

# 基于实体可靠路径与语义增强的知识图谱对齐

王红斌<sup>1,2</sup>, 唐浩东<sup>1,2</sup>, 线岩团<sup>1,2</sup>, 刘博<sup>3</sup>, 顾新亮<sup>3</sup>

(1. 昆明理工大学 信息工程与自动化学院 昆明 650500; 2. 昆明理工大学 云南省人工智能重点实验室 昆明 650500; 3. 富奥汽车零部件股份有限公司, 长春, 130012)

**摘要:** 针对知识图谱实体之间存在大量的多步骤关系路径以表明实体之间的语义关系, 现有的方法很少考虑路径信息, 也忽略了知识图谱关系结构与属性结构之间的邻域异质性问题。本文提出了实体可靠路径信息语义增强模型, 该模型集成了知识图谱中的关系和路径结构信息以及异构信息, 设计了一种初始可靠路径推理算法, 用于捕获和聚合对齐实体及其异构邻居的多源信息, 并将实体的关系结构、属性结构、实体名称路径集信息通过模型聚合进行语义增强, 由此解决知识图谱对齐邻域异质性问题。本文在 3 个数据集 (WK31-15K、DBP-15K 和 DWY-100K) 上评估了实体可靠路径信息语义增强模型, 将该模型与其他较先进的实体对齐方法进行比较, Hits@1 提升了 1.5%~3.2%, 表明本文方法具有更好的性能。

**关键词:** 知识图谱; 实体对齐; 可靠路径信息; 属性结构; 语义增强

**中图分类号:** TP391.1 **文献标志码:** A **文章编号:** 1671-5497(2025)11-3673-13

**DOI:** 10.13229/j.cnki.jdxbgxb.20240177

## Knowledge graph alignment based on entity reliable path and semantic aggregates

WANG Hong-bin<sup>1,2</sup>, TANG Hao-dong<sup>1,2</sup>, XIAN Yan-tuan<sup>1,2</sup>, LIU Bo<sup>3</sup>, GU Xin-liang<sup>3</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Key Laboratory of Artificial Intelligence in Yunnan Province, Kunming University of Science and Technology, Kunming 650500, China; 3. Fawer Automotive Parts Limited Company, Changchun 130012, China)

**Abstract:** There are numerous multi-step relationship paths between entities in the knowledge graph to indicate semantic relationships between entities, as well as the neighborhood heterogeneity between relationship structures and attribute structures. In response to this problem, An entity reliable path information semantic augmentation model is proposed in this paper, which simultaneously captures and aggregates multi-source information of aligned entities and their heterogeneous neighbors, an initial reliable path reasoning algorithm to generate. The model aggregates the relationship structure, attribute structure, and entity name information reliable path of the entities for semantic augmentation, which solves the

收稿日期: 2024-02-23.

基金项目: 云南省重点研发计划项目(202202AD080003).

作者简介: 王红斌(1983-), 男, 教授, 博士. 研究方向: 自然语言处理, 信息检索和机器学习.

E-mail: whbin2007@126.com

通信作者: 线岩团(1981-), 男, 副教授, 博士. 研究方向: 自然语言处理, 信息检索和机器学习.

E-mail: 195426286@qq.com

problem of domain heterogeneity in knowledge graph alignment. The paper evaluated the entity reliable path information semantic augmentation model on three datasets (WK31-15K, DBP-15K and DWY-100K) show that this model is improved by 1.5%~3.2% compared with the state-of-art entity alignment method Hits@1, which shows that the proposed method has better performance.

**Key words:** knowledge graphs; entity alignment; reliable path information; attribute structures; semantic augmentation

## 0 引言

知识图谱以结构化信息表示现实世界中的事实<sup>[1]</sup>,针对不同的需求构建不同知识图谱(例如DBpedia<sup>[2]</sup>、YAGO<sup>[3]</sup>、Freebase<sup>[4]</sup>、Probase<sup>[5]</sup>),广泛应用于智能检索、推荐与问答<sup>[6-8]</sup>、机器翻译等邻域<sup>[9-12]</sup>。然而这些知识图谱仍然远远不够完整。实体对齐(也称为实体匹配<sup>[13-15]</sup>)旨在识别不同知识图谱中的相同实体。传统的实体对齐方法<sup>[16]</sup>利用字符串相似度或者规则挖掘等方法<sup>[17]</sup>实现对齐。目前方法使用图神经网络(GNN)(GCN<sup>[18]</sup>和GAT<sup>[19]</sup>)通过嵌入实体并聚合相邻特征对齐<sup>[15]</sup>。

但是目前方法仍存在局限性。一是基于GNN的模型(HGCN<sup>[20]</sup>、RDGCN<sup>[21]</sup>、NMN<sup>[22]</sup>)没有考虑知识图谱的属性结构,而常见的知识图谱有关系结构和属性结构。二是现有方法很少考虑知识图谱的路径信息(即对齐实体的间接邻域)。本文通过路径信息挖掘对齐实体之间更多相似特征,使其单独作为一个模块进行嵌入。三是现有方法忽略了知识图谱的异质性。不同知识图谱中相同实体的邻域信息差距较大,被称为知识图谱邻域异质性。知识图谱之间的异质性会影响对齐性能。

本文提出了实体可靠路径信息语义增强模型(EPsA),该模型将实体的三元组分为关系三元组与属性三元组,挖掘实体路径信息提高对齐判断的准确性和鲁棒性;同时通过知识图谱语义增强模型强化并聚合实体名称路径信息、关系结构、属性结构信息解决实体异构邻域中的异质性问题。本文主要贡献总结如下:

(1)提出了实体可靠路径信息语义增强模型(EPsA),该模型考虑了知识图谱实体的属性结构与路径信息(即对齐实体的间接邻域)。通过实体的路径信息作为第三个维度嵌入聚合,有助于提高对齐判断的准确性和鲁棒性。

(2)设计了一种聚合多方异构信息的知识图谱语义增强模型,该模型将实体的名称路径信息、关系结构、属性结构三方面信息通过模型聚合进行语义增强,强化实体各结构之间的联系解决实体异构邻域中的异质性问题。

(3)本文方法在3个数据集(WK31-15K、DBP-15K和DWY-100K)上进行实验,实验结果表明EPsA显著优于目前9种最先进的方法。

## 1 相关工作

### 1.1 基于翻译模型

基于翻译的模型TransE<sup>[23]</sup>将实体映射到低维向量以获得实体语义,MTTransE<sup>[24]</sup>使用两个TransE模型分别训练两个知识图谱的向量空间,通过实体训练线性变换对齐这两个向量空间。

在基于翻译模型的基础上,有些方法(例如IPTransE<sup>[25]</sup>、BootEA<sup>[26]</sup>、AHAB<sup>[27]</sup>)开始专注于迭代发现更多新的对齐实体。IPTransE<sup>[25]</sup>采用迭代方法识别不同知识图谱的相同实体。BootEA<sup>[26]</sup>通过寻求所有标记和未标记实体的最大对齐可能性。尽管如此,迭代过程产生的不正确对齐实体会影响后续迭代的准确性。

### 1.2 基于GNNs的实体对齐

基于GNN的方法主要使用GCN或GAT获得两个知识图谱之间的相似结构信息,例如RDGCN<sup>[21]</sup>、NMN<sup>[22]</sup>、MuGNN<sup>[28]</sup>)。其中基于GAT的模型(NAEA<sup>[29]</sup>、KECG<sup>[30]</sup>)采用自注意力机制聚合实体的相邻特征。基于GCN的模型(GCN Align<sup>[31]</sup>、HMAN<sup>[32]</sup>、VR-GCN<sup>[33]</sup>、GMNN<sup>[34]</sup>)使用卷积算子来聚合实体邻居的嵌入,并最终使对齐实体对的嵌入彼此接近。MuGNN<sup>[28]</sup>将邻域结构图与实体嵌入相融合,但它没有明确地利用关系结构的邻域信息。MultiKE<sup>[35]</sup>模型基于多视图学习方法将实体描述视图映射到不同的向量空间。然而,现有方法大多不考虑知识图谱的属性结构,只考虑关系结构,没有考

考虑实体的邻域路径信息,实体路径信息有助于提高对齐判断的准确性和鲁棒性。

### 1.3 考虑属性结构的实体对齐方法

为了获得更多实体的邻域信息,已有方法开始考虑实体的属性结构。例如,图1中蓝色方框表示关系三元组结构,绿色方框表示属性三元组结构。JAPE<sup>[36]</sup>是第一个同时考虑关系和属性结构的模型,后来 AttrE<sup>[37]</sup>、MultiKE<sup>[35]</sup>、GCN Align<sup>[31]</sup>、HMAN<sup>[32]</sup>、AttrGNN<sup>[38]</sup>、COTSAE<sup>[39]</sup>和 JarKA<sup>[40]</sup>也同时考虑了关系和属性结构。KD-CoE<sup>[25]</sup>和 HMAN<sup>[32]</sup>考虑了实体描述信息,IP-TransE<sup>[31]</sup>和 HMAN<sup>[30]</sup>考虑了实体的拓扑特征。RDGCN<sup>[31]</sup>、HGCN<sup>[30]</sup>、MultiKE<sup>[35]</sup>和 AttrGNN<sup>[33]</sup>考虑了实体名称信息作为实体的输入特征。但上述方法未考虑知识图谱异质性。

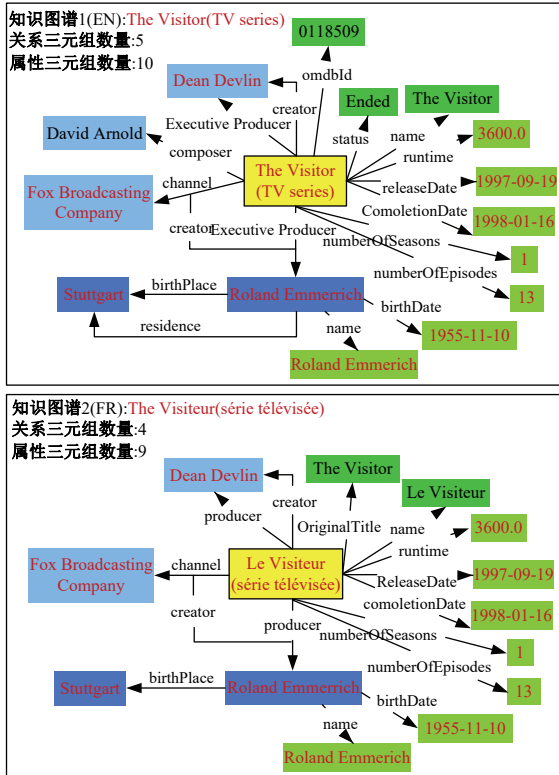


图1 不同知识图谱的相同实体对比

Fig. 1 Comparison of identical entities in different knowledge graphs

## 2 问题定义

关于符号的使用。本文将一个知识图谱表示为6个部分  $KG=(E, P, R, M, V, X, Y)$ , 其中  $E, P, R, M, V$  分别表示实体、实体路径信息、关系、属性与属性值。 $X \subseteq E \times R \times E$  表示关系三元组的集合; $Y \subseteq E \times M \times V$  表示属性三元组的

集合; $T_{path} \subseteq E \times P \times E$  表示路径三元组的集合。 $KG_1=(E_1, P_1, R_1, M_1, V_1, X_1, Y_1)$ 、 $KG_2=(E_2, P_2, R_2, M_2, V_2, X_2, Y_2)$  表示要对齐的两个知识图谱。本文将  $KG_1$  与  $KG_2$  合并成一个原始知识图谱  $KG_a=(E_a, P_a, R_a, M_a, V_a, X_a, Y_a)$ , 其中,  $E_a=E_1 \cup E_2$ , 其他相关属性同理。最终目标是找到一组对齐的实体  $L=\{(ei, ej) \in E_1 \times E_2 | ei \equiv ej\}$ , 其中  $\equiv$  表示为实体对齐。本文使用粗体小写字母表示嵌入向量, 使用粗体大写字母表示嵌入矩阵, 表1列出了具体符号及其说明。

表1 符号与说明

Table 1 Notations and descriptions

| 符号            | 说明                   |
|---------------|----------------------|
| $X^{e\_init}$ | 实体名称嵌入初始矩阵           |
| $X^{pt}$      | 实体路径集输出嵌入矩阵          |
| $X^{v\_init}$ | 属性值嵌入初始矩阵            |
| $X^{at}$      | 基于注意力感知属性三元组的输出嵌入矩阵  |
| $R^d$         | $d$ 维矩阵空间            |
| $L$           | Linear 连接函数          |
| $X^e$         | 实体名称嵌入矩阵             |
| $X^r$         | 关系嵌入矩阵               |
| $X^v$         | 属性值嵌入矩阵              |
| $X^{e_a}$     | 关系三元组与属性三元组语义增强输入矩阵  |
| $\odot$       | 乘法运算                 |
| $\sigma$      | ReLU 函数              |
| $X^p$         | 实体路径嵌入矩阵             |
| $X^{rt}$      | 基于注意力感知关系三元组的输出嵌入矩阵  |
| $X^m$         | 属性嵌入矩阵               |
| $X^{e_{out}}$ | 基于多头注意力感知聚合的语义增强输出矩阵 |
| $//$          | 矢量连接                 |
| $\eta$        | LeakyReLU 非线性函数      |

## 3 本文方法

本文提出了一种用于实体对齐的实体可靠路径信息语义增强模型(EPsA), 该模型挖掘了不同知识图谱实体的邻域路径信息, 考虑了知识图谱实体的路径信息, 将实体的关系结构、属性结构、实体名称路径信息三方面通过模型聚合进行语义增强。该模型既提高了对齐判断的准确性和鲁棒性, 又通过增强实体各结构之间的联系解决了实体异构邻域中的异质性问题。图2显示了本文EPsA模型的总体架构, 本文模型分为6个部分: (1) 实体嵌入。将知识图谱的实体名称、实体路径集、属性值进行预对齐嵌入, 将嵌入的实体信息分为实体名称嵌入、实体路径嵌入、关系三元组嵌入与属性三元组嵌入模块。(2) 实体路径匹配嵌入。分别将实体的名称与路径集嵌入并进行实体路径匹配, 获得对齐实体相应路径信息。(3) 关

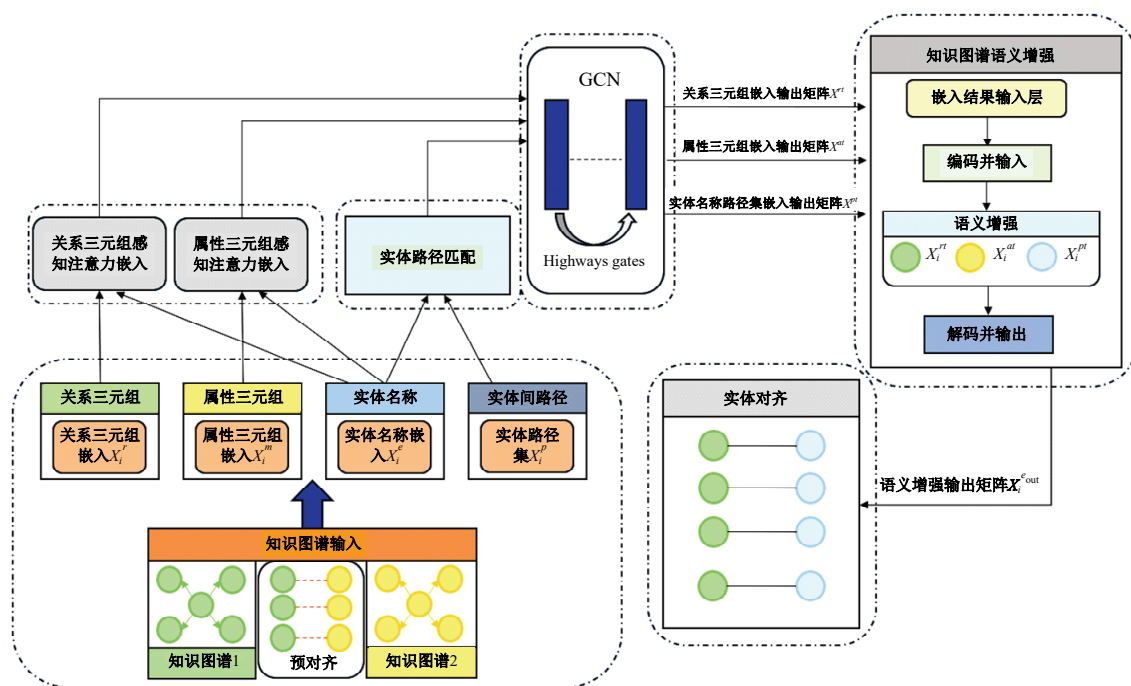


图 2 EPSA 的总体架构

Fig. 2 Overall architecture of EPSA model

系三元组与属性三元组感知注意力嵌入。本模块使用一种变体注意力机制模型,分别在关系结构与属性结构模块上聚合实体的异构邻域特征。(4) GCN 聚合模块。使用搭配 Highway gates 的双层 GCN 模型进一步聚合异构邻域的结构信息,并分别输出实体的路径结构、关系结构、属性结构的嵌入结果。(5) 知识图谱语义增强。本模块将路径结构、关系结构、属性结构的两个嵌入结果进行语义增强,强化实体各部分之间的联系并聚合其输出结果。(6) 实体对齐与训练。通过计算两个实体之间语义增强输出矩阵的曼哈顿距离实现实体对齐。训练的目的在于使类似的多语言实体对尽可能对齐,并且通过替换头尾实体产生负三元组,通过训练增大负三元组与正确对齐三元组的曼哈顿距离。本文代码与数据集已公开至 Github <https://github.com/tanghaodong/EPSA.git>。

### 3.1 实体嵌入

实体的名称都是由单词或字符序列组成,所以我们将使用 Skip-Gram<sup>[42]</sup> 中的文字嵌入预嵌入实体与属性值。

$$EM(w) = \begin{cases} word\_embed(w), & w \text{ 具有词嵌入} \\ char\_embed(w), & \text{其他} \end{cases} \quad (1)$$

式中:  $EM(w)$  为已获得单词嵌入的嵌入函数;  $word\_embed(\cdot)$  为预训练的单词嵌入;  $char\_embed(\cdot)$  为预训练的字符嵌入的平均向量。

$embed(\cdot)$  为预训练的字符嵌入的平均向量。

设定  $H_{e_i} = (w_1, w_2, \dots, w_n)$  表示实体名称的集合,  $e_i \in \epsilon_a$ , 构建实体名称嵌入  $x_i^e \in X^{e, init} \in R^{|\epsilon_a| \times d}$  如下所示:

$$x_i^e = com \left( \sum_{w_j \in H_{e_i}} EM(w_j) \right) \quad (2)$$

式中:  $com(\cdot)$  为 SUM 组成函数, 同样属性值嵌入  $x_i^v \in X^{v, init} \in R^{|\epsilon_v| \times d}$  也使用这种方法进行构造。

在实体嵌入后将嵌入的模块信息分为实体名称嵌入、实体路径嵌入、关系三元组嵌入和属性三元组嵌入 4 个模块。将实体名称嵌入  $x^e$  和实体路径嵌入  $x^p$  放入实体路径匹配模块, 实体名称嵌入  $x^e$  和关系三元组放入关系模块, 将实体名称嵌入  $x^e$ 、属性值嵌入  $x^v$  和属性三元组放入属性模块。

### 3.2 实体路径匹配嵌入

通过实体嵌入模块获得实体名称与实体路径集嵌入。由于已知预对齐实体的小范围邻域之间匹配次数高的路径可以被认为是可靠的, 其可以用来匹配其他实体大范围路径的判断<sup>[37]</sup>。图 3 为实体路径匹配架构图。

通过建立每个预对齐实体对之间的路径邻域匹配, 推导出匹配路径, 最后选择匹配次数较多的路径形成可靠路径集。图 4 为预对齐实体之间的路径邻域匹配示意图。

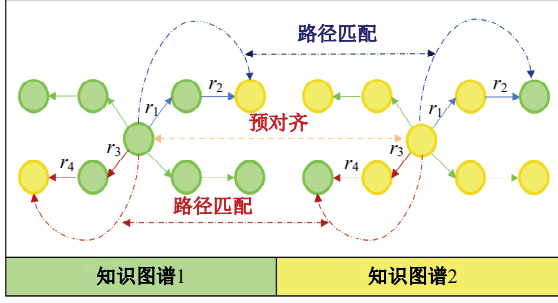


图 3 实体路径匹配架构图

Fig. 3 Overall architecture of entity path matching model

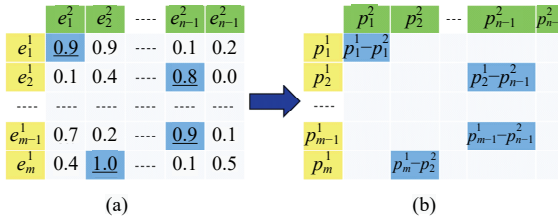


图 4 预对齐实体之间的路径邻域匹配示意图

Fig. 4 Illustration of the path neighborhood matching between the pre-alignment entity

首先,获得预对齐实体对  $(e_a^1, e_a^2)$ , 相似矩阵  $S$  表示路径邻域  $PN(e_a^1)$  与  $PN(e_a^2)$  之间的相似度, 其中  $PN(\cdot)$  表示实体的路径邻域。如图 4(a) 所示, 先选取  $S$  的每一行相似度最大的实体作为匹配邻居, 例如  $e_1^1$  的相似度最高的匹配邻居为  $e_2^2$ , 但是可能存在多个相互匹配的实体, 例如  $e_1^1$  和  $e_{m-1}^1$  与  $e_{n-1}^2$  同时匹配, 因此邻居匹配需要一对一的约束条件:

(1) 实体匹配邻居的相似度必须达到一定的阈值;

(2) 将满足阈值的相似度值由高到低排序, 进行一对一匹配。由此可得, 如果  $e_2^1$  与  $e_{n-1}^2$  相似度比  $e_{m-1}^1$  与  $e_{n-1}^2$  相似度低, 那么选择  $e_{m-1}^1$  与  $e_{n-1}^2$  相互匹配。显然, 并不是所有的实体都能找到匹配的邻居。

其次, 根据匹配的实体对  $(e_i^1, e_j^2)$ , 对其实体路径进行匹配得到  $(p_k^1, p_l^2) \in X^p$ , 如图 4(b) 所示:

$$e_i^1 \leftrightarrow e_j^2 \Rightarrow (e_a^1, p_k^1, e_i^1) \leftrightarrow (e_a^2, p_l^2, e_j^2) \Rightarrow p_k^1 \leftrightarrow p_l^2 \quad (3)$$

式中:  $\leftrightarrow$  为匹配关系;  $(e_a^1, p_k^1, e_i^1) \leftrightarrow (e_a^2, p_l^2, e_j^2)$  为路径三元组。

最后, 统计每条匹配路径的匹配次数,  $p_k^1 \leftrightarrow p_l^2$  选择匹配次数较多的路径形成实体路径集输出嵌入矩阵  $X^{pt}$ 。

### 3.3 关系三元组感知注意力嵌入

关系三元组感知注意力模块。  $X^r \in R^{|R_s| \times 2d}$  表示关系三元组的嵌入矩阵,  $t$  表示第  $t$  个实体, 实体的关系  $r_t$  的嵌入通过从关系三元组中平均其头部实体  $H_t$  和尾部实体  $T_t$  的嵌入近似, 如下所示:

$$X_t^r = \sigma \left[ \frac{\sum_{e_i \in H_t} b_i^1 x_i^e}{|H_t|} \parallel \frac{\sum_{e_j \in T_t} b_j^2 x_j^e}{|T_t|} \right] \quad (4)$$

式中:  $x_i^e, x_j^e \in X^{e, \text{init}}$  为实体的名称嵌入  $e^i$  与  $e^j$ ;  $|H_t|$  与  $|T_t|$  分别为关系三元组中头实体与尾实体的集合大小;  $b^1 \in R^{|e|}$  为头实体的权重向量参数, 是为了分辨不同的头实体对关系  $r_t$  的不同贡献。同理  $b^2$  表示尾实体的权重向量参数。

本文使用了感知注意力机制 GATs 学习与关联实体中具有不同关系的邻居的权重, 将邻域间的关系添加到了注意力系数的计算中, 公式如下:

$$e_{ij}^r = \text{GATs}_{\text{relation}}(e_i, e_j; G_{ij}^r) \quad (5)$$

式中:  $\text{GATs}_{\text{relation}}(\cdot)$  为关系感知注意力模块;  $G_{ij}^r \in \gamma_a$  为头实体  $e^i$  与尾实体  $e^j$  的关系三元组集合  $\gamma_a$ 。

如式(5)所示, 注意力系数不但取决于实体的特征, 还取决于实体之间关系的特征。由于不同知识图谱之间同一实体的邻域不对称性, 从实体  $e^i$  到  $e^j$  与实体  $e^j$  到  $e^i$  的关系会有很大的区别, 这是知识图谱异质性的一个关键特性。所以通过注意力系数  $e_{ij}^r$  通过 soft max 的函数进行细化和归一化:

$$a_{ij}^r = \psi \left\{ \eta \left( \sum_{(e_i, r_i, e_j) \in G_{ij}^r} e_{ij}^r ([x_i^e \parallel x_j^e] \odot X_t^r) \right) \right\} \quad (6)$$

$a_{ij}^r$  为 soft max 系数, 通过 soft max 系数  $a_{ij}^r$  的实体邻域之间不同实体邻居的不同重要性聚合邻域属性。结合 soft max 系数  $a_{ij}^r$ , 可以得到结合邻域实体  $e_i$  的关系邻域结构矩阵:

$$X_i^{rt} = \sigma \left( \sum_{e_j \in N_i^r} a_{ij}^r x_j^e \right) \quad (7)$$

本文通过混合实体的名称嵌入和关系感知注意力嵌入  $X^{rt}$  获得最终输出, 其中  $X^{rt}$  为关系邻域结构矩阵  $X_i^{rt}$  集合,  $\beta_1$  为最终结果中平衡  $X^{rt}$  与  $X^{e, \text{init}}$  的权衡参数。

$$X^{rt} = \beta_1 X^{rt} + X^{e, \text{init}} \quad (8)$$

### 3.4 属性三元组感知注意力嵌入

属性三元组感知注意力嵌入模块与关系三元

组的相似。 $X^m \in R^{|\mathcal{M}_s| \times 2d}$  表示属性三元组的嵌入矩阵,实体的属性  $x_g$  的嵌入  $X_g^m$  通过从关系三元组中近似平均其有关联的头部实体  $H_g$  和尾部属性值  $V_g$  的嵌入构建属性三元组嵌入  $X_g^m$ ,如下所示:

$$X_g^m = \sigma \left[ \frac{\sum_{e_h \in H_g} c_h^1 x_h^e}{|H_g|} // \frac{\sum_{v_k \in V_g} c_k^2 x_k^v}{|V_g|} \right] \quad (9)$$

式中: $x_h^e$  和  $x_k^v$  分别为头实体  $e^n$  与尾属性值  $v^k$ ;  $|H_g|$  与  $|V_g|$  分别为属性三元组中头实体与尾属性值的集合大小; $c^1$  与  $c^2$  分别为头实体与尾属性值的权重向量参数。

使用属性感知注意力机制 GATs 学习与关联实体中具有不同关系的邻居权重,公式如下:

$$v_{pq}^m = \text{GATs}_{\text{attribute}}(e_p, v_q; G_{pq}^m) \quad (10)$$

式中: $\text{GATs}_{\text{attribute}}(\bullet)$  为属性感知注意力模块; $G_{pq}^m \in M_a$  为头实体  $e^p$  与尾属性值  $v^q$  的属性三元组集合。注意力系数  $v_{pq}^m$  通过 soft max 的函数进行细化和归一化:

$$a_{pq}^m = \psi \left\{ \eta \left( \sum_{(e_p, m_g, v_q) \in G_{pq}^m} v_{pq}^m ([x_p^e || x_q^v] \odot X_g^m) \right) \right\} \quad (11)$$

$a_{pq}^m$  为 soft max 系数,通过 soft max 系数  $a_{pq}^m$  的实体邻域之间不同实体邻居的不同重要性聚合邻域属性。结合 soft max 系数  $a_{pq}^m$ ,可以得到结合邻域属性  $v_p$  的属性结构邻域矩阵:

$$X_p^{at} = \sigma \left( \sum_{v_q \in N_p^m} a_{pq}^m X_p^{rt} \right) \quad (12)$$

本文通过混合实体的名称嵌入和属性感知注意力嵌入  $X^{at}$  获得最终输出,其中  $X^{at}$  为属性结构邻域矩阵  $X_p^{at}$  集合,  $\beta_2$  为最终结果中平衡  $X^{at}$  与  $X^{e.\text{init}}$  的权衡参数。

$$X^{at} = \beta_2 X^{at} + X^{e.\text{init}} \quad (13)$$

### 3.5 GCN 聚合

本文得到了分别包含两个知识图谱的实体路径信息、关系结构信息、属性结构信息的实体嵌入  $X^{pt}$ 、 $X^{rt}$ 、 $X^{at}$ 。分别将 3 种结构的嵌入结果使用添加 Highway gates 的双层 GCN 模型聚合对应的异构邻域结构信息。而添加 Highway gates 的目的是控制多层神经网络中传输到节点的邻域信息的噪声<sup>[43]</sup>。

$$X^{(f,l)} = \sigma(\tilde{D}\tilde{A}\tilde{D}X^{(f,l-1)}W^{(f,l)}) \quad (14)$$

$$T(X^{(f,l)}) = \psi(X^{(f,l)}W_T^{(f,l)} + b_T^{(f,l)}) \quad (15)$$

$$X^{(f,l+1)} = T(X^{(f,l)}) \odot X^{(f,l)} + (1 - T(X^{(f,l)})) \odot X^{(f,l-1)} \quad (16)$$

式中: $f = \{pt, rt, at\}$  表示各模块嵌入; $\tilde{A} = A + I$  为关系结构的邻接矩阵,  $I$  为单位矩阵; $\tilde{D}$  为  $\tilde{A}$  的对角度数矩阵; $W^{(f,l)}$  为 GCN 中表示第  $l$  层的相关参数; $X^{(f,l)}$  为第  $l$  层的输出; $W_T^{(f,l)}$  与  $b_T^{(f,l)}$  分别为权重矩阵和偏差向量。将实体路径结构、关系结构、属性结构的结果分别输入 GCN 中聚合不同特征以便进行语义增强时更好地聚合三者之间的联系与特点。

### 3.6 知识图谱语义增强

本文设计了一种新的实体对齐模型,称为知识图谱语义增强模型。该模型通过结合实体各异构信息并通过语义增强强化实体各结构之间的联系以解决知识图谱的异构邻域问题。首先将实体路径结构、关系结构、属性结构的输出结果  $X^{pt}$ 、 $X^{rt}$ 、 $X^{at}$  使用 MFSA<sup>[44]</sup> 中的模型编码器方法将其嵌入并进行上下文编码,关系结构为:

$$H_i^{rt} = [\text{LSTM}_{f(X_i^{rt})}, \text{LSTM}_{b(X_i^{rt})}] \quad (17)$$

关系结构的输出通过两层 BiLSTM 捕获来自不同方向的邻域信息进行编码形成最终的上下文表示。BiLSTM 模型可以自动有效地从输入序列中学习知识图谱邻居特征。使用 BiLSTM 模型编码可以为后文语义增强并聚合提供便利。路径结构与属性结构的输出  $X^{pt}$ 、 $X^{at}$  也用同样的方法进行编码。

EPSA 为每个实体构建更小的子图,将实体各结构的编码结果  $H_i^{pt}$ 、 $H_i^{rt}$ 、 $H_i^{at}$  放入矩阵空间构建更小的子图,从而通过语义增强强化实体各结构之间的联系实现更精细的实体对齐。以下是语义增强的公式:

$$w_i = \frac{\exp(\cos(H_i^{rt}, H_i^{at}))}{\sum_k \exp(\cos(H_k^{rt}, H_k^{at}))} \quad (18)$$

$$X_i^{e_n} = \sigma\{L(w_i[H_i^{rt} // H_i^{at}])\} \quad (19)$$

式中: $w_i$  为语义增强权重参数,用来平衡关系结构和属性结构的编码结果之间的差异。得到实体路径结构、关系结构、属性结构的编码结果  $H_i^{pt}$ 、 $H_i^{rt}$ 、 $H_i^{at}$  之间的语义增强权重参数后,将其放入矩阵空间构建更小的子图,然后通过 Linear 连接函数进行连接聚合并使用 ReLU 函数激活。

使用 MFSA<sup>[44]</sup> 中的解码器方法将语义增强后的结果进行解码并输出  $X_i^{e_{\text{out}}}$ 。

### 3.7 实体对齐与训练

本文通过知识图谱语义增强模块聚合得到结果  $X_i^{e_{out}}$  后,使用曼哈顿距离  $L_1$  测量候选实体对的相似性。曼哈顿距离越小,表示候选实体对对齐的概率越高。为了捕捉实体各个方面的特征,先前的方法(例如 HGCN<sup>[20]</sup>、HMAN<sup>[32]</sup> 和 NMN<sup>[22]</sup>)通常将实体的多个嵌入连接起来,然后在损失函数中用连接的嵌入。本文对比不同知识图谱的两个实体之间最终输出结果之间的曼哈顿距离判断是否实体对齐,以下为损失函数  $d_f(e_1, e_2)$ :

$$d_f(e_1, e_2) = \|x_{e_1}^f - x_{e_2}^f\|_{L_1} \quad (20)$$

式中:实体  $e_1$  与实体  $e_2$  分别来自知识图谱  $G_1$  与知识图谱  $G_2$ ;  $f = \{X_i^{e_{out}}\}$ 。

本文的训练目标是保持对齐实体对(正对)的嵌入距离尽可能小,而未对齐实体对的嵌入距离(负对)尽可能大。 $K$ -nearest<sup>[45]</sup>最近邻采样是最常用的负三元组生成方法,生成基于关系感知嵌入,属性感知嵌入、语义增强嵌入输出的负三元组  $\mathbb{L}'_{pt}, \mathbb{L}'_{rt}, \mathbb{L}'_{at}, \mathbb{L}'_{e^{out}}$ , 最大化负对和正对之间的距离。在训练数据的指导下,通过反向传播对模型进行优化,以学习实体的嵌入。损失函数为:

$$\begin{aligned} \mathcal{L} = & \sum_{(a, b, c) \in k, (a', b', c') \in \mathbb{L}'_{pt}} a'_+ + \theta. + \\ & \sum_{(a, b, c) \in \mathbb{L}, (a'', b'', c'') \in \mathbb{L}'_{rt}} a''_+ + \theta. + \\ & \sum_{(a, b, c) \in \mathbb{L}, (a'', b'', c'') \in \mathbb{L}'_{at}} a''_+ + \theta. + \\ & \sum_{(a, b, c) \in \mathbb{L}, (a''', b''', c''') \in \mathbb{L}'_{e^{out}}} a'''_+ \end{aligned} \quad (21)$$

式中:  $[\cdot]_+$  表示  $\max\{0, \cdot\}$ ;  $\mathbb{L}$  为对齐实体的集合;

$\gamma_1, \gamma_2, \gamma_3, \gamma_4 > 0$  表示用于分辨正三元组与负三元组的超参数;  $\theta$  为平衡 3 个损失部分的折衷参数。

## 4 实验结果与分析

### 4.1 数据集

本文在 WK31-15K<sup>[30]</sup>、DBP-15K<sup>[35]</sup> 和 DWY-100K<sup>[45,18]</sup> 3 个知识图谱对齐常用的数据集上评估了 EPSA 模型。表 2 列出了所有数据集的统计数据。

WK31-15K 包含从 DBpedia 收集的 2 个跨语言数据集:英语-德语(EN-DE)、英语-法语(EN-FR)。每对跨语言知识图谱有 2 个版本:V1 是稀疏的, V2 的密度是 V1 的 2 倍。WK31-15K 数据集包含 4 个子集:EN-DE(V1)、EN-DE(V2)、EN-FR(V1) 和 EN-FR(V2)。

DBP-15K 也像 WK31-15K 一样地收集来自跨语言知识图谱 DBpedia 中的实体, 包含 3 个子集:中文-英语(ZH-EN)、日语-英语(JA-EN)和法语-英语(FR-EN)。而 DBP-15K 中 ZH-EN 的属性值无法获取,无法构建属性三元组,因此在模型中只能使用 JA-EN 和 FR-EN 2 个数据集。为了区别于 WK31-15K,将这 2 个子集分别标记为 JA-EN(DBP) 和 FR-EN(DBP)。

DWY-100K 包含 2 个大型单语言知识图谱:DBpedia 和 Wikidata (DBP-WD)、DBpedia 和 YAGO3 (DBP-YG), 每个知识图谱都有 100K 个对齐的实体对。

表 2 数据集统计

Table 2 Statistics of datasets

| 数据集      |            | 实体        | 关系      | 关系三元组 | 属性      | 属性值   | 属性三元组   |
|----------|------------|-----------|---------|-------|---------|-------|---------|
| WK31-15K | EN-DE(V1)  | EN        | 15 000  | 215   | 47 676  | 286   | 49 956  |
|          |            | DE        | 15 000  | 131   | 50 419  | 194   | 57 661  |
|          | EN-DE(V2)  | EN        | 15 000  | 169   | 84 867  | 171   | 45 805  |
|          |            | DE        | 15 000  | 96    | 92 632  | 116   | 57 652  |
|          | EN-FR(V1)  | EN        | 15 000  | 267   | 47 334  | 308   | 45 783  |
|          |            | FR        | 15 000  | 210   | 40 864  | 404   | 39 772  |
| DBP-15K  | EN-FR(V2)  | EN        | 15 000  | 193   | 96 318  | 189   | 36 391  |
|          |            | FR        | 15 000  | 166   | 80 112  | 221   | 32 412  |
|          | JA-EN(DBP) | JA        | 19 814  | 1 299 | 77 214  | 5 311 | 84 814  |
|          |            | EN        | 19 780  | 1 153 | 93 484  | 4 791 | 71 948  |
|          | FR-EN(DBP) | FR        | 19 661  | 903   | 105 998 | 4 136 | 93 680  |
|          |            | EN        | 19 993  | 1 208 | 115 722 | 5 844 | 88 202  |
| DWY-100K | DBP-WD     | DBpedia   | 100 000 | 330   | 463 294 | 356   | 310 728 |
|          |            | WikiPedia | 100 000 | 220   | 448 774 | 730   | 765 610 |
|          | DBP-YG     | DBpedia   | 100 000 | 302   | 428 952 | 341   | 365 617 |
|          |            | YAsGO3    | 100 000 | 31    | 502 563 | 33    | 567 597 |

#### 4.2 DBP-15K 和 WK31-15K 上的整体性能对比分析

在 DBP-15K 和 WK13-15K 数据集上,将 EP-SA 与之前的 9 种先进的比对模型进行了比较: MTransE<sup>[24]</sup>、IPTransE<sup>[23]</sup>、JAPE<sup>[36]</sup>、BootEA<sup>[26]</sup>、AttrE<sup>[37]</sup>、RDGCN<sup>[21]</sup>、NMN<sup>[22]</sup>、RAGA<sup>[46]</sup>、RHGT<sup>[47]</sup>。在数据集 WK31-15K 和 DBP-15K 中,为减少训练时间,本文只选择 20% 作为训练数据,并将剩余部分留给每次运行的验证(10%)和测试(70%)。嵌入维度大小=300。在训练期间,每 10 个训练周期为每个实体生成 50 个最接近的候选者作为负样本。其他配置为:学习率=0.01,  $\beta_1 = \beta_2 = 1$ ,  $\gamma_1 = \gamma_2 = 5$ ,  $\theta = 0.5$ 。使用了 100 次训练周期的耐心早期停止策略,即如果验证损失在 100 个连续时期内没有减少,则停止训练。EPSA 所有比较模型在跨语言数据集上的比较结果如表 3 所示。

表 3 显示了 EPSA 和所有比较模型在跨语言数

据集上的比较结果,每个表格加粗的数字表示为当前数据集最佳性能。在 WK31-15K 中由 2 个跨语言数据集中实体稀疏密度不同的 4 个数据集组成,每个数据集分别由 2 个不同语言知识图谱组成。由于英语、法语、德语语种上类似,而目前预嵌入中会将数据集统一翻译成英文进行嵌入对齐,其跨语言数据集中实体相似度高,所以本模型 WK31-15K 上取得较好结果。而 DBP-15K 中日语、中文与英文语义差距较大,并且此数据集三元组数量与结构不统一,难以实现精准对齐,目前此数据集性能仍有较大提升空间。如表 3 所示,EP-SA 在所有数据集中与目前其他较优方法比较实现了最佳性能。

#### 4.3 DWY-100K 上的整体性能对比分析

由于只有少数模型在 DWY-100K 数据集上进行了评估,本文将 EPSA 与以下模型进行了比较: MultiKE<sup>[35]</sup>、RDGCN<sup>[21]</sup>、NMN<sup>[22]</sup>、COT-SAE<sup>[39]</sup>、RHGT<sup>[47]</sup>,在数据集 DWY-100K 中,由于与上述数据集相同的原因,选择 30% 作为训练数据,将剩余部分用于测试(70%)。并使用与

表 3 所有模型在 WK31-15K 和 DBP-15K 数据集上的交叉验证结果

Table 3 Cross-validation results of all models on WK31-15K and DBP-15K datasets

%

| 模型      | EN-DE(V1)    |              |              | EN-DE(V2)    |              |              | EN-FR(V1)    |              |              |
|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|         | Hits@1       | Hits@5       | MRR          | Hits@1       | Hits@5       | MRR          | Hits@1       | Hits@5       | MRR          |
| MTransE | 30.70        | 51.85        | 40.73        | 19.37        | 35.21        | 27.42        | 24.77        | 46.77        | 35.16        |
| IPTranE | 35.00        | 51.54        | 43.00        | 47.64        | 67.82        | 57.15        | 16.92        | 32.00        | 24.32        |
| JAPE    | 28.80        | 51.27        | 39.45        | 16.72        | 32.93        | 25.00        | 26.24        | 49.71        | 37.20        |
| BootEA  | 67.50        | 82.00        | 74.00        | 83.35        | 91.25        | 86.99        | 50.70        | 71.82        | 60.34        |
| AttrE   | 57.12        | 68.74        | 59.74        | 65.09        | 81.65        | 72.62        | 48.16        | 67.12        | 56.90        |
| RDGCN   | 81.98        | 87.56        | 84.60        | 81.61        | 86.98        | 84.10        | 80.53        | 87.66        | 83.70        |
| NMN     | 85.57        | 90.45        | 87.70        | 85.18        | 89.57        | 87.10        | 85.12        | 90.74        | 87.66        |
| RAGA    | 87.90        | 94.28        | 90.80        | 81.34        | 89.15        | 84.9         | 82.71        | 91.55        | 86.70        |
| RHGT    | 92.18        | 96.32        | 94.40        | 93.80        | 97.20        | 95.30        | 90.92        | 95.54        | 93.00        |
| EPSA    | <b>94.93</b> | <b>97.17</b> | <b>96.09</b> | <b>97.06</b> | <b>98.84</b> | <b>97.88</b> | <b>93.57</b> | <b>96.22</b> | <b>94.84</b> |
| 模型      | EN-FR(V2)    |              |              | JA-EN(DBP)   |              |              | FR-EN(DBP)   |              |              |
|         | Hits@1       | Hits@5       | MRR          | Hits@1       | Hits@5       | MRR          | Hits@1       | Hits@5       | MRR          |
| MTransE | 24.02        | 43.60        | 33.69        | 20.41        | 40.52        | 30.36        | 19.74        | 40.37        | 29.72        |
| IPTranE | 23.63        | 44.91        | 33.97        | 27.92        | 52.72        | 39.61        | 31.22        | 57.42        | 43.46        |
| JAPE    | 29.24        | 52.48        | 40.23        | 23.86        | 44.59        | 34.00        | 22.98        | 45.22        | 33.66        |
| BootEA  | 66.00        | 85.00        | 74.54        | 52.71        | 71.89        | 61.68        | 57.61        | 77.27        | 66.62        |
| AttrE   | 53.51        | 74.62        | 63.10        | 35.96        | 60.31        | 47.52        | 40.21        | 66.09        | 52.22        |
| RDGCN   | 87.12        | 92.88        | 89.80        | 81.22        | 87.98        | 84.40        | 80.88        | 88.08        | 84.20        |
| NMN     | 89.29        | 94.28        | 91.57        | 84.29        | 90.47        | 87.00        | 83.46        | 90.10        | 86.40        |
| RAGA    | 88.95        | 95.36        | 91.90        | 79.29        | 89.12        | 83.80        | 85.27        | 93.17        | 88.90        |
| RHGT    | 94.95        | 98.00        | 96.30        | 88.64        | 94.30        | 91.20        | 88.92        | 95.59        | 91.90        |
| EPSA    | <b>96.57</b> | <b>98.60</b> | <b>97.66</b> | <b>89.56</b> | <b>95.93</b> | <b>92.28</b> | <b>89.82</b> | <b>95.64</b> | <b>92.25</b> |

NMN [17]模型中相同的实体初始名称嵌入。为了降低训练的复杂性,本文设置了嵌入大小  $d=100$ ,并学习线性变换以获得实体和属性值在 DWY-100K 中的初始名称嵌入。学习率为 0.005,其他配置与上述相同。DWY-100K 上所有模型的性能结果如表 4 所示。

由于 DBpedia 和 YAGO 都源自维基百科,导致 77.60% 的已发布等效实体的名称完全相同,而其余实体的名称非常相似,因此 Hits@1 在表 4 中基本能达到 90% 以上。而由于目前的最新性能已经接近于 1,提升空间已经很小了,本文模型也只能接近于最新性能。

#### 4.4 消融实验

表 5 显示了 EPSA 模型不同变体的消融实验的性能比较。EPSA“w/o EP”是对实体路径信息模块的消融实验,EPSA“w/o SA”是对知识图谱语义增强模块的消融实验,以分别证实本文实体路径信息模块与知识图谱语义增强模块性能的提升。当删除实体路径信息模块时,性能会下降,Hit@1 下降 0.9%~1.6%。但通过知识图谱语义增强模块聚合不同异质信息性能后,相比目前最佳模型效果依然得到提升;当删除知识图谱语义增强模块时,性能 Hit@1 下降 0.7%~1.8%。但通过实体路径信息模块获得实体路径匹配信息后,相比目前最佳模型效果依然得到提升。

**表 4 DWY-100K 数据集上所有模型的总体性能**  
**Table 4 Overall performance of all models on DWY-100K datasets** %

| 模型       | DBP-WD |        |       | DBP-YG |        |       |
|----------|--------|--------|-------|--------|--------|-------|
|          | Hits@1 | Hits@5 | MRR   | Hits@1 | Hits@5 | MRR   |
| MultitKE | 91.86  | 96.26  | 93.55 | 88.03  | 95.32  | 90.68 |
| RDGCN    | 97.90  | 99.13  | —     | 94.75  | 97.34  | —     |
| NMN      | 98.12  | 99.20  | —     | 96.00  | 98.27  | —     |
| COTSAE   | 92.68  | 97.86  | 94.57 | 94.39  | 98.74  | 96.14 |
| RSA      | 98.54  | 99.29  | —     | 97.20  | 97.90  | —     |
| RHGT     | 99.26  | 99.86  | 99.50 | 96.58  | 98.86  | 97.40 |
| EPSA     | 99.42  | 99.27  | 99.77 | 98.72  | 98.91  | 99.25 |

**表 5 EPSA 不同模块的消融实验**

**Table 5 Ablation experiments on different modules of EPSA** %

| 模型       | EN-DE(V1) |        |       | EN-DE(V2)  |        |       | EN-FR(V1)  |        |       |
|----------|-----------|--------|-------|------------|--------|-------|------------|--------|-------|
|          | Hits@1    | Hits@5 | MRR   | Hits@1     | Hits@5 | MRR   | Hits@1     | Hits@5 | s     |
| EPSA     | 94.93     | 97.17  | 96.09 | 97.06      | 98.84  | 97.88 | 93.57      | 96.22  | 94.84 |
| RHGT     | 92.18     | 96.32  | 94.40 | 93.80      | 97.20  | 95.30 | 90.92      | 95.54  | 93.00 |
| (w/o EP) | 93.31     | 96.51  | 95.50 | 95.21      | 97.80  | 96.90 | 92.21      | 96.15  | 94.50 |
| (w/o SA) | 93.75     | 96.85  | 95.08 | 95.73      | 98.47  | 96.96 | 92.45      | 95.89  | 94.62 |
| 模型       | EN-FR(V2) |        |       | JA-EN(DBP) |        |       | FR-EN(DBP) |        |       |
|          | Hits@1    | Hits@5 | MRR   | Hits@1     | Hits@5 | MRR   | Hits@1     | Hits@5 | MRR   |
| EPSA     | 96.57     | 98.60  | 97.66 | 89.56      | 95.93  | 92.28 | 89.82      | 95.64  | 92.25 |
| RHGT     | 94.95     | 98.00  | 96.30 | 88.64      | 94.30  | 91.20 | 88.92      | 95.59  | 91.90 |
| (w/o EP) | 95.10     | 98.14  | 96.80 | 86.95      | 92.69  | 89.65 | 87.23      | 93.31  | 90.50 |
| (w/o SA) | 95.25     | 98.58  | 97.11 | 87.32      | 92.86  | 89.82 | 87.65      | 93.72  | 91.25 |
| 模型       | DBP-WD    |        |       | DBP-YG     |        |       |            |        |       |
|          | Hits@1    | Hits@5 | MRR   | Hits@1     | Hits@5 | MRR   |            |        |       |
| EPSA     | 99.42     | 99.27  | 99.77 | 98.72      | 98.91  | 99.25 |            |        |       |
| RHGT     | 99.26     | 99.86  | 99.50 | 96.58      | 98.86  | 97.40 |            |        |       |
| (w/o EP) | 98.49     | 99.27  | 99.60 | 98.94      | 98.99  | 98.45 |            |        |       |
| (w/o SA) | 98.21     | 99.59  | 99.57 | 98.55      | 99.26  | 98.36 |            |        |       |

#### 4.5 EN-DE(V2)上的性能对比分析

为了更好地评估 EPSA 的有效性,本文测试了 EN-DE(V2)上的 Hits@[1,3,5,10,50,100] 的性能对比,实验结果如图 5 所示。

由图 5 可知,EPSA 在所有评估指标上优于其他模型。尤其是当 Hits@ $\leq 0$  时,EPSA 明显优于其他模型。

为了进一步分析不同大小训练数据集下 EPSA 对比试验的有效性,本文以 EN-DE(V2)数据集为例进行进一步评估。本文将训练数据集的比例从 5% 改变到 30%,每一次增长 5%,验证数据集保持在 10%,实验结果如图 6 所示。

图 6 说明了不同比例的训练数据集的 Hits@[1,5] 性能。EN-DE(V2)上所有 Hits@1 的性能

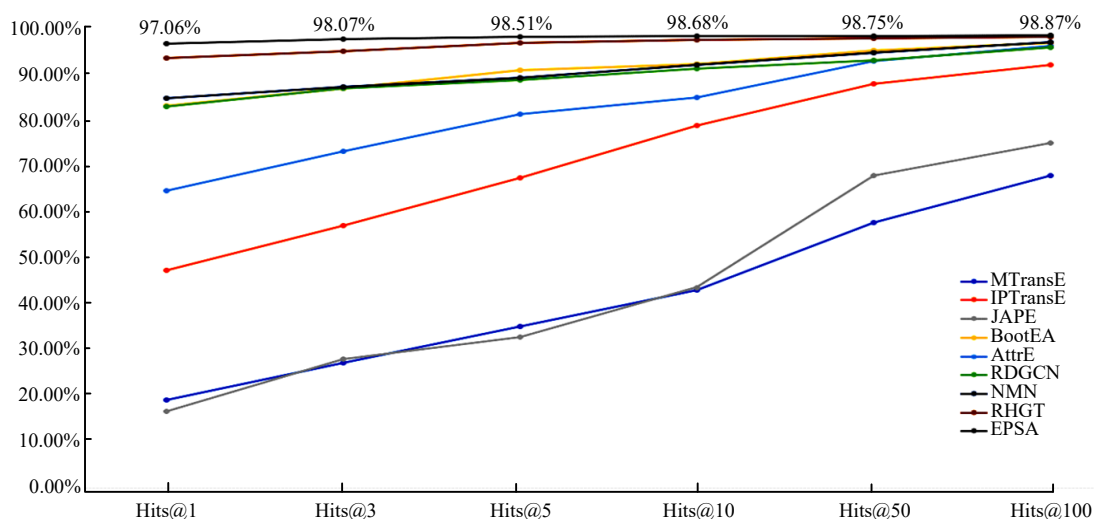


图 5 EPSA 与所有模型在 EN-DE (V2) 上的所有性能 Hits@k 比较

Fig. 5 Hits@k performance of EPSA and all comparison models on EN-DE (V2)

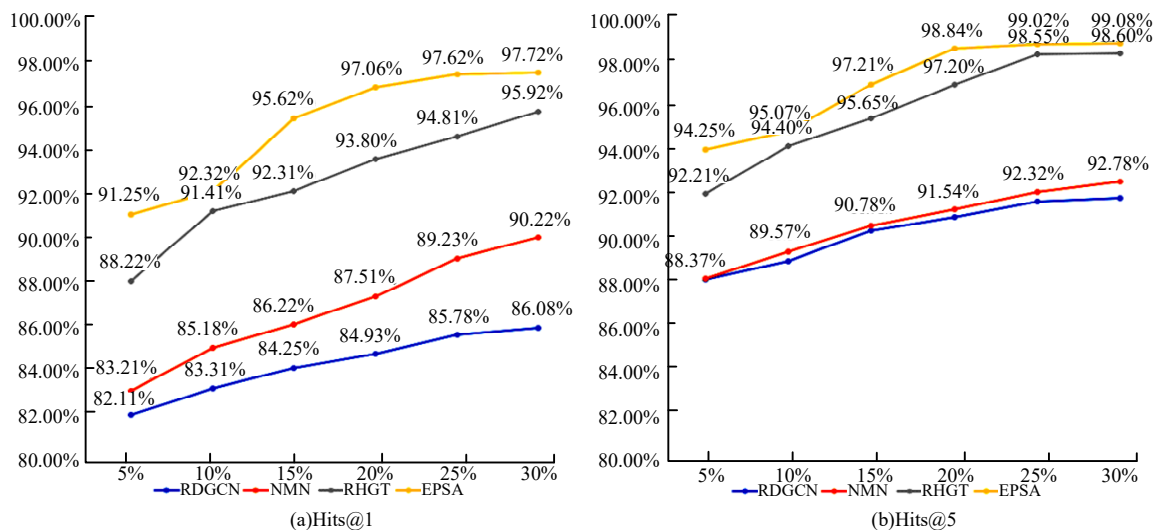


图 6 EPSA 与在 EN-FR (V2) 上不同比例训练数据集的 Hits@ [1, 5] 性能。

Fig. 6 Hits @ [1, 5] performance of EPSA and training datasets with different proportions on EN-FR (V2)

都随着比例的提高而提高,而 EPSA 的性能依然保持了领先,并且 EPSA 具有很强的鲁棒性,模型在较少的训练数据下仍能获得令人满意的结果。

## 5 结束语

对齐实体不仅具有相似的关系结构,还存在大量的多步骤关系路径,以表明实体之间的语义关系。本文提出一种用于实体对齐的实体可靠路径信息语义增强模型(EPSA),该模型挖掘不同知识图谱实体的邻域路径信息,考虑了知识图谱实体的路径信息,将实体的关系结构、属性结构和实体名称路径信息三方面通过模型聚合进行语义增强。该方法考虑了知识图谱实体的路径信息,提高对齐判断的准确性和鲁棒性,同时将每个实

体的关系结构、属性结构、实体路径结构的输出结果进行编码并嵌入低维空间中,然后使用语义增强模型聚合实体信息,由此增强实体各结构之间的联系,以解决实体异构邻域中的异质性问题。与最先进的方法相比,该模型在各种情况下都能获得最佳的对比结果,在 Hits@1 上高于性能最佳的模型。

## 参考文献:

- [1] 刘峤,李杨,段宏,等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
- 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] 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.
- [3] Suchanek F, Kasneci G, Weikum G. YAGO: a core of semantic knowledge unifying WordNet and wikipedia[C]//*Proceedings of the 16th International Conference on World Wide Web*, New York, USA, 2007: 697–706.
- [4] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]//*Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, New York, USA, 2008: 1247–1250.
- [5] Xu B, Xu Y, Liang J, et al. CN-dbpedia: a never-ending Chinese knowledge extraction system[C]//*Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Berlin: Springer, 2017: 428–438.
- [6] 张婷婷, 欧阳丹彤, 孙成林, 等. 融合协同过滤的神经 Bandits 推荐算法[J]. *吉林大学学报: 理学版*, 2024, 62(1): 92–99.  
Zhang Ting-ting, Ouyang Dan-tong, Sun Cheng-lin, et al. Neural bandits recommendation algorithm based on collaborative filtering[J]. *Journal of Jilin University (Science Edition)*, 2024, 62(1): 92–99.
- [7] 孟令鑫, 才华, 付强, 等. 基于关系记忆与路径信息的多跳知识图谱问答算法[J]. *吉林大学学报: 理学版*, 2024, 62(6): 1391–1400.  
Meng Ling-xin, Cai Hua, Fu Qiang, et al. Multi-hop knowledge graph question answering algorithm based on relational memory and path information[J]. *Journal of Jilin University (Science Edition)*, 2024, 62(6): 1391–1400.
- [8] 李鑫, 王文迪, 张伟, 等. 基于知识嵌入技术的制度文件推荐算法[J]. *吉林大学学报: 理学版*, 2024, 62(6): 1377–1383.  
Li Xin, Wang Wen-di, Zhang Wei, et al. Recommendation algorithm for institutional documents based on knowledge embedding technology[J]. *Journal of Jilin University (Science Edition)*, 2024, 62(6): 1377–1383.
- [9] 化青远, 彭涛, 崔海, 毕海嘉. 基于知识图谱中路径推理的多轮对话模型[J]. *吉林大学学报: 理学版*, 2025, 63(1): 76–82.  
Hua Qing-yuan, Peng Tao, Cui Hai, et al. Multi-round conversational model based on path reasoning in knowledge graph[J]. *Journal of Jilin University (Science Edition)*, 2025, 63(1): 76–0082.
- [10] 何山, 肖晰, 张嘉玲. 面向领域知识图谱的实体关系抽取模型仿真[J]. *吉林大学学报: 理学版*, 2025, 63(2): 465–471.  
He Shan, Xiao Xi, Zhang Jia-ling. Simulation of entity relationship extraction model for domain knowledge graph[J]. *Journal of Jilin University (Science Edition)*, 2025, 63(2): 465–471.
- [11] 费敏学, 黄东岩, 郭晓新. 改进蜣螂算法优化机器学习模型[J]. *吉林大学学报: 理学版*, 2025, 63(4): 1117–1121.  
Fei Min-xue, Huang Dong-yan, Guo Xiao-xin. Improve dung beetle algorithm to optimize machine learning model[J]. *Journal of Jilin University (Science Edition)*, 2025, 63(4): 1117–1121.
- [12] 汪雨竹, 彭涛, 朱蓓蓓, 等. 基于元学习的小样本知识图谱补全[J]. *吉林大学学报: 理学版*, 2023, 61(3): 623–630.  
Wang Yu-zhu, Peng Tao, Zhu Bei-bei, et al. Few-shot knowledge graph completion based on meta learning[J]. *Journal of Jilin University (Science Edition)*, 2023, 61(3): 623–630.
- [13] Lu W, Wang P, Ma X, et al. Enrich cross-lingual entity links for inline wikis via multi-modal semantic matching[J]. *Information Processing & Management*, 2020, 57(5): 102271.
- [14] 王雪鹏, 刘康, 何世柱, 等. 基于网络语义标签的多源知识库实体对齐算法[J]. *计算机学报*, 2017, 40(3): 701–711.  
Wang Xue-peng, Liu Kang, He Shi-zhu, et al. Multi-source knowledge bases entity alignment by leveraging semantic tags[J]. *Chinese Journal of Computers*, 2017, 40(3): 701–711.
- [15] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network[C]//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Anchorage, USA, 2019: 793–803.
- [16] 庄严, 李国良, 冯建华. 知识库实体对齐技术综述[J]. *计算机研究与发展*, 2016, 53(1): 165–192.  
Zhuang Yan, Li Guo-liang, Feng Jian-hua. A survey on entity alignment of knowledge base[J]. *Journal of Computer Research and Development*, 2016, 53(1): 165–192.

- [17] 乔晶晶, 段利国, 李爱萍. 融合多种特征的实体对齐算法[J]. 计算机工程与设计, 2018, 39(11): 3395-3400.
- Qiao Jing-jing, Duan Li-guo, Li Ai-ping. Entity alignment algorithm based on multi-features[J]. Computer Engineering and Design, 2018, 39(11): 3395-3400.
- [18] Kipf N, Welling M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 2016: 160902907.
- [19] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[C]//Proceedings of the 6th International Conference on Learning Representations, Vancouver, Canada, 2018: 171010903v3.
- [20] Wu Y, Liu X, Feng Y, 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, 2019: 240-249.
- [21] Wu Y, Liu X, Feng Y, et al. Relation-aware entity alignment for heterogeneous knowledge graphs[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 2019: 5278-5284.
- [22] 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, Virtual, 2020: 6477-6487.
- [23] Bordes A, Usunier N, García D A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013: 2787-2795.
- [24] Chen M, Tian Y, Chang K. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Jeju, South Korea, 2018: 3998-4004.
- [25] Zhu Q, Wei H, Sisman B, et al. Collective multi-type entity alignment between knowledge graphs[C]//Proceedings of the 2020 World Wide Web Conference, Taipei, China, 2020: 2241-2252.
- [26] Kipf T, Welling M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 2017: 160902907.
- [27] Chen L, Gu W, Tian X, et al. AHAB: aligning heterogeneous knowledge bases via iterative blocking[J]. Information Processing & Management, 2019, 56(1): 1-13.
- [28] Cao Y, Liu Z, Li C, et al. Multi-channel graph neural network for entity alignment[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019: 1452-1461.
- [29] 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, Virtual, 2020: 6477-6487.
- [30] Sun Z. A benchmarking study of embedding-based entity alignment for knowledge graphs[C]//Proceedings of the VLDB Endowment, Tokyo, Japan, 2020: 2326-2340.
- [31] Wang Z, Lan X, Zhang Y, 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, 2018: 349-357.
- [32] 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, Hong Kong, China, 2019: 4430-4440.
- [33] Ye R, Li X, Fang Y, et al. A vectorized relational graph convolutional network for multi-relational network alignment[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 2019: 4135-4141.
- [34] Xu K, Wang L, Yu M, et al. Cross-lingual knowledge graph alignment via graph matching neural network[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019: 3156-3161.
- [35] Zhang Q, Sun Z, Hu W, et al. Multi-view knowledge graph embedding for entity alignment[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 2019: 5429-5435.
- [36] Sun Z, Hu W. Cross-lingual entity alignment via

- joint attribute-preserving embedding[C]//Proceedings of the International Semantic Web Conference, Vienna, Austria, 2017: 628-644.
- [37] Trisedya B, Qi J, Zhang R. Entity alignment between knowledge graphs using attribute embeddings [C]//Proceedings of the 33th AAAI Conference on Artificial Intelligence, Hawaii, USA, 2019: 297-304.
- [38] Liu Z, Cao Y, Pan L, et al. Exploring and evaluating attributes, values, and structures for entity alignment[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Virtual, 2020: 6355-6364.
- [39] Yang K, Liu S, Zhao J, et al. COTSAE: co-training of structure and attribute embeddings for entity alignment[C]//Proceedings of 34th AAAI Conference on Artificial Intelligence, New York, USA, 2020: 3025-3032.
- [40] Chen B, Zhang J, Tang X, et al. JarKA: modeling attribute interactions for cross-lingual knowledge alignment[C]//Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, Singapore, Singapore, 2020: 845-856.
- [41] Chen M, Tian Y, Chang K, et al. Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Jeju, South Korea, 2018: 3998-4004.
- [42] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, USA, 2013: 3111-3119.
- [43] Rahimi A, Cohn T, Baldwin T. Semi-supervised user geolocation via graph convolutional networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018: 2009-2019.
- [44] Liu P, Li H, Wang Z, et al. Multi-features based semantic augmentation networks for named entity recognition in threat intelligence[J]. International Conference on Pattern Recognition, 2022, 7: 250243626.
- [45] Sun Z, Hu W, Zhang Q, et al. Bootstrapping entity alignment with knowledge graph embedding[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 4396-4402.
- [46] Zhu R B, Ma M, Wang P. RAGA: relation-aware graph attention networks for global entity alignment [C]//Advances in Knowledge Discovery and Data Mining: 25th Pacific-Asia Conference, Virtual, 2021: 501-513.
- [47] Cai W S, Ma W J, Zhan J Y, et al. Entity alignment with reliable path reasoning and relation-aware heterogeneous graph transformer[C]//Proceedings of the 31st International Joint Conference on Artificial Intelligence and the 25th European Conference on Artificial Intelligence, Vienna, Austria, 2022: 1930-1937.