

自适应特征融合的迭代实体对齐方法

李婷婷, 邵 斐, 温天晓, 董 飒

(吉林大学 计算机科学与技术学院, 符号计算与知识工程教育部重点实验室, 长春 130012)

摘要: 针对知识图谱实体对齐任务中缺乏训练数据以及长尾实体对齐准确率较低的问题, 提出一种基于自适应特征融合策略的迭代实体对齐方法, 并设计一种迭代策略自动扩充训练数据的规模. 该方法使用知识图谱的结构信息, 并利用关系、属性和实体名称信息作为语义信息辅助对齐, 从而提升对齐效果. 在数据集上的实验结果表明, 该模型在知识图谱实体对齐任务中效果良好.

关键词: 知识图谱; 实体对齐; 迭代策略; 自适应特征融合

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)03-0629-07

Iterative Entity Alignment Method for Adaptive Feature Fusion

LI Tingting, SHAO Fei, WEN Tianxiao, DONG Sa

(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education,
College of Computer Science and Technology, Jilin University, Changchun 130012, China)

Abstract: Aiming at the problems of insufficient training data and low accuracy of long-tail entity alignment in the task of knowledge graph entity alignment, we proposed an iterative entity alignment method based on an adaptive feature fusion strategy and designed an iterative strategy to automatically expand the scale of the training data. This method utilized the structural information of the knowledge graph and utilized relationships, attributes, and entity name information as semantic information to assist alignment and improve alignment effectiveness. The experimental results on the dataset show that the proposed model performs well in the task of knowledge graph entity alignment.

Keywords: knowledge graph; entity alignment; iterative strategy; adaptive feature fusion

1 引言与预备知识

知识图谱(knowledge graph, KG)的目的是保证搜索引擎的搜索结果足够准确全面, 进而提高搜索引擎的检索能力^[1]. 知识图谱与其前身语义网络^[2]相比, 知识图谱更侧重于描述实体之间的关联. 知识图谱可视为一个有向图, 图中的节点表示实体, 边表示实体间的关系, 知识图谱中的每条知识可表示为一个“主-谓-宾”形式的三元组, 根据谓语的不同形式又可分为关系三元组和属性三元组两种. 知识图谱目前不仅作为搜索引擎的数据库使用, 而且已被广泛应用于知识问答、推理决策、推荐系统、社交网络等其他领域. 不同领域一般都会根据自身的需求有针对性地构建知识图谱, 由于缺乏统一的构建标准, 因此不同知识图谱之间存在异构和冗余问题. 通常两个知识图谱间的信息存在互补的情

收稿日期: 2023-07-12.

第一作者简介: 李婷婷(1986—), 女, 汉族, 硕士, 副研究员, 从事人工智能的研究, E-mail: ttleec@jlu.edu.cn. **通信作者简介:** 董 飒(1985—), 女, 满族, 博士, 高级工程师, 从事数据挖掘和高性能计算的研究, E-mail: dongsa@jlu.edu.com.

基金项目: 吉林省科技发展计划项目(批准号: 20230201083GX).

况, 所以为充分利用知识图谱蕴含的信息, 对不同的知识图谱进行融合生成一个规模更大的知识图谱非常必要, 而知识图谱融合技术中的关键是实体对齐.

知识图谱实体对齐任务的目标是匹配两个不同知识图谱中指代同一客观概念的实体. 目前, 知识图谱实体对齐任务的主流方法是基于图表示学习的实体对齐方法, 其在性能和效率上均远优于传统实体对齐方法. 基于图表示学习的实体对齐方法核心思想是将知识图谱嵌入到低维向量空间中, 通过计算实体向量之间的距离判断实体是否对齐. 这种方法的关键是如何有效地嵌入知识图谱, 目前主要有两类模型用于知识图谱嵌入. 一类是基于 TransE 系列模型的嵌入方法, 代表性的工作有 CTEA^[3], JETEA^[4], ESEA^[5]等, 该类方法的核心思想是将三元组中的关系视为头实体到尾实体的平移向量, 即对于一个三元组 (h, r, t) , 应保证 $h+r \approx t$, 其中 h 表示头实体向量, t 表示尾实体向量, r 表示头实体与尾实体之间的关系向量. 另一类是基于图神经网络系列模型的嵌入方法, 代表性的工作有 RNM^[6], RAC^[7], IPEA^[8]等, 这类方法通过聚合邻居实体信息到中心实体, 从而获得更具表达性的中心实体的嵌入表示.

目前主流的知识图谱实体对齐方法在对长尾实体对齐上的准确率较低, 这是因为目前的实体对齐方法主要依赖知识图谱的结构信息进行对齐, 即使用到其他信息时, 通常也都以固定的权重去融合多种信息, 并且结构信息仍占较大权重, 但长尾实体的结构信息十分匮乏. 此外, 实体对齐任务需要大量预先对齐的实体对充当标记数据, 但这些标记数据的获取并非易事, 人工标注的成本极高. 因此, 本文提出一种基于自适应特征融合策略的迭代实体对齐模型——AFFIEA. 该模型在融合不同方面信息时, 采用一种自适应特征融合策略, 根据实体结构信息的丰富程度动态为不同方面信息分配融合权重, 由于长尾实体的结构信息比较匮乏, 此时将增大其他信息的权重, 该策略可有效提高长尾实体对齐的准确率; 并设计一种迭代策略自动获取标记数据, 很好地解决了标记数据不足的问题.

知识图谱是一种有向图, 其可以被形式化定义为 $KG=(E, R, A, T)$, 其中 E 表示实体集合, R 表示关系集合, A 表示属性集合, T 表示三元组集合. 知识图谱实体对齐任务可以描述为: 给定两个待对齐的知识图谱 $KG_1=(E_1, R_1, A_1, T_1)$ 和 $KG_2=(E_2, R_2, A_2, T_2)$, KG_1 称为源知识图谱, KG_2 称为目标知识图谱, $S=\{(e_1, e_2) \mid e_1 \in KG_1, e_2 \in KG_2, e_1 \equiv e_2\}$ 是事先给定的预对齐实体对的集合, 也称为种子集. 知识图谱实体对齐的目标就是根据种子集 S , 得到分别出现在源知识图谱 KG_1 和目标知识图谱 KG_2 中但表示现实世界中同一概念的实体对的集合, 即最终对齐结果集合 $Align=\{(e_1, e_2) \mid e_1 \in E_1, e_2 \in E_2, e_1 \equiv e_2\}$. 其中 \equiv 表示两个实体间的等价关系, $S \subseteq Align$.

2 自适应特征融合的迭代实体对齐模型

2.1 整体框架

本文提出的 AFFIEA 模型整体架构如图 1 所示. AFFIEA 模型使用图注意力网络 (graph attention network, GAT) 对结构信息进行建模, 通过计算结构嵌入的余弦相似度得到结构相似性矩阵. 对于关系和属性, 本文将它们简单地视为词袋特征, 并通过只包含一个输入层和一个输出层的前馈神经网络获得关系嵌入和属性嵌入. 对于实体名称信息, 本文通过平均化实体名字符串的预训练 Glove 向量得到实体名称特征, 同样也通过上述的前馈神经网络获得实体名称嵌入. 为更好表现实体的语义信息, 本文先将关系嵌入、属性嵌入和实体名称嵌入融合生成一个统一的语义嵌入, 根据语义嵌入生成实体的语义相似性矩阵. 然后使用自适应特征融合策略对结构相似性矩阵和语义相似性矩阵进行融合, 生成实体相似性矩阵, 根据实体相似性矩阵得到实体对齐结果. 最后, 利用设计好的迭代策略从对齐结果中筛选出尽可能正确的对齐实体对加入种子集, 使用更新后的种子集开启下一轮迭代, 重复该过程, 直到新产生的对齐实体对数量低于给定值 S 时结束迭代.

2.2 结构嵌入

图注意力网络 GAT 是一种处理结构化数据的典型神经网络^[9]. 因此, 本文使用 GAT 建模 KG_1 和 KG_2 的结构信息. 对于实体 e_i 的隐藏状态 h_i , 通过聚合其带有自环的一跳邻居 N_i 获得, 计算公式为

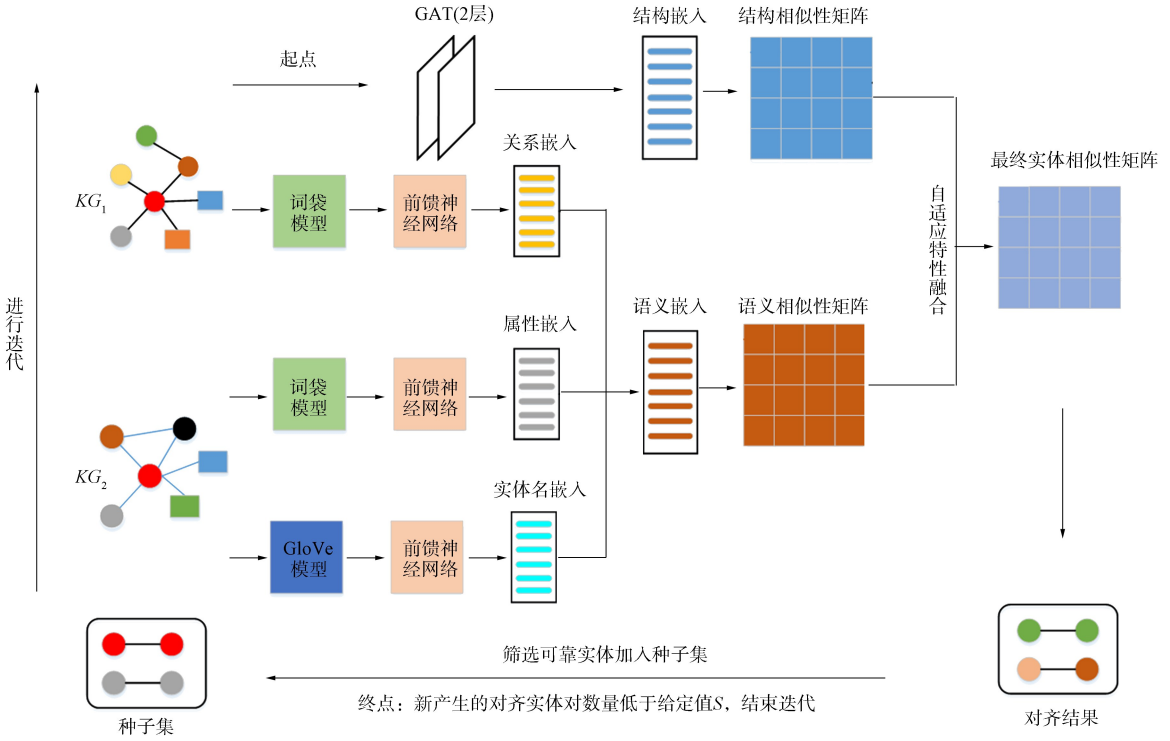


图 1 AFFIEA 模型的整体框架

Fig. 1 Overall framework of AFFIEA model

$$h_i = \text{ReLU}\left(\sum_{j \in N_i} \alpha_{ij} h_j\right), \tag{1}$$

其中: $h_j \in \mathbb{R}^d$ (d 是隐藏层大小) 为实体 e_j 的隐藏状态; α_{ij} 为注意力系数, 表示实体 e_j 对实体 e_i 的重要性, 其通过自注意力机制计算得到:

$$\alpha_{ij} = \frac{\exp\{\text{LeakyReLU}(\mathbf{p}^T(\mathbf{W}h_i \oplus \mathbf{W}h_j))\}}{\sum_{m \in N_i} \exp\{\text{LeakyReLU}(\mathbf{p}^T(\mathbf{W}h_i \oplus \mathbf{W}h_m))\}}, \tag{2}$$

式中 $\mathbf{W} \in \mathbb{R}^{d \times d}$ 为权重矩阵, $\mathbf{p} \in \mathbb{R}^{2d}$ 为可学习的参数, \oplus 表示拼接操作. 为减少计算量增加模型的可扩展性, 本文将 \mathbf{W} 限制为对角矩阵. 为稳定自注意力的学习过程, 本文并行执行 K ($K=2$) 个独立的式(1)过程, 并将这些特征拼接以获得实体 e_i 的结构嵌入表示 h_i^g , 计算公式为

$$h_i^g = \bigoplus_{k=1}^K \text{ReLU}\left(\sum_{j \in N_i} \alpha_{ij}^k h_j\right), \tag{3}$$

其中 α_{ij}^k 为由第 k 个注意力机制计算得到的归一化注意力系数. 在具体实验中, 本文使用一个两层的 GAT 模型聚合多跳内的邻居信息, 并使用最后一个 GAT 层的输出作为邻居结构的嵌入表示.

2.3 关系、属性和实体名称嵌入

本文使用词袋模型显式地对关系和属性进行建模. 设计两个 N -hot 向量 \mathbf{v}_r 和 \mathbf{v}_a 表示关系和属性, N -hot 向量可反映一个实体连接了哪些关系和属性以及它们的数量. 为防止 N -hot 向量的维度过高, 本文在构建词典时只保留出现频率排在前 K 的关系和属性, 故 \mathbf{v}_a 和 \mathbf{v}_r 均为 K 维向量.

因为 GAT 模型在建模关系和属性时, 会自然地引入邻居的相关信息, 产生噪声. 所以本文使用仅包含一个输入层和一个输出层的简单前馈神经网络获取关系和属性的嵌入, 计算公式为

$$h_i^r = \omega_r v_i^r + b_r, \tag{4}$$

$$h_i^a = \omega_a v_i^a + b_a, \tag{5}$$

其中 h_i^r 和 h_i^a 分别表示实体 e_i 的关系嵌入和属性嵌入, v_i^r 和 v_i^a 分别表示实体 e_i 的关系和属性的词袋特征向量, $\omega_r, \omega_a, b_r, b_a$ 为可学习的参数.

对于实体名称, 本文使用预训练的词向量模型 GloVe^[10] 获得实体名称的初始特征向量 \mathbf{v}_n . 同样通

过前馈神经网络获得实体名称嵌入, 计算公式为

$$\mathbf{h}_i^n = \omega_n \mathbf{v}_i^n + b_n. \quad (6)$$

2.4 自适应特征融合

目前主流的实体对齐方法大多数选择在嵌入表示层面去融合多种信息, 即为实体生成一个统一的嵌入表示. 该方法的不足之处是可能无法保持实体的某些原始信息, 两个实体可能在某一特定嵌入下十分接近, 但在统一嵌入下相距很远. 所以, 受文献[11]的启发, 本文选择在相似性矩阵层面对各种信息进行融合. 根据前面得到的结构嵌入, 通过计算余弦相似度的方式得到对齐实体间的结构相似性矩阵 \mathbf{M}_g , 将关系嵌入、属性嵌入和实体名称嵌入以一种简单的加权拼接方式聚合在一起生成语义嵌入, 语义嵌入的计算公式为

$$\mathbf{h}_i^s = \bigoplus_{m \in \mathbf{M}} \left(\frac{e^{\omega_m}}{\sum_{j \in \mathbf{M}} e^{\omega_j}} \mathbf{h}_i^m \right), \quad (7)$$

其中: $\mathbf{M} = (R, A, N)$; ω_m 表示 m 信息的可训练权重, 在加权拼接前对输入的嵌入进行归一化. 同样通过计算语义嵌入间的余弦相似度得到实体间的语义相似性矩阵 \mathbf{M}_s . 然后将 \mathbf{M}_g 和 \mathbf{M}_s 融合生成最终的实体相似性矩阵 \mathbf{M} .

知识图谱中存在大量的长尾实体, 长尾实体通常是指在数据集中出现频率较低的实体, 由于这些实体的数据量较小, 很可能在训练实体对齐模型时未能充分捕捉到它们的特征信息, 导致对其进行准确对齐困难. 此外, 长尾实体常缺乏充分的标注信息, 或者标注信息质量较低, 从而导致在实体对齐过程中对这些实体的特征和语义信息理解不准确, 进而影响对齐准确率. 对于长尾实体, 由于其信息稀疏, 很可能无法获得足够的上下文信息以学习其特征表示, 从而导致在实体对齐任务中难以准确地捕捉到其语义信息. 长尾实体的结构信息十分匮乏, 而结构信息对于对齐实体最重要, 在融合不同信息时结构信息所占权重较大, 且权重固定不变, 从而导致长尾实体的对齐准确率很低. 针对上述问题, 可考虑采用数据增强、更智能的特征表示学习方法、结合上下文信息的实体对齐模型等方法提高长尾实体对齐的准确率. 所以在生成实体相似性矩阵 \mathbf{M} 时, 本文使用一种可根据实体结构信息的丰富程度动态为结构信息和语义信息分配权重的策略. Zeng 等^[12]研究表明, 实体的度数越高, 实体与种子实体的关联程度越大, 此时结构信息越有效. 基于此结论, 本文设计一种权重计算方法, 计算公式为

$$\text{weight}_g = L / (1 + \mathbf{M} \times e^{-N(\text{Degree} + \text{Neighbour})}), \quad (8)$$

$$\text{weight}_s = 1 - \text{weight}_g, \quad (9)$$

其中: L, M, N 为超参数; Degree 为实体的度数; Neighbour 为实体与种子实体的关联度, 可表示为

$$\text{Neighbour} = A_1 \times W_1 + \lg(A_2 \times W_2), \quad (10)$$

A_1 和 A_2 分别为距离种子实体一跳和两跳的实体数, W_1 和 W_2 为超参数.

2.5 迭代策略

为克服训练数据不足的问题, 本文设计一种迭代学习策略自动地从对齐结果中挑选可靠实体对加入种子集. 由于迭代过程不可避免地会引入错误的对齐实体对, 这些错误的实体对在之后的迭代中又会导致产生更多的错误实体对, 出现错误累积情况. 所以迭代策略的关键是如何尽可能地筛选出正确的对齐实体对. 因此, 本文设计了两点策略. 1) 设置距离阈值 θ : 即使一对实体满足互为最近邻的条件, 他们之间的相似度仍可能很低, 所以在此基础上设置一个距离阈值 θ , 只有该对实体之间的距离在低于阈值 θ 时, 才将其视为可靠实体对. 2) 试用期策略: 每经过 m 个 epoch 进行一轮建议, 将满足前两个条件的实体对放入候选集合, 如果该实体对在后续的 n 轮建议中仍满足前两个条件(即试用期), 才将其加入种子集. 每经过 $m \times n$ 个 epoch 更新一次候选集合. 在数据集上验证这两点策略的有效性, 迭代方法的终止条件, 即模型的收敛条件是: 当迭代过程中新生成的对齐实体数低于 S 值时, 终止迭代过程.

3 实验结果与分析

3.1 数据集

本文在数据集 DBP15K 上进行实验. 数据集 DBP15K 是一个被广泛使用的跨语言知识图谱实体

对齐的基准数据集, 它包含来自 DBpedia^[13] 的 4 个特定语言的知识图谱, 相应的存在 3 个版本的跨语言知识图谱实体对齐数据集, 分别是中-英、法-英和日-英版本. 每个版本包含 1.5 万个流行实体间的连接, 通常每种语言涉及到的实体数量远超 1.5 万, 每个数据集中包含大量的三元组, 并且属性三元组在数据集中的占比很高. 本文使用 30% 的实体对作为训练集(种子集), 其余实体对作为测试集. 数据集 DBP15K 的信息列于表 1.

表 1 数据集 DBP15K 的信息

Table 1 Information of DBP15K dataset

数据集	语言	实体数	关系数	属性数	关系三元组	属性三元组	对齐实体数
DBP15K _{中-英}	中	66 469	2 830	8 113	153 929	379 684	15 000
	英	98 125	2 317	7 173	237 674	567 755	
DBP15K _{日-英}	日	65 744	2 043	5 882	164 373	354 619	15 000
	英	95 680	2 096	6 066	233 319	497 230	
DBP15K _{法-英}	法	66 858	1 379	4 547	192 191	528 665	15 000
	英	105 889	2 209	6 422	278 590	576 543	

3.2 评价指标

采用 Hits@ k 和 MRR 作为模型的评价指标, 其中: Hits@ k 表示正确的对齐实体在候选实体中排在前 k 位的比例, Hits@ k 值越大表示模型效果越好, 本文实验中 $k=1, 10$; MRR 表示对齐结果中所有正确对齐实体排名的倒数的平均值, MRR 值越大表示模型的效果越好.

3.3 参数设置

设用于建模结构信息的 GAT 模型的层数为 2 层, GAT 每个隐藏层的维度为 300; 关系、属性和实体名称的嵌入向量的维度均为 100; 自适应特征融合的超参数 $M=1.5$, $N=1$, L 的取值由训练集的规模确定, 超参数 $W_1=0.8$, $W_2=0.1$; 模型的迭代轮数为 1 000, 批大小为 512, 后 500 轮训练中将迭代过程试用期策略中的 m 和 n 值设为 5 和 10, 迭代过程中的实体间距离阈值 $\theta=0.4$, 结束迭代过程的条件值 $S=30$. 本文使用 AdamW^[14] 优化模型, 学习率设为 0.000 5.

3.4 实验结果

选择 BootEA^[15], MRAEA^[16], CEAFF^[17], RNM^[6] 作为本文模型 AFFIEA 的对比模型. BootEA 模型在实体对齐任务中引入了 BootStrapping 的思想, 通过迭代的方式扩充种子集, 并且 BootEA 模型允许对新生成的对齐实体对进行编辑或删除, 从而缓解迭代过程中可能出现的错误累积问题. MRAEA 模型通过为知识图中的关系创建一个逆关系, 从而使知识图中的关系数量加倍, 将节点的关系信息及其邻居信息聚合在一起作为节点特征. CEAFF 模型的核心是利用了强化学习的思想, 在强化学习的框架下, 设计了一致性和排他性约束表征相互依赖并限制集体对齐. RNM 模型利用邻域匹配增强实体对齐, 除在匹配邻域时比较邻居节点外, 模型还从连接关系中挖掘有用的信息, 此外, RNM 模型的迭代过程以半监督的方式利用实体对齐与关系对齐之间的积极交互.

表 2 列出了 AFFIEA 模型与对比模型在 MRR, Hits@1 和 Hits@10 三个指标上的性能. 为模型对比的公平性, BotEA, MRAEA, 和 RNM 三个模型均使用迭代策略, CEAFF 模型中使用一种自适应特征融合策略. 由表 2 可见, 本文模型 AFFIEA 与其他模型相比实验结果最好, 表明了 AFFIEA 模型的有效性. 在与 CEAFF 模型的对比中, AFFIEA 模型在法-英数据集上的性能略高于 CEAFF, 在其余两个数据集上的性能 AFFIEA 模型显著优于 CEAFF 模型, 这得益于 AFFIEA 模型采用的迭代策略, 说明了迭代策略对实体对齐任务的有效性. 在与同为迭代模型的比较中, AFFIEA 模型性能也最好, 说明本文设计的迭代策略相比于其他迭代策略具有优越性, 同时也表明本文的自适应特征融合策略具有重要作用. AFFIEA 模型的性能相比于 BootEA 模型提升 29 个百分点, 相比于 MRAEA 模型提升 16 个百分点, 性能提升明显, 这是因为 BootEA 模型中仅利用了结构信息, 而 MRAEA 模型也只是利用了关系的类型信息, 但 AFFIEA 模型除结构信息外还利用了关系、属性和实体名称信息, 说明额外的辅助信息的确可以增强实体的嵌入表示.

表 2 不同模型的实验结果对比

Table 2 Comparison of experimental results of different models

模型	法-英			日-英			中-英		
	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10	MRR
BootEA	0.653	0.874	0.731	0.622	0.854	0.701	0.629	0.847	0.703
MRAEA	0.780	0.948	0.849	0.758	0.934	0.826	0.757	0.930	0.827
CEAFF	0.933	0.981	0.951	0.803	0.911	0.842	0.736	0.869	0.785
RNM	0.938	0.981	0.954	0.872	0.944	0.872	0.840	0.919	0.870
AFFIEA	0.943	0.985	0.963	0.893	0.967	0.918	0.871	0.953	0.906

3.5 消融实验

本文在数据集 DBP15K 的法-英版本数据集上进行 AFFIEA 模型的消融实验, 结果列于表 3. 由表 3 可见, 当去掉迭代模块时, AFFIEA 模型的 Hits@1 指标下降了 9.1 个百分点, 表明了本文迭代策略的有效性. 自适应特征融合策略也给 AFFIEA 模型带来了 4.4 个百分点的性能提升. 对于不同辅助信息, 当去掉结构信息时, 模型性能下降最多, 下降了 42.3 个百分点, 说明结构信息对实体对齐任务最重要. 关系信息和属性信息分别给 AFFIEA 模型带来了 3.1 个百分点和 4.4 个百分点的性能提升, 实体名称信息为 AFFIEA 模型提供了 13.8 个百分点的性能提升, 说明除结构信息外, 不同信息对实体对齐任务的贡献程度有差距, 实体名称信息要更重要.

表 3 AFFIEA 模型的消融实验结果

Table 3 Results of ablation experiments of AFFIEA model

模型	DBP15K _{法-英}			模型	DBP15K _{法-英}		
	Hits@1	Hits@10	MRR		Hits@1	Hits@10	MRR
AFFIEA-结构	0.520	0.811	0.612	AFFIEA-迭代策略	0.852	0.956	0.882
AFFIEA-关系	0.912	0.976	0.931	AFFIEA-自适应	0.899	0.968	0.926
AFFIEA-属性	0.899	0.969	0.925	特征融合			
AFFIEA-实体名称	0.805	0.948	0.858	AFFIEA	0.943	0.985	0.963

注: “-”表示去掉某一模块.

3.6 自适应特征融合对长尾实体对齐的影响

在消融实验中已验证了自适应特征融合的有效性, 下面进一步研究自适应特征融合对长尾实体对齐的影响. 在数据集 DBP15K_{法-英} 上进行实验, 对比固定特征融合和自适应特征融合两种策略的性能. 根据实体度数由低到高的顺序划分出第 1 组实体、第 2 组实体、第 3 组实体. 图 2 为不同特征融合策略的 Hits@1 对比结果. 由图 2 可见, 在各组实验中, 自适应特征融合的性能均优于固定权重融合, 在第 1~3 组实体中, 二者之间的性能差距分别为 0.047, 0.035, 0.031, 在第 1 组实体中二者之间的性能差距最大, 说明实体结构信息越匮乏, 自适应特征融合策略的性能越好, 从而验证了自适应特征融合策略可提高长尾实体对齐的准确率.

综上所述, 针对知识图谱实体对齐任务中缺乏训练数据以及长尾实体对齐准确率较低的问题, 本文提出了一个基于自适应特征融合策略的迭代实体对齐模型——AFFIEA. 该模型除利用实体的结构信息外, 还利用实体的关系、属性和实体名称作为辅助信息增强实体的嵌入表示. AFFIEA 模型利用迭代的方式自动扩充种子集的规模, 通过设计互为最近邻、距离阈值 θ 和试用期 3 个策略保证迭代过程中尽可能地筛选出正确的对齐实体对, 迭代策略克服了实体对齐任务缺少训练数据的问题. 为提高长尾实体对齐的准确率, 使用了一种可根据实体结构信息的丰富程度动态为结构信息和语义信息分

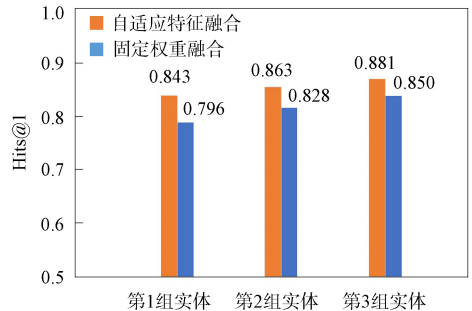


图 2 不同特征融合策略的 Hits@1 对比结果

Fig. 2 Comparison results of Hits@1 for different feature fusion strategies

配聚合权重的自适应特征融合方法. 将 AFFIEA 模型在数据集 DBP15K 上进行实验, 与其他基线模型相比实验结果最好, AFFIEA 模型的消融实验结果验证了模型各模块的有效性. 此外, 本文还研究了自适应特征融合对长尾实体对齐的影响, 实验结果证实了本文策略的有效性.

参 考 文 献

- [1] LIU Q, LI Y, DUAN H, et al. Knowledge Graph Construction Techniques [J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [2] BERNERS-LEE T, HANDLER J, LASSILA O. The Semantic Web [J]. Scientific American, 2003, 284(5): 34-43.
- [3] YAN Z H, PENG R, WANG Y Q, et al. CTEA: Context and Topic Enhanced Entity Alignment for Knowledge Graphs [J]. Neurocomputing, 2020, 410(3): 419-431.
- [4] SONG X, ZHANG H, BAI L. Entity Alignment between Knowledge Graphs Using Entity Type Matching [C]// Knowledge Science, Engineering and Management. Berlin: Springer, 2021: 578-589.
- [5] JIANG T T, BU C Y, ZHU Y, et al. Combining Embedding-Based and Symbol-Based Methods for Entity Alignment [J]. Pattern Recognition, 2022, 124: 108433-1-108433-14.
- [6] ZHU Y, LIU H Z, WU Z H, et al. Relation-Aware Neighborhood Matching Model for Entity Alignment [C]// National Conference on Artificial Intelligence. Palo Alto: AAAI, 2021: 4749-4756.
- [7] ZENG W X, ZHAO X, TANG J Y, et al. Reinforced Active Entity Alignment [C]//Proceedings of the 30th ACM International Conference on Information & Knowledge Management. New York: ACM, 2021: 2477-2486.
- [8] YANG L Y, LÜ C, WANG X, et al. Collective Entity Alignment for Knowledge Fusion of Power Grid Dispatching Knowledge Graphs [J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(11): 1990-2004.
- [9] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks [EB/OL]. (2017-10-30) [2023-04-11]. <https://arxiv.org/abs/1710.10903>.
- [10] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global Vectors for Word Representation [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). [S. l.]: ACL, 2014: 1532-1543.
- [11] ZENG W X, ZHAO X, TANG J Y, et al. Reinforcement Learning-Based Collective Entity Alignment with Adaptive Features [J]. ACM Transactions on Information Systems (TOIS), 2021, 39(3): 1-31.
- [12] ZENG W X, ZHAO X, WANG W, et al. Degree-Aware Alignment for Entities in Tail [C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 811-820.
- [13] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia: A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia [J]. Semantic Web, 2015(6): 1-29.
- [14] LOSHCHILOV I, HUTTER F. Decoupled Weight Decay Regularization [EB/OL]. (2017-11-14)[2023-02-10]. <https://arxiv.org/abs/1711.05101>.
- [15] SUN Z Q, HU W, ZHANG Q H, et al. Bootstrapping Entity Alignment with Knowledge Graph Embedding [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. New York: ACM, 2018: 4396-4402.
- [16] MAO X, WANG W T, XU H M, et al. MRAEA: An Efficient and Robust Entity Alignment Approach for Cross-Lingual Knowledge Graph [C]//Proceedings of the 13th International Conference on Web Search and Data Mining. New York: ACM, 2020: 420-428.
- [17] ZENG W X, ZHAO X, TANG J Y, et al. Reinforcement Learning Based Collective Entity Alignment with Adaptive Features [J]. ACM Transactions on Information Systems, 2021, 39(3): 26-1-26-31.

(责任编辑: 韩 啸)