

面向多知识图谱融合的实体对齐优化方法

王硕, 李克, 李泽霖

(北京联合大学智慧城市学院, 北京 100101)

摘要: 实体对齐是对不同来源的知识图谱进行知识融合时的一个关键步骤, 现有方法通常只考虑两两图谱间的实体对齐, 而很多实际应用场景中需要对多个图谱进行融合。在解决多图谱实体对齐问题时, 现有的方法往往只能将其转化为多个图谱对的对齐任务, 忽略了多个图谱的等价实体间内在联系和约束, 从而影响了最终的对齐性能。针对上述问题, 在系统分析现有实体对齐优化方法的基础上, 利用等价实体在多个图谱间的传递性约束特征, 提出一种面向多知识图谱融合的实体对齐优化方法(MGEAO), 并通过与现有的两图谱间实体对齐方法相结合, 提出一种通用的多图谱实体对齐优化框架。首先根据每个知识图谱中各实体的嵌入表示计算得到各图谱对的实体预对齐矩阵, 然后经过多图谱间对齐优化方法修正预对齐矩阵得到最终结果。其中对齐优化方法融合了双向归一化(BN)、延迟接受算法(DAA)、关系实体感知调整算法(REA)和传递性约束优化算法(TCO)。在 DBP15K、FB15K 和 YAGO15K 等多个数据集上的实验表明, 和不采用多图谱间对齐优化的基线模型相比, 该优化方法性能显著提升, 其中 Hits@1 和 Hits@10 指标分别最大可提升 18.8 和 18.05 百分点。

关键词: 实体对齐; 知识融合; 知识图谱; 传递性约束; 结构信息

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0070058

Entity Alignment Optimization Method for Multiple Knowledge Graphs Fusion

WANG Shuo, LI Ke, LI Zelin

(Smart City College, Beijing Union University, Beijing 100101, China)

【Abstract】 Entity Alignment (EA) is a key step in the fusion of Knowledge Graphs (KGs). Existing EA methods only consider EA between two KGs, whereas many scenarios require the EA across multiple KGs. Existing methods transform the EA tasks of multiple KGs to several pairwise EA tasks, while ignoring the inherent connections and constraints among equivalent entities across all KGs. To address this problem, based on an analysis of existing EA optimization methods, Entity Alignment Optimization for Multiple knowledge Graphs Fusion (MGEAO) is proposed by leveraging the transitivity constraints of equivalent entities across multiple KGs. A general framework for EA optimization across multiple KGs is proposed by combining it with existing pairwise EA methods. First, the pre-alignment matrix for each KG pair is computed on the basis of entity embeddings. The matrix is then corrected to obtain the final alignment through multiple KGs alignment optimization, which integrates Bidirectional Normalization (BN), Deferred Acceptance Algorithm (DAA), Relation-Entity Aware adjustment (REA) and Transitivity Constrained Optimization (TCO). Experiments on the DBP15K, FB15K, and YAGO15K datasets indicate that the performance is significantly improved in relation to that of baseline EA models, i.e., Hits@1 and Hits@10 are improved by up to 18.8 and 18.05 percentage points, respectively.

【Key words】 Entity Alignment (EA); knowledge fusion; Knowledge Graph (KG); transitivity constraint; structural information

0 引言

近年来,知识图谱(KG)已经广泛应用于许多现实场景中,比如商品推荐^[1]、智能问答^[2]、知识推理^[3]等基于知识图谱的下游任务。大规模通用知识图谱如 DBpedia^[4]、YAGO^[5]、ConceptNet^[6]和 Wikidata^[7]等的出现,为这些下游任务的发展提供了基础。

单个知识图谱所包含的信息往往有限,因此通

常会将多源的知识图谱融合为一个更加完备的知识图谱,即知识融合。实体对齐^[8]是知识融合的一个关键步骤,其目的是发现不同知识图谱中表示相同现实世界对象的等价实体。

目前研究者提出了许多类型的实体对齐方法。研究初期主要采用基于相似度和基于关系推理两种方法进行实体对齐。近年来,基于表示学习的实体对齐方法取得了不错的效果并成为主流方法。这类

基金项目: 国家自然科学基金(61972040)。

作者简介: 王硕(CCF 学生会员),男,硕士研究生,主研方向为知识图谱;李克(通信作者),教授、博士;李泽霖,硕士研究生。

收稿日期: 2024-07-01

修回日期: 2024-10-21

E-mail: like@buu.edu.cn

方法首先利用翻译模型或图神经网络(GNN)^[9]等知识表示学习方法得到实体的低维向量表示,然后将不同图谱中的实体嵌入到同一个向量空间中,通过计算向量间距离发现对齐的实体对。

现有的实体对齐方法一般研究的是两个图谱间的实体对齐任务,而在实际应用中往往需要将多个图谱进行融合。比如将多个百科类图谱如百度百科、维基百科、互动百科等融合为一个百科类图谱;或者对于电商平台,往往需要将多个供应商的商品子图谱融合为一个全品类的商品知识图谱;又或者对于电商比价网站,需要在多个电商平台的商品图谱间建立同一商品实体的等价关系从而提供比价服务。根据应用场景的需要,多图谱间的实体对齐任务又可以分为有中心的多图谱实体对齐和无中心的多图谱实体对齐。前者是将多个边缘图谱中的实体向一个中心图谱的实体进行对齐,后者则是两两图谱间的实体对齐。

现有的实体对齐方法通常并不考虑多图谱间的实体对齐场景,因而在解决此类对齐任务时,通常只能转化为多个图谱对之间的实体对齐问题来求解。由于其忽略了多个图谱的等价实体间的内在联系和约束,往往会影响最终的对齐性能。

针对以上问题,本文提出了一种面向多知识图谱融合的实体对齐优化方法(MGEAO),该方法融合了多种优化方法来提升实体对齐的准确率:使用双向归一化(BN)^[10]处理图谱间的双向匹配问题,使对齐矩阵包含双向的信息;使用延迟接受算法(DAA)^[11]优化预对齐矩阵,从全局上保证每个实体都得到最优的匹配;使用关系实体感知调整算法(REA),利用实体对齐和关系对齐之间的正交互作用修正预对齐矩阵;使用传递性约束优化算法(TCO),利用传递性规则建立多个图谱间的连接,将多图谱间的实体联系起来,通过结构信息和距离信息等约束挖掘等价实体,实现更加准确的实体对齐。本文的主要贡献如下:

1)提出了一种面向多知识图谱融合的传递性约束优化算法,基于传递性规则建立多图谱间的实体关联关系从而提高实体对齐性能。

2)提出了一种关系实体感知调整算法,利用实

体间的关系和实体本身的交互,更有效地利用关系信息来增强对齐性能。

3)提出了一种通用的面向多图谱融合的实体对齐优化框架,该框架可用于基于表示学习的各类实体对齐方法,作为其后处理过程来提升对齐性能。

4)在 DBP15K 多语言图谱数据集和 FB15K、DB15K 和 YAGO15K 多源图谱数据集上分别进行了实体对齐实验,本文优化方法相比基线模型各类评价指标均得到了显著提升。

1 相关工作

近年来,随着知识表示学习技术的发展,基于知识表示学习的实体对齐方法引起了广泛关注。这些方法首先将实体和关系编码到低维向量空间中,然后将这些向量映射到同一个向量空间,通过计算不同实体向量的距离得到对齐的实体对。目前基于表示学习的实体对齐模型一般包括 3 个阶段,即嵌入表示、预对齐和对齐优化调整。

预对齐是指先通过计算向量空间中实体间距离获得初步对齐结果,其形式多表示为对齐矩阵,并按距离最近原则进行等价实体的匹配。

嵌入表示阶段在获得实体和关系的向量表示过程中,主要有两种方法:一种是基于翻译模型的方法,例如 TransE^[12]、MTransE^[13]、PTransE^[14]和 IPTransE^[15]等;另一种是基于 GNN 的方法,例如 EMGCN^[16]、HGCN^[17]、RDGCN^[18]和 EVA^[19]等。前者根据三元组之间的内部约束关系学习向量表示,后者通过图卷积和图注意力网络融合邻域节点的信息得到向量表示。

通过计算不同实体向量的距离就能得到实体对,然而仅通过计算距离得到的对齐结果往往面临着一对多等错误对齐的问题。

现有的许多模型一般在预对齐后通过对齐优化调整来修正错误的对齐、提高准确性。目前实体对齐模型中优化调整所采用的方法主要分为表 1 中所示的几类。本文对各类实体对齐模型中的对齐优化调整算法进行了梳理,如表 2 所示,其中,√表示使用该方法,×表示不使用该方法。

表 1 对齐优化调整所采用的方法

Table 1 Methods adopted for alignment optimization adjustment

基本原则	主要思想	代表性方法
1-1 约束性原则	将对齐任务转换为匹配任务,且一个图谱中实体只能与另一图谱中最多一个实体匹配	DAA ^[11] , Sinkhorn ^[20]
双向一致性原则	实体对齐是两个知识图谱间的双向匹配问题	BN ^[10]
关系实体一致性原则	利用实体对齐和关系对齐之间的正交互作用	关系感知邻居匹配 ^[21] , 模糊对齐调整算法 ^[22] , 对齐矩阵融合算法 ^[23]

表 2 实体对齐模型中的对齐优化调整算法

Table 2 Alignment optimization and adjustment algorithms in entity alignment models

模型	DAA	Sinkhorn	BN	关系实体感知类
CEA ^[24]	✓	×	×	×
SEU ^[25]	×	✓	×	×
LightEA ^[26]	×	✓	×	×
RAGA ^[10]	✓	×	✓	×
TTEA ^[27]	✓	×	✓	×
OTIEA ^[28]	✓	×	✓	×
MCLEA ^[29]	×	×	×	×
RNM ^[21]	×	×	×	✓
MOGNN ^[22]	✓	×	✓	✓
MHEA ^[23]	✓	×	✓	✓

CEA^[24]模型为了捕捉实体间的相互依赖关系,减少实体对齐中一对多或对齐失败等问题,将实体对齐表述为经典的稳定匹配问题,使用 DAA 延迟接受方法来保证每个实体都得到匹配。

DAA 的工作原理如下:

第 1 步,每个源实体向它最喜欢(即相似度最高)的目标实体配对,然后目标实体与目前为止它最喜欢的源实体进行临时匹配。

第 2 步,每个未匹配的源实体向其尚未配对过的最喜欢的目标实体配对,然后每个目标实体也会选择最优的源实体进行临时匹配。对应有两种情况:一种是接受此轮源实体的配对请求,并且若此目标实体在之前轮次已经有临时匹配的源实体,则该源实体会被断开匹配,进入下一轮的匹配中;另一种是由于该目标实体更喜欢之前已临时配对的源实体而不是此轮的源实体,从而拒绝此轮的源实体。最后所有未能匹配到目标实体的源实体进入下一轮匹配。

重复第 2 步,直到每个源实体都配对成功。

图 1 中的示例简单模拟了 DAA 的工作原理(彩色效果见《计算机工程》官网 HTML 版,下同)。假设有 3 个源实体 x_1, x_2, x_3 和 3 个目标实体 y_1, y_2, y_3 , 给定相似矩阵 S , 其中矩阵中的值代表对应实体对的相似程度,值越大代表两个实体越相似。

第 1 轮, x_1 和 x_2 都向 y_1 配对, 由于 y_1 更喜欢 x_1 , 则 y_1 接受 x_1 , 两者临时匹配, x_3 向 y_2 配对并临时匹配。

第 2 轮, x_2 向除了已经配对过的 y_1 之外最喜欢的 y_2 配对, 由于 y_2 更喜欢 x_2 , 因此 y_2 会拒绝 x_3 , 接受 x_2 的配对。

第 3 轮, x_3 向未配对过的 y_3 配对并临时匹配。

至此,所有源实体都配对成功。

可以看到,不使用 DAA 会出现 y_1 与 x_1 和 x_2 同时配对的情况,并导致 x_3 和 y_2 错误对齐,使用 DAA 可有效解决这些问题。

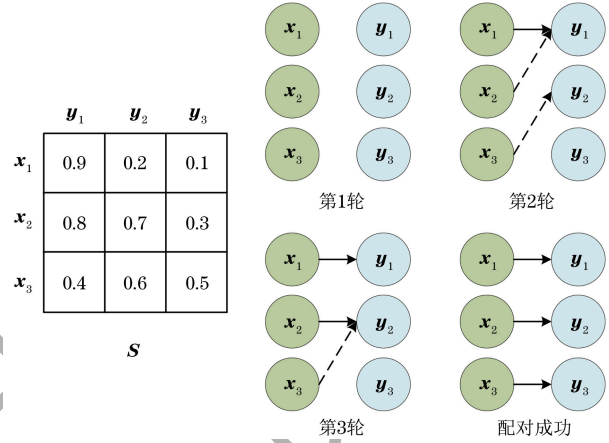


图 1 DAA 原理示例

Fig.1 Schematic diagram of DAA

考虑到基于 GNN 的实体对齐方法有可解释性差和效率低的缺点, MAO 等^[25]基于同构假设,将实体对齐问题转换为分配问题,并使用 SEU (Simple but Effective Unsupervised entity alignment method)来处理,将时间复杂度从 $O(n^3)$ 提升到 $O(n^2)$ 。

考虑到 Sinkhorn 中的指数归一化会导致对齐矩阵中包含许多无限接近 0 的值,这些值对应的实体对齐概率非常小,删除这些值对结果影响较小, LightEA^[26]模型中提出了一种稀疏的 Sinkhorn 方法,可将时间复杂度降为 $O(n)$ 。

RAGA^[10]模型考虑到实体对齐的局部匹配可能会导致多对一的错误对齐,设计了全局对齐算法,并且认为实体对齐是两个图谱间的双向匹配问题,对齐矩阵中对应的行和列分别代表了两个方向的对齐信息。将对齐矩阵的行和列都使用 Softmax 处理,并将横向和纵向结果相加得到新的对齐矩阵,从而得到融合了两个方向的对齐信息,最后采用 DAA 对更新后的对齐矩阵进行全局对齐。BN 的计算公式如式(1)所示:

$$S_{i,j}^s = \frac{\exp(S_{i,j})}{\sum_{j'=1}^{|E_2|} \exp(S_{i,j'})} + \frac{\exp(S_{i',j})}{\sum_{i'=1}^{|E_1|} \exp(S_{i',j})} \quad (1)$$

式中: $S_{i,j}$ 表示预对齐矩阵 S 中第 i 行 j 列对应的相似距离值,即源图谱第 i 个实体和目标图谱第 j 个实体的距离; $S_{i,j}^s$ 表示经过 Softmax 的行和列并相加得到的相似距离值。

TTEA^[27]和 OTIEA^[28]模型参考了 RAGA 的

思想,也使用了 BN 和 DAA 调整对齐矩阵,并且强调了关系三元组在嵌入时的不可分割性。

MCLEA^[29]模型在实体对齐中利用多模态的信息对实体进行联合表示,除了使用名称和属性信息外,还使用了图像信息以增强对齐准确率。

关系作为实体间的桥梁对实体对齐有重要作用,RNM^[21]模型利用实体对齐和关系对齐之间的正交互作用,提出了关系感知邻域匹配和实体感知关系匹配,并设置迭代策略更新预对齐对,从而有效利用了关系信息。其思想是如果两个图谱中的一对实体所对应的邻域实体和对应关系已经预对齐,那么这对实体对齐的概率应该变大。同理,如果两个图谱中的实体和邻域实体已经预对齐,那么其之间的关系对齐的概率也应该变大。

受 RNM 启发,本文在 MOGNN^[22] 和 MHEA^[23]模型中进行了尝试,利用实体对齐和关系对齐的内在一致性,先后提出了模糊对齐调整算法和对齐矩阵融合算法进行对齐矩阵的优化调整,最后利用 DAA 调整得到最终对齐结果,有效减少了实体对齐中存在的一对多问题。在此基础上,本文沿用该思想提出了关系实体感知调整算法。

然而,上述实体对齐的调整优化方法均是从两个图谱间对齐的角度进行优化调整。在多图谱融合任务中,也仅考虑两两图谱间的对齐优化,忽略了同一物理实体在多个图谱上的相似性特征,包括实体属性的相似性和邻域结构的相似性。针对上述不足,本文提出了面向多知识图谱的实体对齐优化调整方法,以解决多图谱实体对齐时信息利用不充分的问题。

2 问题描述

一个知识图谱可定义为 $G=(E,R,T)$,其中, E 是实体集合, R 是关系集合, T 包括关系三元组集合和属性三元组集合,给定 N 个图谱 $\{G_k | k=1, 2, \dots, N, N \geq 3\}$,本文将多图谱的实体对齐分为有

中心图谱的对齐和无中心图谱的对齐。

对于有中心图谱的实体对齐, G_1 为中心图谱,对应有 $N-1$ 个边缘图谱 $G_n (n \in [2, N])$,即 G_1 是源图谱, G_n 是目标图谱。对齐任务是找出 G_1 和 G_n 中存在对应关系的实体对,可表示为 $L = \{e_i^1 \equiv e_j^n | e_i^1 \in G_1, e_j^n \in G_n\}$,其中, e_i^1 和 e_j^n 分别是 G_1 和 G_n 中的第 i 和 j 个实体;而对于无中心图谱的对齐, N 个图谱两两进行实体对齐,即 G_m 和 $G_n (m, n \in k, m \neq n)$ 进行实体对齐,其中, G_m 是源图谱, G_n 是目标图谱,实体对齐的任务是找出 G_m 和 G_n 中存在对应关系的实体对。

根据实体在多图谱中的等价性质,本文将实体分为 3 类:非对齐实体(NAE),该实体仅存在于一个图谱中,在其他各图谱中均没有等价实体;部分对齐实体(PAE),该实体仅在部分图谱中存在等价实体;全对齐实体(FAE):该实体在各图谱中均有等价实体。示例如图 2 所示(以三图谱为例)。

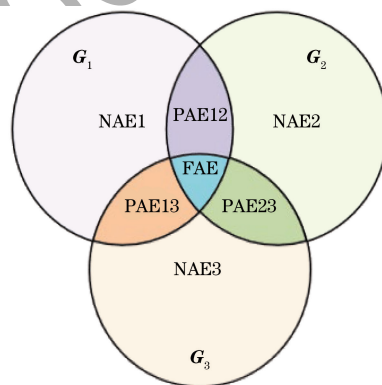


图 2 多知识图谱数据集中的实体分布

Fig.2 Distribution of entities in multiple knowledge graphs dataset

3 多知识图谱的实体对齐优化方法

本文提出的面向多知识图谱融合的实体对齐优化方法与嵌入表示和预对齐过程共同组成一种通用的多知识图谱的实体对齐优化框架,如图 3 所示。

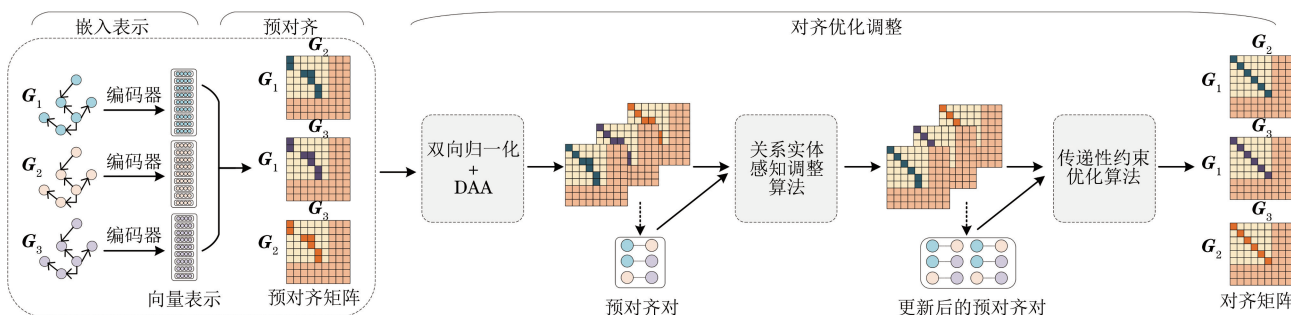


图 3 多知识图谱的实体对齐优化框架

Fig.3 Entity alignment optimization framework for multiple knowledge graphs

具体过程如下:

1) 通过嵌入表示获得图谱中实体和关系的向量表示。不限制编码器的种类, 可使用现有的基于表示学习的各类实体对齐模型中的编码器。

2) 通过计算实体嵌入间的距离得到预对齐矩阵, 完成预对齐的过程。

3) 通过多种优化算法对预对齐矩阵进行调整优化。首先对预对齐矩阵的行和列做归一化并相加, 即通过 BN 得到融合了双向对齐信息的矩阵, 然后通过 DAA 进行全局对齐保证每个实体都得到匹配, 这样经过 BN 和 DAA 的优化, 可以初步提高对齐的准确率, 得到初步优化的预对齐矩阵, 并生成初步的实体预对齐对和关系预对齐对; 利用实体和关系的正相互作用, 使用关系实体感知调整算法进一步优化并得到新的预对齐矩阵, 并更新预对齐对的集合。

4) 利用传递性规则建立多图谱的连接, 使用传递性约束优化算法得到最终的对齐矩阵。

3.1 关系实体感知调整算法

本节利用实体对齐和关系对齐之间的正交互作用, 使用预对齐的实体对和关系对对预对齐矩阵进行优化调整。

首先要找出距离最近的预对齐对。为了保证预对齐对的准确性, 使用阈值对其进行过滤; 然后根据预对齐对对预对齐矩阵进行调整, 让实体的邻域实体中更为匹配的实体对间的距离更小。

调整优化的过程主要包括两部分:

一部分基于如下假设: 如果两个头实体有对齐关系, 若其对应的关系和尾实体都有对齐关系, 那么这两个实体对齐的概率应被提高。与 RNM 不同的是, 本文进行优化时已经确定了被优化的这对头实体, 只需要找出满足关系和尾实体都能对齐的个数进行计算, 而 RNM 中考虑了图谱对所有存在这种对齐关系的头实体对, 考虑了多对一情况, 使用了概率累加和的方式进行计算。

另一部分则是基于如下假设: 如果两个关系有对齐关系, 若其对应的头实体和尾实体也都有对齐关系, 那么两个关系对齐的概率应该被提高, 即让距离矩阵中的值更小, 在初始距离矩阵上对相应位置进行优化, 减去相应的分值。计算公式如下:

$$d_{ij}^e = \|x_i - x_j\|_1 - \theta_e \cdot \frac{|N|}{|N_i| + |N_j|} \quad (2)$$

$$d_{ij}^r = \|r_i - r_j\|_1 - \theta_r \cdot \frac{|S|}{|S_i| + |S_j|} \quad (3)$$

式中: x_i 和 x_j 分别表示第 i 和 j 个实体的向量; r_i 和 r_j 分别表示第 i 和 j 个关系的向量; N_i 和 N_j 分别是实体 i 和 j 的邻域实体关系集合; N 是满足预对齐关系的实体关系集合; S_i 和 S_j 分别是关系 i 和 j 对应的实体集合; S 是满足预对齐关系的实体集合; θ_e 和 θ_r 是超参数, 通过设置其大小可以控制优化调整的程度; d_{ij}^e 和 d_{ij}^r 分别是经过计算得到的新的实体之间的距离和关系之间的距离。

具体步骤如算法 1 所示。

算法 1 关系实体感知调整算法

输入 实体表示 X_e , 关系表示 X_r , 迭代次数 T_1 , 实体集合 E_1 和 E_2 , 关系集合 R_1 和 R_2 , 邻域实体集合 N_1 和 N_2 , 实体距离阈值 λ_e , 关系距离阈值 λ_r

输出 实体预对齐对 L_e , 实体对齐矩阵 S_e

1. 根据 X_e 计算实体预对齐矩阵 S_e
2. 根据 X_r 计算关系预对齐矩阵 S_r
3. 初始化实体和关系预对齐对 $L_e, L_r \leftarrow \emptyset$
4. 初始化实体关系集合和实体集合 $N, S \leftarrow \emptyset$
5. for $t \leftarrow 1$ to T_1 do
6. for e_i in E_1 do
7. $j = \operatorname{argmin}_j d_{ij}^e$
8. if $d_{ij}^e < \lambda_e$ then
9. $L_e \leftarrow L_e \cup (e_i, e_j)$
10. end if
11. end for
12. for r_i in R_1 do
13. $j = \operatorname{argmin}_j d_{ij}^r$
14. if $d_{ij}^r < \lambda_r$ then
15. $L_r \leftarrow L_r \cup (r_i, r_j)$
16. end if
17. end for
18. for e_i in E_1 and e_j in E_2 do
19. if $(n_{ie}, n_{je}) \in L_e$ and $(n_{ir}, n_{jr}) \in L_r$
20. $N \leftarrow N \cup (n_i, n_j)$
21. end if
22. end for
23. 使用式(2)得到新的实体对齐矩阵 S_e
24. for r_i in R_1 and r_j in R_2 do
25. if $(s_{ie}, s_{je}) \in L_e$ then
26. $S \leftarrow S \cup (s_i, s_j)$
27. end if
28. end for
29. 使用式(3)得到新的关系对齐矩阵 S_r
30. end for

在算法 1 中, 实体距离阈值 λ_e 和关系距离阈值 λ_r 可以把距离较远的过滤掉, 从而减少运算量, n_{ie} 和 n_{je} 分别代表实体 e_i 和 e_j 的邻域实体, n_{ir} 和 n_{jr}

代表与 n_{ie} 和 n_{je} 对应的关系, s_{ie} 和 s_{je} 分别代表关系 r_i 和 r_j 对应的实体。

3.2 传递性约束优化算法

通过关系实体感知调整算法得到更新的两两图谱间的预对齐矩阵和预对齐后, 本节进一步利用传递性规则传递实体间的对齐信息, 把多图谱间的实体联系起来进行实体对齐优化。

传递性约束规则是在表 1 的 3 种基本实体对齐优化原则之外的第 4 种原则, 适用于多图谱对齐场景。其基本思想是: 如果 $A=B$ 且 $A=C$, 则 $B=C$ 。如图 4 所示的示例中, 假设有 G_1 、 G_2 和 G_3 3 个知识图谱, 每个图谱有若干实体和关系, 图谱两两进行融合。若通过计算实体间的距离后得到实体 A_1 和 B_1 对齐, 实体 B_1 和 C_1 也对齐, 则实体 A_1 和 C_1 也对齐的概率是比较大的, 若预对齐结果中实体 A_1 和实体 C_1 没有对齐, 那么其中一定存在着错误的对齐需要进行修正。

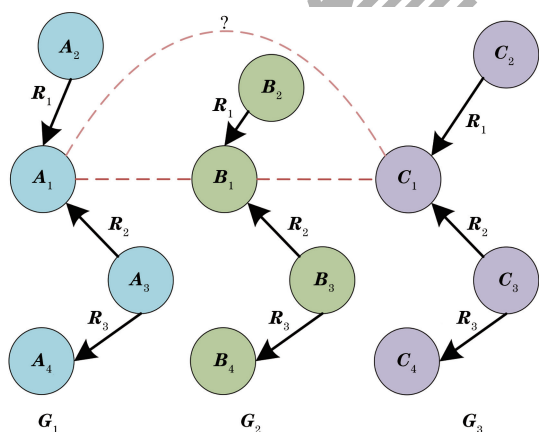


图 4 实体对齐中的传递性规则

Fig. 4 Transitive rules in entity alignment

针对实体 A_1 和 C_1 没有对齐的问题, 本文设计了 3 种处理方法:

方法 1: 直接利用传递性规则的约束, 将实体 A_1 和实体 C_1 对齐。

方法 2: 假设不同图谱中等价的实体其对应的相邻实体也具有一定的相似性, 通过计算实体 A_1 和 C_1 的邻域实体的对齐情况, 即计算其一阶邻域实体对齐的比例, 利用图谱间的结构相似性约束比较两个实体邻域的这种相似性, 并通过设置阈值确定实体 A_1 和 C_1 是否对齐。

方法 3: 利用预对齐矩阵中的距离信息, 即通过距离值的大小判断。如果实体 A_1 和 C_1 在预对齐矩阵中的距离不是最小, 但只要其足够小, 排名达到设定的阈值, 则可以认为实体 A_1 和 C_1 对齐。

这样就可以利用传递性规则在多图谱之间建立关联关系, 让符合条件的实体对的距离更加接近, 从而对预对齐矩阵进行优化, 公式如下:

$$d_{ij} = \|x_i - x_j\|_1 - \theta \quad (4)$$

式中: x_i 和 x_j 分别表示第 i 和 j 个实体的向量; θ 是超参数, 可以控制距离优化的程度; d_{ij} 是经过计算得到的新的实体之间的距离值。

具体步骤如算法 2 所示(以三图谱为例)。

算法 2 传递性约束优化算法

输入 实体预对齐矩阵 S_1 、 S_2 和 S_3 , 实体预对齐对 L_{12} 、 L_{32} 和 L_{13} , 循环次数 T_2 , 邻域相似阈值 γ_n , 距离排序阈值 d_n

输出 实体对齐矩阵 S_{1e} 、 S_{2e} 和 S_{3e}

1. for $i, j, k \leftarrow 1$ to T_2 do
2. if (e_i, e_k) in L_{13} and (e_k, e_j) in L_{32} then
3. if (e_i, e_j) not in L_{12} then
4. if $\gamma_{ij} \geq \gamma_n$ or $d_{ij} \leq d_n$ then
5. $L_{12} \leftarrow L_{12} \cup (e_i, e_j)$
6. 使用式(4)更新预对齐矩阵 S_1
7. end if
8. end if
9. end if
10. 同理对 L_{32} 和 L_{13} 进行优化
11. 使用式(4)更新预对齐矩阵 S_2 和 S_3
12. 得到更新后的对齐矩阵 S_{1e} 、 S_{2e} 和 S_{3e}
13. end for

在算法 2 中, 邻域相似阈值 γ_n 和距离排序阈值 d_n 分别对应方法 2 和方法 3 中所述情况, 方法 1 中所述情形不需要设置阈值。在后文的消融实验中会对 3 种方法做进一步的讨论。经过算法 2 处理, 实体预对齐矩阵被进一步优化并得到最终的对齐结果。

4 实验结果及分析

4.1 数据集

本实验使用了多个实体对齐数据集。首先是实体对齐研究中普遍使用的数据集 DBP15K, 包括多种语言的知识图谱数据集, 实验使用了汉语、英语和法语 3 种语言的知识图谱。原始数据集中的图谱都是成对的, 其中中-英数据集和法-英数据集中的英文图谱不完全相同。为了实验的需要, 本文对两个原始英文图谱中的实体和关系做了合并去重处理得到新的唯一的英文图谱, 并更新了 3 种语言数据集的实体编号, 最终得到更新后的 DBP15K 数据集; 另外还使用了 FB15K、DB15K 和 YAGO15K 3 种数据集做了跨图谱的实体对齐。各个数据集的统计数据如表 3 所示。

表 3 各个数据集的统计数据
Table 3 Statistics for various datasets 单位:个

数据集	实体数量	关系数量	关系三元组数量
DBP15K _{ZH} (中)	19 388	1 701	70 414
DBP15K _{EN} (英)	34 342	2 121	210 864
DBP15K _{FR} (法)	19 661	1 299	105 998
FB15K	14 951	1 367	592 213
DB15K	12 842	263	99 028
YAGO15K	15 404	32	122 886

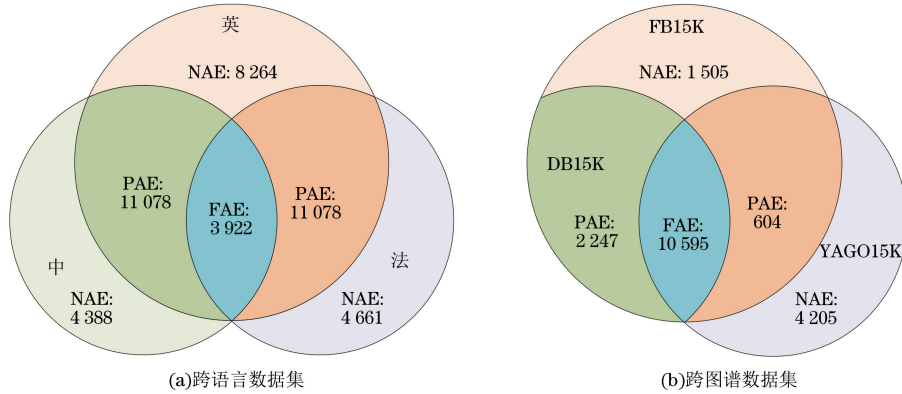


图 5 数据集中的实体分布

Fig.5 Distribution of entities in datasets

4.2 基线模型

本节将本文提出的面向多知识图谱融合的实体对齐优化方法在 6 个不同的实体对齐模型中进行了实验,包括 RAGA、TTEA、OTIEA、MCLEA、MOGNN 和 MHEA。

4.3 评价指标与实验设置

为了更好地评价多图谱实体对齐任务中的算法性能,本文除了采用目前常用的 3 个指标 Hits@1、Hits@10、平均倒数排名(MRR)之外,还针对无中心的多图谱实体对齐任务定义了两个新指标 FHits@1 (Full-aligned Hits@1) 和 WHits@1 (Weighted average Hits@1)。考虑到有中心的多图谱实体对齐任务是中心图谱和每个边缘图谱间的对齐,即两两对齐,暂不考虑更多的指标。本文的实验针对无中心的多图谱实体对齐任务展开。

FHits@1 针对的是所有的全对齐实体,表示对于图谱中的所有全对齐实体能够最终被正确对齐的比例,该指标可以直观地衡量传递性所带来的性能提升;WHits@1 针对的是图谱对中所有的实体对(即全对齐实体和部分对齐实体的总和),计算它们能被正确对齐的比例,即对各图谱对的 Hits@1 进行加权平均,该指标代表了整体的对齐水平。FHits@1 和 WHits@1 计算公式如下:

$$I_{\text{FHits@1}} = \frac{C_{\text{FAE}}}{N_{\text{FAE}}} \quad (5)$$

数据集中各类实体的分布情况如图 5 所示。左图的中-法数据集间没有部分对齐实体,是因为在对原始 DBP15K 数据集处理时是根据中-英和法-英的等价实体对提取得到的中-法的实体对,故中-法间都是全对齐实体;右图中的 DB15K 数据集之所以没有非对齐实体,是因为 FB15K-DB15K 这对数据集中,DB15K 中的所有实体均与 FB15K 中的实体有对齐关系。

$$I_{\text{WHits@1}} = \frac{\sum_{i=1}^{C_N^2} N_{\text{TAE}_i}}{C_N^2 N_{\text{FAE}} + \sum_{i=1}^{C_N^2} N_{\text{PAE}_i}} \quad (6)$$

式中: C_{FAE} 表示被正确对齐的全对齐实体数量; N_{FAE} 表示所有全对齐实体数量; N 表示图谱数量; C_N^2 表示图谱对的数量; N_{TAE_i} 表示第 i 个图谱对中正确对齐的实体数量; N_{PAE_i} 表示第 i 个图谱对中部分对齐实体的数量。

针对 DBP15K 跨语言数据集,本文实验使用 30% 的预对齐实体对作为训练集,剩余的 70% 作为测试集,使用 80 个 Epoch 进行训练;针对 FB-DB-YAGO 跨图谱数据集,使用 80% 的预对齐实体对作为训练集,剩余的 20% 作为测试集,使用 200 个 Epoch 进行训练。关系实体感知调整算法中取 $\theta_e = 10$, $\theta_r = 200$, $\lambda_e = 30$, $\lambda_r = 2$,传递性约束优化算法中取 $\theta = 5$, $\gamma_n = 0.2$, $d_n = 30$ 。

4.4 实验结果

4.4.1 对比实验 1(DBP15K 数据集)

本节采用 DBP15K 进行跨语言知识图谱实体对齐实验。基线模型分别采用和不采用 MGEAO 处理的实验结果如表 4 和图 6 所示。其中基线模型中使用过的优化方法不在 MGEAO 中重复使用,指标得到提升的用加粗格式标注,提升最大的用下划线标注,下同。

表 4 跨语言数据集的实体对齐结果对比

Table 4 Comparison of entity alignment results in cross-lingual datasets

Model	ZH-EN			FR-EN			ZH-FR		
	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR
RAGA	89.72	94.85	0.914 2	96.35	98.21	0.969 7	86.92	93.77	0.891 5
RAGA+MGEOA	92.89	96.71	0.942 4	97.64	98.91	0.980 8	94.79	96.72	0.954 4
TTEA	89.90	95.10	0.916 9	96.72	98.70	0.973 9	88.82	94.06	0.906 4
TTEA+MGEOA	93.14	97.15	0.945 5	98.04	99.26	0.984 6	95.41	97.09	0.959 5
OTIEA	89.71	94.92	0.915 0	97.00	98.87	0.976 5	87.98	94.46	0.901 3
OTIEA+MGEOA	93.04	97.11	0.944 7	98.16	99.37	0.986 0	95.48	97.30	0.960 7
MCLEA	96.84	98.76	0.975 1	99.48	99.82	0.996 1	95.63	98.14	0.965 0
MCLEA+MGEOA	97.82	99.22	0.983 5	99.71	99.91	0.998 0	98.72	99.23	0.989 6
MOGNN	95.58	97.46	0.962 4	99.42	99.72	0.995 3	95.12	97.30	0.957 8
MOGNN+MGEOA	96.24	97.90	0.968 2	99.56	99.77	0.996 5	98.03	98.72	0.982 5
MHEA	98.11	99.25	0.984 9	99.46	99.86	0.996 1	97.67	99.13	0.981 8
MHEA+MGEOA	98.44	99.45	0.988 1	99.61	99.90	0.997 1	99.16	99.56	0.993 3
Improve best/percentage points	3.33	2.19	0.297 0	1.32	0.70	0.111 0	7.87	3.03	0.062 9

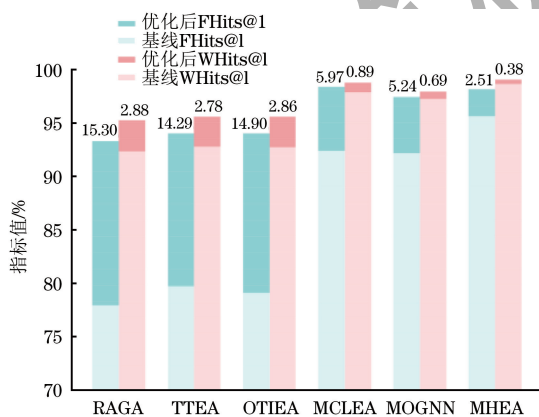


图 6 跨语言数据集的实体对齐结果对比

Fig. 6 Comparison of entity alignment results in cross-lingual KGs

由表 4 可以看出,基线模型在加入 MGEOA 后性能均得到了显著提升,其中 Hits@1、Hits@10 和 MRR 分别最大可提高 7.87、3.03 百分点和 0.062 9,证明了该优化方法的有效性。由于本节使用的数据集与基线的数据集在英语实体构成上存在一些差异,且为了对比各基线模型采用 MGEOA 处理后的性能差异,因此本文对 6 个基线模型均进行了复现。

可以看到 FR-EN 的指标提升相较于 ZH-FR 和 ZH-EN 略低,原因是法语图谱中包含更多的关系三元组,基线模型的指标已经很高,导致提升有限。ZH-FR 的各项指标提升最高,部分原因可能是由于中法数据集中对齐的均为全对齐实体,而 ZH-EN 和 FR-EN 中还包含大量部分对齐实体,而传递性约束优化算法更多的是优化全对齐实体,从而使 ZH-FR 的指标提升较为理想。基于此,本

文提出了第 2 类评价指标来进一步观察优化方法的有效性。

如图 6 所示, FHits@1 指标的提升非常明显,其中 FHits@1 和 WHits@1 指标分别最大可提高 15.3 和 2.88 百分点。FHits@1 的增加意味着 MGEOA 使得更多的全对齐实体被正确对齐,证明 MGEOA 在优化全对齐实体时很有效,传递性约束在其中起到较大的作用。另外,由于实体间存在着一对一关系,当全对齐实体被优化方法纠正对齐后,部分对齐实体也会受其影响被纠正对齐,因为修正一个错误对齐的同时也会影响与它有关的其他的错误对齐,即全对齐实体的正确对齐也会使部分对齐实体的对齐性能提升。

从指标的结果上来看, FHits@1 的初始结果低,它会提升到和 WHits@1 相近的水平,而 FHits@1 的提升也会使得 WHits@1 得到提升;从两个指标的计算公式上看, FHits@1 针对的是全对齐实体,而 WHits@1 针对的是全对齐实体和部分对齐实体,后者的分母明显更大,所以较前者提升幅度不明显, WHits@1 的结果也代表了多图谱对齐的整体水平。两个指标的提升进一步验证了 MGEOA 的有效性。

4.4.2 对比实验 2(FB-DB-YAGO 数据集)

本节进一步在数据集 FB15K、DB15K 和 YAGO15K 中进行跨图谱的实体对齐实验,结果如表 5 所示。可以看出,加入 MGEOA 后各指标均有不同比例的提高。其中 Hits@1、Hits@10 和 MRR 指标分别最大可提高 18.8、18.05 百分点和 0.184 6。可以看到在 FB15K-DB15K 和 FB15K-

YAGO15K 两对跨图谱数据集中的提升幅度比较明显,可能是由于其初始对齐指标结果较低,有着很大的提升空间。而由于 DB15K-YAGO15K 初始对

齐结果正确率已经很高,使得方法所带来的优化效果不明显。总体来看,提出的方法在跨图谱的异构数据集中有着不错的表现。

表 5 跨图谱数据集的实体对齐结果对比

Table 5 Comparison of entity alignment results in cross-graph datasets

Model	FB15K-DB15K			FB15K-YAGO15K			DB15K-YAGO15K		
	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR
RAGA	63.94	71.34	0.667 6	53.15	57.75	0.550 0	99.20	99.43	0.992 9
RAGA+MGEAO	71.42	77.18	0.736 0	69.50	72.26	0.706 0	99.24	99.43	0.993 2
TTEA	65.03	73.60	0.683 0	58.33	64.81	0.607 1	99.39	99.48	0.994 4
TTEA+MGEAO	74.53	80.88	0.769 8	73.51	77.58	0.751 0	99.39	99.48	0.994 4
OTIEA	66.67	75.04	0.697 5	59.54	66.41	0.622 9	99.62	99.67	0.996 5
OTIEA+MGEAO	75.66	81.74	0.778 3	74.68	78.74	0.763 1	99.67	99.72	0.996 9
MCLEA	81.78	90.93	0.849 3	75.44	88.34	0.798 5	99.48	99.62	0.995 5
MCLEA+MGEAO	87.66	94.16	0.899 0	86.78	93.75	0.891 2	99.53	99.62	0.995 8
MOGNN	60.12	64.06	0.616 4	46.54	48.68	0.474 3	99.53	99.58	0.995 4
MOGNN+MGEAO	67.10	70.52	0.684 1	64.81	66.06	0.653 2	99.53	99.58	0.995 4
MHEA	62.69	65.58	0.637 6	48.46	50.11	0.492 6	99.53	99.53	0.995 3
MHEA+MGEAO	70.25	72.82	0.711 6	67.26	68.16	0.677 2	99.53	99.53	0.995 3
Improve best/percentage points	9.50	7.28	0.086 8	18.80	18.05	0.184 6	0.05	0.05	0.000 4

4.4.3 消融实验

为了进一步验证传递性约束优化算法(TCO)中针对实体 A_1 和 C_1 没有对齐的问题所设置的三种处理方法的有效性,使用 DBP15K 数据集在 RAGA 模型上进行测试,将基线、方法 1、方法 2、方法 3、联合方法 2 和方法 3 的 5 种结果进行了对比,

如图 7 所示。

由图 7 可以看出,相较于基线模型,各方法都有不同程度的性能提高。方法 2 和方法 3 都比方法 1 更优,说明在使用传递性进行优化时,利用好实体的邻域结构信息会得到更好结果,联合使用方法 2 和方法 3 时效果最优。

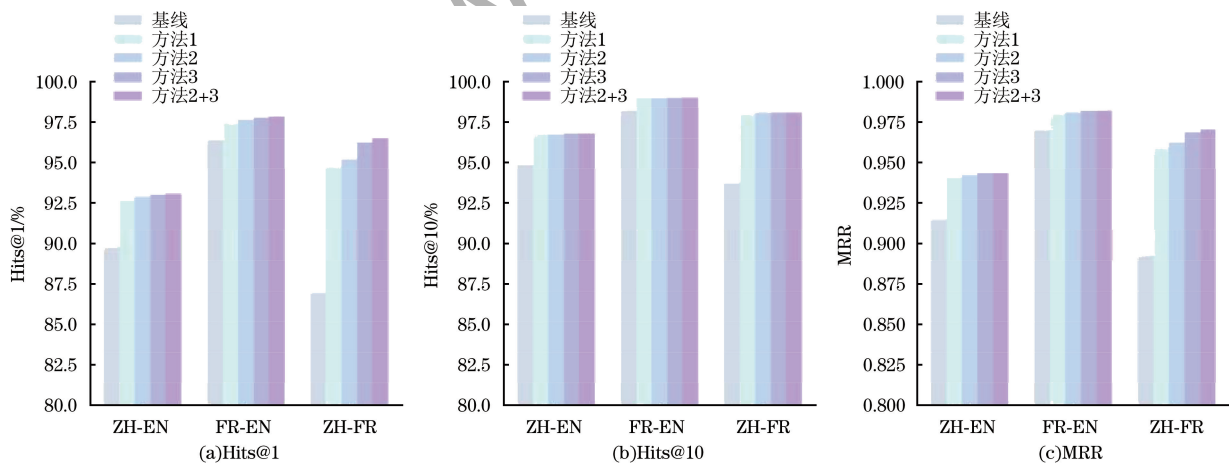


图 7 TCO 在不同处理方法下的实体对齐性能的对比

Fig. 7 Comparison of entity alignment performance of TCO under different processing methods

为了进一步验证 MGEAO 中各优化方法的有效性,本节使用 DBP15K 数据集在 RAGA 基线模型上进行测试,对比了使用 BN、DAA、REA、TCO、BN+DAA、REA+TCO、BN+DAA+REA、BN+DAA+TCO 以及 MGEAO (即 BN+DAA+REA+TCO) 的 9 种结果,如表 6 所示。

由表 6 可以看出,使用 DAA 优于 BN,单独

使用 BN 的效果并不理想,但是两者联合使用后的结果会有提升,说明 BN 是有效的。在使用 BN 和 DAA 后,无论是和关系实体感知调整联合,还是和传递性约束优化联合,都会使指标得到提升,验证了 REA 和 TCO 算法的有效性,并且 MGEAO 的效果达到最优,验证了所提方法的有效性。

表 6 不同对齐优化方法的实验结果对比

Table 6 Comparison of experimental results of different alignment optimization methods

Model	ZH-EN			FR-EN			ZH-FR		
	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR	Hits@1/%	Hits@10/%	MRR
BN	77.98	93.12	0.834 8	87.23	96.19	0.906 1	73.84	91.95	0.803 0
DAA	86.87	93.83	0.891 8	93.67	96.98	0.947 9	83.93	92.75	0.868 7
REA	81.71	95.19	0.868 5	88.17	97.30	0.916 2	77.45	93.55	0.832 7
TCO	79.70	93.31	0.847 1	89.88	96.86	0.924 7	81.60	93.11	0.858 0
BN+DAA	89.72	94.85	0.914 2	96.35	98.21	0.969 7	86.92	93.77	0.891 5
REA+TCO	82.65	95.34	0.874 8	89.87	97.63	0.928 2	85.43	95.05	0.888 8
BN+DAA+REA	92.24	96.43	0.937 2	97.15	98.75	0.976 9	89.29	95.19	0.912 6
BN+DAA+TCO	90.56	95.18	0.920 9	97.08	98.49	0.975 3	92.90	95.45	0.937 6
MGEAO	92.89	96.71	0.942 4	97.64	98.91	0.980 8	94.79	96.72	0.954 4

5 结束语

参考文献

现有的各类实体对齐模型通常仅面向两个图谱间的对齐任务,并没有考虑多图谱间实体对齐的场景。虽然也可以将多图谱间的对齐任务转化为多个两两图谱的对齐任务并直接采用上述模型进行处理,但普遍存在不能充分利用多图谱间内在信息和约束关系、对齐效果欠佳等问题。因此,本文提出了一种面向多知识图谱融合的实体对齐优化方法(MGEAO)和一种通用的多图谱实体对齐优化框架,并设计了更有效的指标来评价多图谱对齐性能。MGEAO首先通过BN和DAA初步优化对齐矩阵,然后通过所提出的关系实体感知调整算法,利用实体和关系的交互作用进一步优化调整预对齐结果,最后基于传递性规则提出并应用传递性约束优化算法建立多图谱间的连接,优化对齐矩阵从而有效提升最终的多图谱间实体对齐性能。该框架的通用性体现在其可应用于基于表示学习的现有各类对齐模型中,使用MGEAO作为其后处理过程对预对齐矩阵进行优化。

本文采用了6种基线模型在DBP15K上做了跨语言的实体对齐实验,在FB15K、DB15K和YAGO15K数据集上做了跨图谱的实验,均获得了显著的性能提升。消融实验也验证了各个优化方法的有效性。

在后续研究中,将会进一步探索如何更有效地表达和利用多图谱间的内在的相似性和传递性约束关系,包括引入强化学习和表示学习等方法对本文所提方法进行增强,从而进一步提升多图谱的实体对齐性能;此外,将新方法在更大规模和更实际的业务场景下的多图谱数据集上进行更广泛的有效性验证也是下一步的工作之一。

- [1] CAO Y X, WANG X, HE X N, et al. Unifying knowledge graph learning and recommendation: towards a better understanding of user preferences [C] // Proceedings of the World Wide Web Conference. New York, USA: ACM Press, 2019: 151-161.
- [2] 陈晓军, 向阳. 企业风险知识图谱的构建及应用[J]. 计算机科学, 2020, 47(11): 237-243.
CHEN X J, XIANG Y. Construction and application of enterprise risk knowledge graph [J]. Computer Science, 2020, 47(11): 237-243. (in Chinese)
- [3] WANG H W, ZHANG F Z, ZHAO M, et al. Multi-task feature learning for knowledge graph enhanced recommendation [C] // Proceedings of the World Wide Web Conference. New York, USA: ACM Press, 2019: 2000-2010.
- [4] BIZER C, LEHMANN J, KOBILAROV G, et al. DBpedia-a crystallization point for the web of data[J]. Journal of Web Semantics, 2009, 7(3): 154-165.
- [5] SUCHANEK F M, KASNECI G, WEIKUM G. YAGO: a large ontology from Wikipedia and WordNet [J]. Journal of Web Semantics, 2008, 6(3): 203-217.
- [6] SPEER R, CHIN J, HAVASI C. ConceptNet 5.5: an open multilingual graph of general knowledge [C] // Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: ACM Press, 2017: 4444-4451.
- [7] VRANDEČIĆ D, KRÖTZSCH M. Wikidata: a free collaborative knowledgebase [J]. Communications of the ACM, 2014, 57(10): 78-85.
- [8] 朱红, 王阔然, 朱彤. 基于多侧面信息表征联合的实体相似性度量及对齐方法[J]. 计算机工程, 2025, 51(3): 64-75.
ZHU H, WANG K R, ZHU T. Entity similarity metrics and alignment method based on the union of multi-side information representations [J]. Computer Engineering, 2025, 51(3): 64-75. (in Chinese)
- [9] WANG Z C, LV Q S, LAN X H, et al. Cross-lingual knowledge graph alignment via graph convolutional networks [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: Association for Computational Linguistics, 2018: 349-357.
- [10] ZHU R B, MA M, WANG P. RAGA: relation-aware graph attention networks for global entity alignment [C] // Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Berlin, Germany: Springer, 2021: 501-513.
- [11] ROTH A E. Deferred acceptance algorithms: history, theory, practice, and open questions [J]. International

- Journal of Game Theory, 2008, 36(3/4): 537-569.
- [12] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2013: 2787-2795.
- [13] CHEN M H, TIAN Y T, YANG M H, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, 2017: 1511-1517.
- [14] LIN Y K, LIU Z Y, LUAN H B, et al. Modeling relation paths for representation learning of knowledge bases[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, 2015: 705-714.
- [15] ZHU H, XIE R B, LIU Z Y, et al. Iterative entity alignment via joint knowledge embeddings[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization, 2017: 4258-4264.
- [16] TAM N T, TRUNG H T, YIN H Z, et al. Entity alignment for knowledge graphs with multi-order convolutional networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(9): 4201-4214.
- [17] WU Y T, LIU X, FENG Y S, et al. Jointly learning entity and relation representations for entity alignment [C] //Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 240-249.
- [18] WU Y T, LIU X, FENG Y S, et al. Relation-aware entity alignment for heterogeneous knowledge graphs [C] //Proceedings of the 28th International Joint Conference on Artificial Intelligence. Macao, China: International Joint Conferences on Artificial Intelligence Organization, 2019: 5278-5284.
- [19] LIU F Y, CHEN M H, ROTH D, et al. Visual pivoting for (unsupervised) entity alignment [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: ACM Press, 2021: 4257-4266.
- [20] MENA G, BELANGER D, LINDERMAN S, et al. Learning latent permutations with gumbel-sinkhorn networks [EB/OL]. [2024-04-02]. <https://arxiv.org/abs/1802.08665>.
- [21] ZHU Y, LIU H Z, WU Z H, et al. Relation-aware neighborhood matching model for entity alignment [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: ACM Press, 2021, 35(5): 4749-4756.
- [22] 陈富强, 寇嘉敏, 苏利敏, 等. 基于图神经网络的多信息优化实体对齐模型[J]. 计算机科学, 2023, 50(3): 34-41.
- CHEN F Q, KOU J M, SU L M, et al. Multi-information optimized entity alignment model based on graph neural network[J]. Computer Science, 2023, 50(3): 34-41. (in Chinese)
- [23] 李泽霖, 吕兆峰, 陈富强, 等. 基于多跳信息融合的实体对齐模型[J]. 计算机工程, 2024, 50(9): 142-152.
- LI Z L, LU Z F, CHEN F Q, et al. Entity alignment model based on multi-hop information fusion [J]. Computer Engineering, 2024, 50(9): 142-152. (in Chinese)
- [24] ZENG W X, ZHAO X, TANG J Y, et al. Collective entity alignment via adaptive features[C]//Proceedings of the IEEE 36th International Conference on Data Engineering (ICDE). Dallas, USA: IEEE Press, 2020: 1870-1873.
- [25] MAO X, WANG W T, WU Y B, et al. From alignment to assignment: frustratingly simple unsupervised entity alignment [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican Republic: Association for Computational Linguistics 2021: 2843-2853.
- [26] MAO X, WANG W T, WU Y B, et al. LightEA: a scalable, robust, and interpretable entity alignment framework via three-view label propagation[C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics 2022: 825-838.
- [27] ZHANG Z S, TAN C X, WANG H H, et al. Type-enhanced ensemble triple representation via triple-aware attention for cross-lingual entity alignment [EB/OL]. [2024-04-02]. <https://arxiv.org/abs/2305.01556>.
- [28] ZHANG Z S, TAN C X, ZHAO X Y, et al. OTIEA: ontology-enhanced triple intrinsic-correlation for cross-lingual entity alignment [EB/OL]. [2024-04-02]. <https://arxiv.org/abs/2305.01561>.
- [29] LIN Z X, ZHANG Z H, WANG M, et al. Multi-modal contrastive representation learning for entity alignment[C]//Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, Republic of Korea: Association for Computational Linguistics 2022: 2572-2584.

文字编辑 金胡考
栏目编辑 赖玉玲