

基于元学习的跨语言知识图谱实体对齐框架

陈壮壮, 邓怡辰, 余敦辉, 肖奎

引用本文

陈壮壮, 邓怡辰, 余敦辉, 肖奎. [基于元学习的跨语言知识图谱实体对齐框架](#)[J]. 计算机科学, 2026, 53(1): 271-277.

CHEN Zhuangzhuang, DENG Yichen, YU Dunhui, XIAO Kui. [Cross-language Knowledge Graph Entity Alignment Based on Meta-learning](#) [J]. Computer Science, 2026, 53(1): 271-277.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于邻域匹配概率与类型商图的实体对齐解释方法](#)

Explanation Method for Entity Alignment Based on Neighborhood Matching Probability and Type Quotient Graph

计算机科学, 2025, 52(12): 260-270. <https://doi.org/10.11896/jsjcx.241100081>

[基于变体注意力的关系与属性感知实体对齐](#)

Relationship and Attribute Aware Entity Alignment Based on Variant-attention

计算机科学, 2025, 52(11): 230-236. <https://doi.org/10.11896/jsjcx.240800140>

[基于两阶段时空对齐的小样本视频行为识别](#)

Few-shot Video Action Recognition Based on Two-stage Spatio-Temporal Alignment

计算机科学, 2025, 52(8): 251-258. <https://doi.org/10.11896/jsjcx.240900127>

[基于元增量学习的开放集识别方法](#)

Open Set Recognition Based on Meta Class Incremental Learning

计算机科学, 2025, 52(5): 187-198. <https://doi.org/10.11896/jsjcx.240600162>

[基于地理时空关联和社会影响的兴趣点推荐](#)

Point-of-interest Recommendation Based on Geospatial-Temporal Correlations and Social Influence

计算机科学, 2025, 52(5): 128-138. <https://doi.org/10.11896/jsjcx.240200099>

基于元学习的跨语言知识图谱实体对齐框架

陈壮壮¹ 邓怡辰³ 余敦辉^{1,2} 肖奎^{1,2}

1 湖北大学计算机与信息工程学院 武汉 430062

2 大数据智能分析与行业应用湖北省重点实验室 武汉 430062

3 武昌首义学院 武汉 430064

(654414209@qq.com)

摘要 跨语言知识图谱实体对齐是连接不同语言知识图谱的关键步骤,在多语言信息检索、数据融合等任务中有重要作用。然而,现有的实体对齐方法依赖知识图谱中的多种信息,难以很好地处理稀疏知识图谱实体对齐任务,并且对新的语言的适应性较差。针对该问题,提出了基于元学习的跨语言实体对齐框架。该框架总体分为外循环与内循环两个阶段:在外循环阶段,通过基于任务相似度的采样方法选取多个任务,然后对模型进行多任务联合训练,构建教师模型;在内循环阶段,利用外循环阶段训练好的教师模型指导学生模型进行训练和实体对齐任务,提升学生模型实体对齐的性能和泛化性。在 SRPRS 和 WK31-60K 数据集上的实验结果表明,所提框架在实体对齐问题中,Hits@1 指标平均提升 3.5%,Hits@10 指标平均提升 4.0%,MRR 指标平均提升 6.3%。

关键词: 元学习;跨语言知识图谱;实体对齐;外循环;内循环;泛化能力

中图分类号 TP391

Cross-language Knowledge Graph Entity Alignment Based on Meta-learning

CHEN Zhuangzhuang¹, DENG Yichen³, YU Dunhui^{1,2} and XIAO Kui^{1,2}

1 School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

2 Hubei Key Laboratory of Big Data Intelligent Analysis and Application, Wuhan 430062, China

3 Wuchang Shouyi College, Wuhan 430064, China

Abstract Cross-language knowledge graph entity alignment is a key step in connecting knowledge graphs of different languages, and it plays an important role in tasks such as multilingual information retrieval and data fusion. However, the existing entity alignment methods rely on a variety of information in the knowledge graph, which cannot handle the entity alignment task of the sparse knowledge graph well, and has poor adaptability to new languages. To solve this problem, a cross-language entity alignment framework based on meta-learning is proposed. The framework is generally divided into two stages, the outer loop and the inner loop. In the outer loop stage, multiple tasks are selected through the sampling method based on task similarity, and then the model is jointly trained with multiple tasks to construct the teacher model. In the inner loop stage, the teacher model trained in the outer loop stage is used to guide the student model to carry out the training and entity alignment tasks, in order to improve the entity alignment performance and generalization of the student model. Experimental results on the SRPRS and WK31-60K dataset show that the proposed framework improves the Hits@1 index by 3.5%, Hits@10 index by 4.0%, and MRR index by 6.3% on average in the entity alignment problem.

Keywords Meta-learning, Cross-language knowledge graph, Entity alignment, Outer loop, Inner loop, Generalization ability

1 引言

近年来,互联网的快速发展促使各领域建立了越来越多包含互补信息的跨语言知识图谱。网络上,语义数据的数量不断增加,跨语言信息交流愈加频繁。知识图谱作为结构化的知识库,在跨语言信息检索^[1]、推荐系统^[2]和多语言问答

系统^[3]等领域发挥着至关重要的作用。如何对齐不同语言知识图谱中的实体,使得跨语言知识图谱之间可以高效协调,是亟待解决的问题。

传统的实体对齐方法在处理语言差异和变体时存在明显的局限性,它们忽略了实体背后的深层语义信息,导致泛化性差。此外,基于表示学习和基于深度学习的实体对齐方法虽

到稿日期:2024-11-12 返修日期:2025-02-11

基金项目:国家自然科学基金(62377009)

This work was supported by the National Natural Science Foundation of China(62377009).

通信作者:邓怡辰(104238327@qq.com)

然取得了一定的进展,但它们通常需要依靠大量知识图谱内部信息,并且在面对新的跨语言知识图谱实体对齐任务时泛化性有待提高。

受 Chen 等^[4-5]的启发,本文提出了一种基于元学习的跨语言知识图谱实体对齐方法。元学习作为一种学习新任务的学习范式,已在多个领域显示出强大的能力。通过元学习框架,可以学习到多个任务的实体表示和拓扑结构,使得模型具有较好的泛化性,即使在标注数据较少的情况下也能取得良好性能。

本文的主要贡献包括:

1)提出了一种基于元学习的跨语言实体对齐方法,使得模型能够将从已训练任务中吸取的经验应用于新的对齐任务,从而在标注数据有限的情况下也能表现出良好的泛化能力;

2)采用基于相似度的任务采样策略,优化了训练过程,显著提高了模型在跨语言知识图谱实体对齐任务中的鲁棒性和对齐性能;

3)实验表明,所提框架与已有的基准方法相比,Hits@1 指标平均提升 3.5%,Hits@10 指标平均提升 3.0%,MRR 指标平均提升 6%,验证了所提框架的有效性。

2 相关工作

在跨语言知识图谱实体对齐领域,研究者面临着将不同语言知识图谱中的实体进行精确匹配的挑战。这一任务对促进跨语言信息检索、增强机器翻译的准确性以及提升智能问答系统的性能至关重要。实体对齐方法主要分为以下 3 类。

第一类是基于相似度的字符串匹配方法,这类方法通过计算实体名称之间的字符串相似度来识别不同语言知识图谱中的对应实体。常用的算法包括编辑距离和 Jaro-Winkler 等,它们通过比较实体名称的文本序列,寻找相似的模式。然而,这些方法通常忽略了实体背后的深层语义信息,仅依赖于文本的表面形式,导致在处理不同语言或方言的变体时效果不佳,且难以适应新的跨语言实体对齐任务。

第二类方法是知识表示学习的嵌入^[6]方法,它们通过学习实体和关系的向量表示来实现实体对齐,捕捉知识图谱的结构特征。例如:Chen^[7]等提出的 MTransE 通过对不同关系的多重嵌入进行建模,以处理多种关系类型;Sun 等^[8]提出 JAPE,通过联合学习对齐和预测嵌入,旨在捕捉知识图谱中实体和关系的复杂模式;Trouillon 等^[9]提出的 ComplEx 通过将实体和关系映射到复数向量空间来解决对称性和反对称性关系的问题,它可以更有效地处理具有复杂语义的关系。但这些方法通常需要依靠大量知识图谱内部信息,在处理稀疏知识图谱时会遇到性能瓶颈,对新任务的适应性有限。

第三类是基于图神经网络的深度学习方法,这类方法利用图神经网络来聚合实体的邻居信息,生成能够反映实体间复杂关系和结构特征的嵌入表示。例如,GCN-Align^[10],RREA^[11],MMEA^[12],KE-GCN^[13]等模型通过图卷积网络对邻居节点进行聚合,学习实体的图结构信息,以实现更精准的实体对齐。这些方法在处理知识图谱时可能面临全局结构信息不足的问题。此外,它们在处理数据稀疏性问题和适应新

任务方面仍面临挑战。

尽管现有的实体对齐方法在某些方面取得了进展,但泛化能力仍有待提高,并且它们过于依赖知识图谱内部信息。这些缺点限制了它们在稀疏知识图谱实体对齐任务中的应用效果。因此,研究者正在探索新的技术和方法,以提高实体对齐的准确性和效率。

元学习^[14-15]作为一种新兴的学习范式,为跨语言实体对齐提供了新的视角。元学习关注于用以往的学习经验来加速新任务的学习过程,在跨语言知识图谱实体对齐的背景下,这意味着模型能够从先前的对齐任务中学习在先验知识,快速适应新的、数据稀缺的对齐任务。为此,本文提出一种基于元学习框架的多任务学习方法来进行实体对齐,确保在数据稀缺的情况下模型仍能适应新任务,提高实体对齐性能。这对于跨语言知识图谱的构建和应用具有重要意义。

3 问题定义与总框架

3.1 问题定义

跨语言知识图谱实体对齐问题可以形式化定义为:给定两个跨语言知识图谱 $G_1 = \{e_1, r_1, t_1\}$ 和 $G_2 = \{e_2, r_2, t_2\}$,其中 e 代表实体, r 代表关系, t 代表知识图谱中的三元组集合。跨语言知识图谱实体对齐的目的是找到一个映射,将 G_1 和 G_2 中语义相同的实体 e_1 和 e_2 链接起来,使得多语言知识图谱之间的实体可以在语义上对齐,从而让知识图谱中的数据可以在多语言环境下进行交互。

3.2 总框架

该框架主要分为内循环和外循环两个阶段,在外循环阶段采用基于任务相似度的采样方法选择出具有高相似度的任务集,选取出来的具有高相似度的任务在外循环阶段有助于简化教师模型的训练和参数优化过程。这些任务内部的两个知识图谱有着高相似性,使得模型能够更快地学习到通用的知识和特征表示。同时,每个任务之间又存在一定的差异性,能够避免模型学习到过于单一的模式,从而更好地对模型进行训练,构建教师模型。通过多任务联合训练,教师模型积累通用知识,学习到多个任务的共同特征。这些共同特征和通用知识不受限于单个稀疏知识图谱的数据特点,能够指导学生模型更好地理解 and 处理不同类型的实体对齐任务,为学生模型提供更全面的知识基础,使学生模型在进行稀疏知识图谱实体对齐任务时能有更多的数据和先验知识。

内循环阶段,教师模型将指导学生模型进行参数更新,并在特定任务上进行实体对齐。教师模型通过其在外循环多个任务中学习到的通用知识和特征表示,为学生模型提供关于实体对齐的指导。在稀疏知识图谱中,学生模型可能因数据不足而难以准确学习到实体对齐的模式,教师模型的指导可以帮助学生模型利用已有的通用知识,更好地处理不同稀疏知识图谱的实体对齐任务,从而提高模型在稀疏知识图谱上的实体对齐性能,并提高学生模型面对稀疏知识图谱实体对齐任务时的泛化能力和适应性。

如果学生模型的实体对齐性能未达到预期,学生模型会将性能指标、更新后的参数和损失函数值反馈给教师模型。教师模型将通过外循环阶段的多任务训练重新更新参数、学习特

征,以进一步指导学生模型进行实体对齐任务。这种反馈机制使得学生模型能够不断调整和优化自身的参数和学习策略,以适应稀疏知识图谱的特点,提高实体对齐的准确性和效率。

4 方法

4.1 基于任务相似度的采样

给定知识图谱集 $G = \{G_1, G_2, \dots, G_n\}$, 在任务采样阶段,

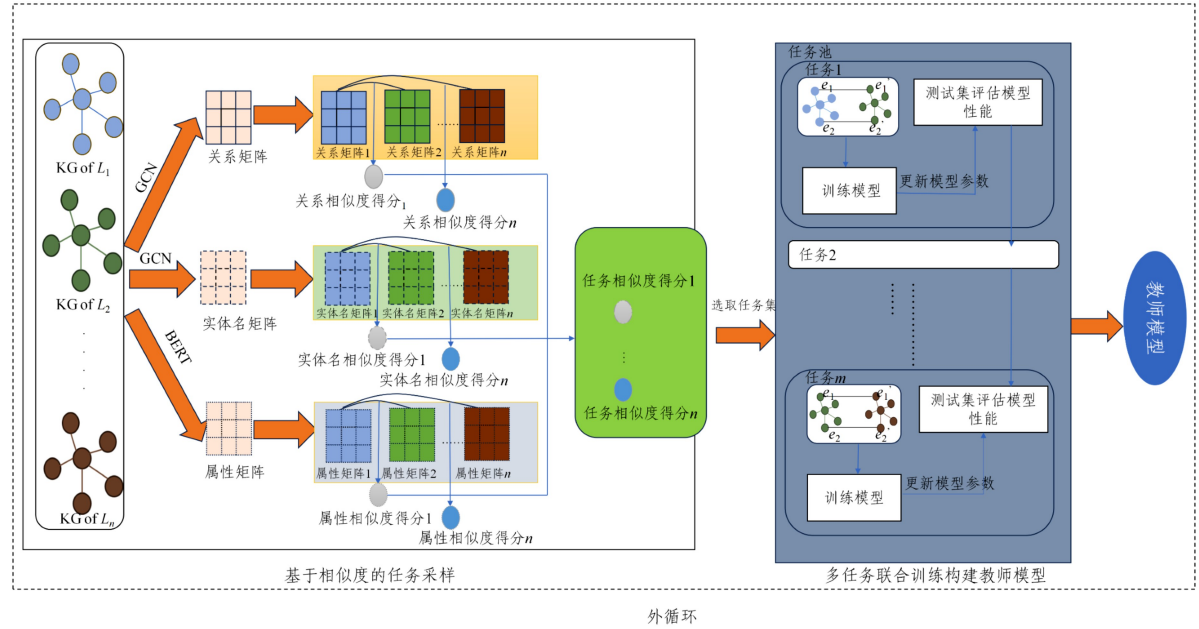


图1 外循环阶段

Fig. 1 Outer loop stage

4.1.1 实体名和关系类型的相似度计算

用 GCN 获取知识图谱的关系矩阵和实体名矩阵, 给定知识图谱 $G_1 = \{e_1, r_1, t_1\}$ 和 $G_2 = \{e_2, r_2, t_2\}$, 关系矩阵分别为 $Relation Matrix_1$ 和 $Relation Matrix_2$, 实体名矩阵分别为 $Entity Name Matrix_1$ 和 $Entity Name Matrix_2$, 通过 Frobenius 范数形式化定义关系相似度和实体名相似度:

$$S_r = \sqrt{\sum_{i,j} (R_{ij}^a - R_{ij}^b)^2} \quad (1)$$

$$S_e = \sqrt{\sum_{i,j} (E_{ij}^a - E_{ij}^b)^2} \quad (2)$$

其中, R_{ij}^a 和 R_{ij}^b 分别表示 $Relation Matrix_1$ 和 $Relation Matrix_2$ 中相应位置的元素, E_{ij}^a 和 E_{ij}^b 分别表示 $Entity Name Matrix_1$ 和 $Entity Name Matrix_2$ 中相应位置的元素。

4.1.2 属性相似度计算

通过 Bert 模型获取知识图谱的语义信息和属性, 并生成相应的属性矩阵。同样地, 给定知识图谱 $G_1 = \{e_1, r_1, t_1\}$ 和 $G_2 = \{e_2, r_2, t_2\}$, 可以形式化定义属性相似度:

$$S_a = \sqrt{\sum_{i,j} (A_{ij}^a - A_{ij}^b)^2} \quad (3)$$

其中, A_{ij}^a 和 A_{ij}^b 分别表示 $Attribute Matrix_1$ 和 $Attribute Matrix_2$ 中相应位置的元素。

4.1.3 相似度加权采样

在计算出实体名相似度、关系相似度和属性相似度得分后, 对这 3 个相似度得分进行加权求和, 得到任务之间的综合相似度, 并根据综合相似度进行采样。

针对每个知识图谱, 通过计算与其他 $n-1$ 个知识图谱之间的属性相似度、实体名相似度和关系相似度, 并对 3 个相似度进行加权求和, 找出与之最相似的另一个知识图谱, 从而形成任务集 $T = \{T_1, T_2, \dots, T_m\} (m \leq n)$ 。每个任务分为训练集和测试集, 在外循环阶段, 每个任务内部的两个知识图谱具有高相似性, 但每个任务之间又存在一定的差异性, 从而可以更好地对模型进行训练。

任务之间的综合相似度可以使用以下加权公式计算:

$$S = w_r S_r + w_e S_e + w_a S_a \quad (4)$$

其中, w_r , w_e 和 w_a 是对应相似度的权重。

Frobenius 范数具有连续性和可微性, 常用于梯度下降算法中, 通过梯度下降, 根据模型的性能自动更新 3 种相似度的权重。这种方法更灵活, 能够根据数据动态调整权重, 避免出现拟合问题, 且其对矩阵的小扰动有较好的稳定性, 这使得它在数值计算中相对可靠, 得到的任务相似度综合得分也有较高的真实性。

选取出来的具有高相似度的 m 个任务在外循环阶段有助于简化模型的训练和参数优化过程。这种策略通过先处理与当前模型知识更匹配的任务, 再逐步过渡到差异较大的任务, 可以有效地引导模型学习, 从而为整个训练过程提供一种更加高效的学习路径。

4.2 基于元学习框架的少样本实体对齐方法

在跨语言知识图谱实体对齐任务中, 元学习过程分为内循环和外循环两个阶段。

外循环训练阶段, 通过多个跨语言实体对齐任务, 提高模型的泛化能力和适应性。首先在 m 个任务上训练模型, 然后计算元损失函数 L_{meta} , 其中包括每个任务的损失函数和一个正则化项, 以鼓励模型参数的一致性更新。元损失函数的计算式为:

$$L_{task} = \sum_{(e_i, e_j)} \max(0, margin - f(e_i)^T f(e_j)) \quad (5)$$

$$L_{\text{meta}} = \sum_{i \in m} \tau_i L_{\text{task}}(i, \theta_0) + \lambda R(\theta_0) \tag{6}$$

$$\theta_0 = \theta_0 - \beta \cdot \nabla L_{\text{task}}(\theta_0) \tag{7}$$

$$\theta_{\text{new}} = \theta_0 \tag{8}$$

其中, $f(e_i)$ 和 $f(e_j)$ 分别是实体 e_i 和 e_j 的特征向量; margin 是一个预设的间隔, 用于确保正确对齐的实体对在嵌入空间中彼此接近; θ_0 是初始模型的参数; θ_{new} 是外循环最终输出的教师模型参数; β 是外循环的学习率; $\nabla L_{\text{task}}(\theta_0)$ 是任务损失函数关于模型参数的梯度; $R(\theta_0)$ 是正则化项; τ_i 是不同实体对齐任务的权重, 用于避免模型过于关注数据量较大的任务, 进而出现过拟合问题。基于相似度的任务选择方法在这一过程中具有重要作用, 因为它可以更智能地选择任务, 并为模型提供更加有针对性的训练数据。

学生模型采用了与教师模型相似的架构, 以便快速吸收和应用教师模型的知识。具体来说, 学生模型包括以下几个主要部分。

1) 输入层: 接收来自知识图谱的实体特征表示, 特征表示为 x 。

2) 隐藏层: 通过多层神经网络结构处理输入特征, 每层都使用非线性激活函数以增强模型的表达能力。假设有 l 层隐藏层, 第 l 层的输出表示为:

$$h_l = \sigma(W_l h_{l-1} + b_l) \tag{9}$$

其中, W_l 是第 l 层的权重矩阵, b_l 是第 l 层的偏置向量, h_{l-1} 是第 $l-1$ 层的输出, $h_0 = x$ 。

3) 输出层: 预测实体对的对齐概率, 表示为:

$$y = W_L h_{L-1} + b_L \tag{10}$$

其中, W_L 是输出层的权重矩阵, b_L 是输出层的偏置向量, h_{L-1} 是最后一层隐藏层的输出。

内循环阶段如图 2 所示, 教师模型指导学生模型在特定任务的训练集上进行训练并指导学生模型调整参数。内循环训练的模型参数更新公式为:

$$\theta_{\tau} = \theta_{\text{new}} - \alpha \cdot \nabla L_{\text{task}}(\theta_{\text{new}}) \tag{11}$$

其中, θ_{new} 是教师模型的参数, α 是内循环中学生模型的学习率, θ_{τ} 是训练后学生模型的参数, $\nabla L_{\text{task}}(\theta_{\text{new}})$ 是任务损失函数关于教师模型参数的梯度。使用学生模型在测试集上评估实体对齐性能, 以确定学生模型对新实体对的对齐能力。这种更新策略使得学生模型能够在保持对当前任务适应性的同时, 学习教师模型在多任务学习中获得的知识。

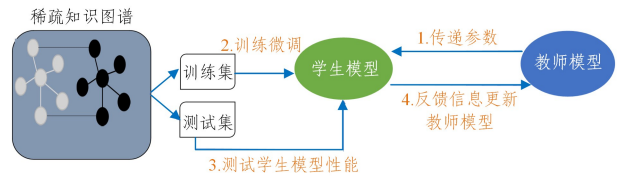


图 2 内循环阶段

Fig. 2 Inner loop stage

如果学生模型的实体对齐性能未达到预期, 则表明模型在适应新任务时遇到了困难。在这种情况下, 学生模型会将性能指标、更新后的参数和损失函数值反馈给教师模型。教师模型通过外循环阶段的多任务训练重新更新参数、学习

特征, 以进一步指导学生模型进行实体对齐任务。这个过程会循环进行, 直到学生模型在新任务上的对齐性能达到让人满意的程度。通过这种内外循环的交替训练和反馈机制, 元学习框架能够有效提升模型在跨语言知识图谱实体对齐任务中的性能和泛化性。

5 实验分析

5.1 数据集和实验设置

为了充分验证基于元学习的实体对齐框架在少样本跨语言实体对齐任务上的有效性, 实验采用了 3 个不同的知识图谱实体对齐任务数据集, 即 DBP15K, SRPRS, WK31-60K。数据集的具体信息如表 1 所列。

表 1 数据集统计信息
Table 1 Dataset statistics

Datasets	Relations	Reltriples	Atttriples
DBP15K _{ZH-EN}	2 830	153 929	379 684
	2 317	237 674	567 755
DBP15K _{JA-EN}	2 043	164 373	354 619
	2 096	233 319	497 230
DBP15K _{FR-EN}	1 379	192 191	528 665
	2 209	278 590	576 543
SRPRS _{EN-FR}	221	36 508	60 800
	177	33 532	53 045
SRPRS _{EN-DE}	222	38 363	55 580
	120	37 377	73 753
WK31-60K _{EN-FR}	445	47 508	69 223
	277	53 532	58 081
WK31-60K _{EN-DE}	477	53 363	60 232
	310	64 377	89 897

1) DBP15K: 包含 3 个语言对, 分别是 ZH-EN(中文-英文)、JA-EN(日文-英文)、FR-EN(法文-英文)。该数据集包含实体、关系三元组及其对应的属性三元组, 用于在外循环阶段对模型进行训练, 构建教师模型。

在 DBP15K 数据集中, 每个语言对任务的数据被随机划分为训练集、验证集和测试集。其中 70% 的数据用于训练集, 15% 用于验证集, 15% 用于测试集。这种划分旨在确保模型在训练过程中有足够的样本学习, 同时也有足够的数据用于模型性能的评估和超参数的调整。

初始学习率设置为 0.003, 以获得平衡的收敛速度。在训练过程中, 采用学习率衰减策略, 经过 100 次迭代后, 学习率会衰减为原来的 0.9 倍, 以保证模型在训练后期能够更细致地逼近最优解。

选择 Adam 优化器进行模型参数的更新。Adam 优化器结合了 AdaGrad 和 RMSProp 的优点, 能够自动调节学习率, 并适应参数的更新频率, 使得模型在训练过程中更加稳定。

为了防止模型过拟合, 引入 L2 正则化。正则化系数设置为 0.01, 这个值在多次实验中被证明能够有效地控制模型的复杂度和保持模型性能。

2) SRPRS: 包含两个语言对, 分别是 EN-FR(英文-法文)和 EN-DE(英文-德文)。

3) WK31-60K: 包含两个语言对, 分别是 EN-FR(英文-法文)和 EN-DE(英文-德文)。

SRPRS 和 WK31-60K 数据集规模较小,旨在评估数据稀缺情况下学生模型的表现。

设置实体和关系的嵌入维度为 300 维,以确保嵌入向量能够捕捉足够的语义信息;元学习框架内循环的学习率设置为 0.005,以加速模型在特定任务上的适应速度;内循环更新次数设置为 50;随机初始化任务相似度权重,并在实验过程中进行更新。

5.2 基准及评价指标

实验将不同种类知识图谱实体对齐方法中具有代表性的 7 种算法作为基准,并在进行实体对齐时将其与元学习框架嵌套,以充分验证所提框架的有效性。具体描述如下:

1)MTransE^[7]:一种基于嵌入的跨语言知识图谱实体对齐模型,通过学习不同语言的实体和关系嵌入,采用距离校准、翻译向量和线性变换 3 种技术对齐不同语言的嵌入空间,优化实体对齐效果。

2)HMAN^[16]:基于嵌入的方法将多语言 KG 中的实体编码到相同的向量空间中,其中等价的实体彼此接近。其应用图卷积网络(GCNs)来组合实体的多方面信息,学习实体嵌入,并使用预训练的 BERT 模型来弥补跨语言的差距,以提高性能。

3)RDGCN^[17]:基于关系感知双图卷积网络的实体对齐方法,通过知识图和它的对偶关系之间的密切交互来整合关系信息,并进一步捕获邻近结构来学习更好的实体表示。

4)JAPE^[8]:通过结合两个知识图谱的结构嵌入和属性嵌入来实现实体对齐。结构嵌入遵循 TransE 模型,学习两个知识库叠加图的实体的向量表示,而属性嵌入遵循 skip-gram 模型,旨在捕获属性之间的相关性,从而进行实体对齐。

5)HGCN^[18]:HGCN 的框架核心是利用图卷积网络学习两种语言知识库中的实体和关系的嵌入表示,以此来实现跨语言的实体对齐。具体来说,HGCN 不依赖预先对齐的关系种子来学习关系表示,而是首先使用 GCN 学习的实体嵌入

来近似它们,然后将关系近似合并到实体中,以迭代学习两者的更好表示。

6)GALA^[19]:通过结合全局对齐信息和局部信息聚合来提高实体对齐的准确性。在半监督学习的场景下,GALA 利用少量已标注的实体对以及大量未标注的数据学习实体之间的对齐关系。

7)RALG^[20]:提出独立于实体学习关系表示,并利用异质线图明确捕捉关系的相关性,以更好地实现跨语言实体对齐。

实验使用平均倒数排名 MRR 和 Hit@k 来评估模型性能。Hit@k 是指结果排名前 k 个中正确实体所占的比例,Hit@k 越大,方法的效果越好。MRR 代表正确对齐实体排名的倒数的平均值, MRR 越大,方法的效果越好。

$$Hits@k=\frac{|\{e_i\in N|rank_i\leqslant k\}|}{|N|}$$

(12)

$$MRR=\frac{1}{|N|}\sum_{i=1}^{|N|}\frac{1}{rank_i}$$

(13)

其中,Hit@k∈[0,1],rank_{*i*} 表示对齐到正确的实体的排名,N 为对齐实体对,|·|为取个数运算,这里的 k 分别取 1 和 10。

5.3 实验结果与分析

在 SRPRS 数据集和 WK31-60K 数据集上的实验结果如表 2 和表 3 所列,其中“O”代表使用基准模型进行实验,“Meta”代表将基准模型嵌入所提出的元学习框架中进行实验。表中结果表明,所提出的基于元学习的跨语言知识图谱实体对齐方法在多个评估指标上优于现有的基线方法。在 SRPRS 数据集上,与基准模型相比,应用元学习框架后 Hits@1 指标平均提升了 3 个百分点,Hits@10 指标平均提升了 3.5 个百分点,MRR 指标平均提升了 5.5 个百分点。在 WK31-60K 数据集上,应用元学习框架后 Hits@1 指标平均提升了 4 个百分点,Hits@10 指标平均提升了 4.5 个百分点,MRR 指标平均提升了 7 个百分点。

表 2 SRPRS 数据集下嵌入元学习框架与基准模型对比

Table 2 Comparison between the embedded meta-learning framework and the benchmark model in the SRPRS dataset

算法	设置	EN-FR			EN-DE		
		Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MTransE	O	21.3	44.7	0.29	10.7	24.8	0.16
	Meta	24.3	48.7	0.35	13.9	27.5	0.22
HMAN	O	40.0	70.5	0.50	52.8	74.8	0.62
	Meta	44.3	73.7	0.56	55.8	77.9	0.67
RDGCN	O	67.2	75.7	0.71	75.9	86.6	0.82
	Meta	70.6	79.2	0.78	78.8	89.2	0.88
JAPE	O	24.1	54.4	0.34	26.8	54.7	0.36
	Meta	27.3	58.1	0.39	29.9	58.8	0.43
HGCN	O	67.0	77.0	0.71	76.3	86.3	0.80
	Meta	69.8	80.3	0.76	79.8	89.5	0.86
GALA	O	72.1	81.2	0.80	68.1	78.3	0.77
	Meta	74.9	84.1	0.84	70.3	81.6	0.82
RALG	O	80.1	88.5	0.83	75.4	78.2	0.81
	Meta	83.2	91.1	0.88	78.3	81.2	0.87

基于任务相似度的采样策略发挥了重要的作用。该策略通过优先选择与当前任务相似度较高的样本,提升了教师模型的构建效率,减少了训练过程中过度关注不相关任务的

现象,进而避免了模型在特定任务上出现过拟合的问题。这表明,合理选择相似任务进行训练,可以优化模型在多任务情境中的性能,从而加速教师模型的收敛过程。

表 3 WK31-60K 数据集下嵌入元学习框架与基准模型对比

Table 3 Comparison between the embedded meta-learning framework and the benchmark model in the WK31-60K dataset

算法	设置	EN-FR			EN-DE		
		Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
MTransE	O	23.3	45.7	0.32	11.7	23.8	0.26
	Meta	27.1	50.2	0.39	15.9	27.9	0.33
HMAN	O	41.2	72.5	0.53	53.9	75.7	0.65
	Meta	45.3	76.7	0.60	57.8	79.9	0.71
RDGCN	O	69.2	77.5	0.73	77.9	85.6	0.84
	Meta	73.4	81.2	0.79	81.8	90.2	0.91
JAPE	O	27.1	56.7	0.35	26.8	56.9	0.38
	Meta	31.3	60.9	0.42	29.9	61.5	0.45
HGCN	O	67.3	79.0	0.73	78.3	88.3	0.82
	Meta	71.2	83.3	0.78	81.8	91.5	0.80
GALA	O	75.1	82.2	0.81	71.1	78.3	0.78
	Meta	78.9	86.7	0.87	74.8	81.6	0.85
RALG	O	82.1	89.5	0.84	77.4	79.2	0.83
	Meta	86.2	93.2	0.90	81.3	83.3	0.89

教师模型在多任务中积累通用知识,学得多个任务的共同特征,从而指导学生模型进行参数更新,使得学生模型在数据稀缺的情况下依然能够迅速适应,表现出较好的对齐效果。

5.4 元学习框架有效性验证

为了进一步验证现有知识图谱实体对齐模型与所提出框架结合后,其实体对齐性能是否有所提升,本文在 SRPRS 数据集上,将元学习框架与具有代表性的 7 个知识图谱实体对齐模型进行结合。7 个对比模型的具体信息如下:

1) 基于知识表示嵌入的实体对齐方法: ComplEx^[9], EPEA^[21], RoadEA^[22]。该类方法主要是将知识图谱中的实体和关系表示为低维向量,通过向量之间的关系实现实体对齐。

2) 基于图神经网络的实体对齐方法: RREA^[11], MMEA^[12], KE-GCN^[13], NMN^[23]。该类方法用神经网络对知识图谱的结构信息进行学习和处理,以实现实体对齐。

实验结果如图 3 所示,可以发现,嵌套元学习框架后获得的 MRR 得分均比原始方法更高,说明所提出元学习框架通过在外循环阶段使用基于任务相似度的采样方法,并进行多任务联合训练构建教师模型,然后在内循环阶段通过教师模式指导学生模型进行训练,有效提升了学生模型的实体对齐性能和泛化性。其在现有方法(基于知识表示的实体对齐方法、基于图神经网络的实体对齐方法)下均具有一定的有效性。

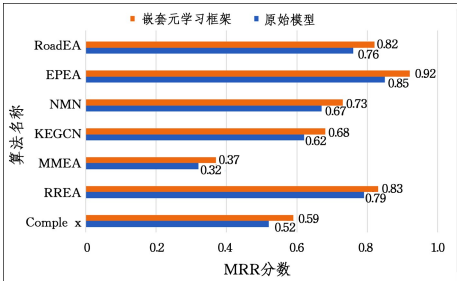


图 3 SRPRS 数据集下嵌入元学习框架与各类模型 MRR 分数对比

Fig. 3 Comparison of MRR scores between the embedded meta-learning framework and various algorithms on the SRPRS dataset

结束语 本文提出了一种基于元学习的跨语言知识图谱实体对齐框架,结合基于任务相似度的采样方法,旨在解决当

前跨语言实体对齐任务中难以处理的数据稀缺问题以及对新任务适应性差的问题。通过引入元学习框架,模型能够在多个跨语言任务上学习通用的知识表征并进行训练,提升了模型的泛化性,即使在少量标注数据的情况下也能保持良好的性能表现。本文为跨语言知识图谱的实体对齐任务提供了新的解决思路,也证明了基于任务相似度采样的元学习框架在稀疏知识图谱实体对齐任务中的广泛应用潜力。然而,本文方法在多任务联合训练阶段没有考虑噪声对构建教师模型的影响,因此仍存在一定的局限性,未来的研究可以进一步提升该框架在多任务训练阶段处理噪声的能力。

参 考 文 献

[1] ZHANG F, YANG L Y, LI J W, et al. A Review of Entity Alignment Research [J]. Journal of Computer Science, 2022, 45(6): 1195-1225.

[2] CAO Y X, WANG X, HE X G, et al. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences[C]// Proceedings of the World Wide Web Conference, 2019: 151-161.

[3] YU L, MADJID N A, DIFALLAH D. CrunchQA: A Synthetic Dataset for Question Answering over Crunchbase Knowledge Graph[C]// 2022 IEEE International Conference on Big Data, IEEE, 2022: 4635-4641.

[4] CHEN X, ZHANG Z, WU Q. A Survey of Meta-Learning Techniques for Knowledge Graphs[J]. ACM Computing Surveys, 2023, 55(4): 1-35.

[5] CHEN X, WU Q, LIU H. Meta-Learning for Knowledge Graph Completion: A Review[J]. Knowledge-Based Systems, 2023, 22(3): 182-193.

[6] WANG T, YANG Z, XU R. A Comprehensive Survey on Knowledge Graph Embedding Techniques[J]. ACM Computing Surveys, 2023, 55(1): 1-34.

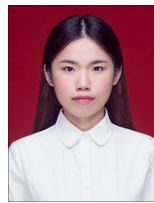
[7] CHEN M, TIAN Y, YANG M, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2016: 213-225.

[8] SUN Z, HU W, LI C. Cross-lingual entity alignment via joint attribute-preserving embedding [C]// The Semantic Web-ISWC

- 2017;16th International Semantic Web Conference. Springer, 2017;1999-2008.
- [9] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]//International Conference on Machine Learning. PMLR, 2016;2071-2080.
- [10] ANG Z, LIU F, ZHANG Z. Cross-lingual knowledge graph alignment via graph convolutional networks[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018;349-357.
- [11] MAO X, LI X, LIU Y, et al. Relational reasoning for knowledge graph entity alignment[C]//Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data. 2021;1121-1133.
- [12] ZHANG J, ZHANG W, CHEN X, et al. Multi-modal entity alignment in knowledge graphs[C]//Proceedings of the 2020 Web Conference. 2020;2164-2174.
- [13] WU Y, WANG X, JIANG P, et al. Knowledge-enhanced graph convolutional networks for entity alignment[C]//Proceedings of the 29th ACM International Conference on Information and Knowledge Management. 2020;1745-1754.
- [14] FINN C, ABBEEL P, LEVINE S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks[C]//Proceedings of the 34th International Conference on Machine Learning (ICML). 2017;70-77.
- [15] SNELL J, SWERSKY K, ZEMEL R S. Prototypical Networks for Few-Shot Learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017;30-39.
- [16] YANG H W, ZOU Y, SHI P, et al. Aligning cross-lingual entities with multi-aspect information [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing(EMNLP-IJCNLP). 2019;2231-2244.
- [17] SANG H, CHEN W, WANG J, et al. RDGCN: Reasonably dense graph convolution network for pedestrian trajectory prediction [J]. Measurement, 2023, 213; 112675.
- [18] ZHU Z, FAN X, CHU X, et al. HGCN: A heterogeneous graph convolutional network-based deep learning model toward collective classification [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020;1161-1171.
- [19] ZHANG X, ZHANG R, CHEN J, et al. Semi-supervised entity alignment with global alignment and local information aggregation[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(10); 10464-10477.
- [20] NG Y, WU J, YU K, et al. Independent Relation Representation With Line Graph for Cross-Lingual Entity Alignment[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(11); 11503-11514.
- [21] WANG Z, YANG J, YE X. Knowledge graph alignment with entity pair embedding[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). Association for Computational Linguistics, 2020;1672-1680.
- [22] SUN Z, HU W, WANG C, et al. Revisiting embedding-based entity alignment: a robust and adaptive method[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(8); 8461-8475.
- [23] WU Y, LIU X, FENG Y, et al. Neighborhood matching network for entity alignment [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020; 6477-6487.



CHEN Zhuangzhuang, born in 1999, postgraduate. His main research interest is knowledge graph.



DENG Yichen, born in 1993, master, lecturer. Her main research interest is embedded development.

(责任编辑:何杨)