong W, Socher R, *et al.* ImageNet: A large-scale hierarchical image database. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009. 248–255.
- 29 Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.
- 30 Liu FY, Ye RT, Wang X, *et al.* HAL: Improved text-image matching by mitigating visual semantic hubs. Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, and the 10th AAAI Symposium on Educational Advances in Artificial Intelligence. New York: AAAI Press, 2020. 11563–11571.
- 31 Conneau A, Lample G, Ranzato MA, *et al.* Word translation without parallel data. arXiv:1710.04087, 2017.

(校对责编: 孙君艳)