

论文复现：利用元模态混合的多模态实体对齐

龙智亮

2024 年 12 月 5 日

摘要

多模态实体对齐 (MMEA) 旨在发现不同知识图谱 (KGs) 中与相关图像关联的相同实体。然而, 当前的 MMEA 算法依赖于知识图谱级别的模态融合策略来进行多模态实体表示, 这忽略了不同实体模态偏好的变化, 从而在面对模态中的噪声 (如模糊图像和关系) 时影响了鲁棒性。本文介绍了 MEAformer, 一种用于元模态混合的多模态实体对齐转换器方法, 它动态预测模态之间的相互关联系数, 以实现更细粒度的实体级别模态融合和对齐, 以及自己的论文复现工作。实验结果表明该模型不仅在多种训练场景中实现了最先进的性能, 包括监督、无监督、迭代和低资源设置, 而且还具有参数数量有限、运行效率高和可解释性等特点。原论文的代码可以在 github.com/zjukg/MEAformer 上获取。而本文的复现代码可以在 github.com/ChinaStark/szucs 上获取。

关键词: 实体对齐; 多模态学习; 知识图谱; 模态混合; 元学习;

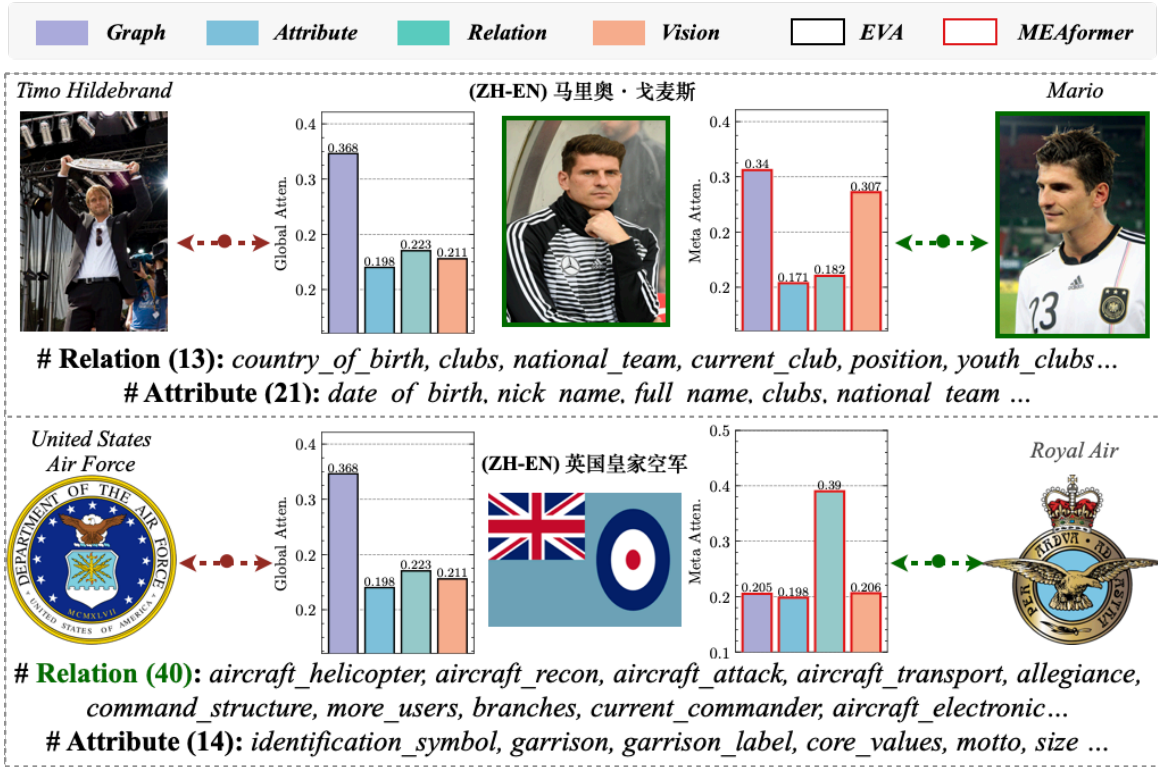


图 1. EVA 中的静态模态权重 [17] (左) 本文的动态元模态权重 (右)。

1 引言

近年来,知识图谱 (KGs) 在诸如问答 [6,7] 和人工智能科学研究 [8,9] 等众多人工智能应用中发挥了越来越重要的作用,提供了常识性知识,因此本文选择实体对齐这个大方向。作为知识图谱整合和构建中的关键任务,实体对齐 (EA) 旨在识别不同知识图谱中的等价实体,同时解决诸如不同的命名习惯、多语言和异构图结构等挑战。为了利用互联网上的视觉内容作为 EA 的补充信息,提出了多模态实体对齐 (MMEA) 也就是本文所选论文所属的细方向,在这种方法中,每个实体都与其名称相关的图像相关联 [3]。

[17] 引入了模态特定的注意力权重学习以确定模态的重要性; [4] 整合视觉特征以指导关系和属性学习; [16] 应用 KL 散度在联合和单模态实体嵌入的输出分布之间减少模态差异。然而,这些方法都学习知识图谱级别的模态融合权重,忽略了每个实体内部模态差异 (例如,节点度数或关系数量) 和模态偏好 (例如,模态缺失或模糊性)。这些缺陷在某种程度上影响了它们的鲁棒性。

在这篇工作中,该作者探索了一种多模态实体混合特征的新范式,并且命名为 MEAformer。具体来说,是一种新型多模态实体对齐变换方法,它动态预测每个实体之间模态的相对互权重 (见图 1 中的例子)。为了实现类似于元学习的范式,提出了动态跨模态加权 (DCMW) 模块为模态校正生成元模态权重,通过浅层交叉注意力网络实现模态间的相互评分。此外,还采用模态自适应对比学习目标,以有限的预对齐实体数量高效地分离模态信息,并引入模态感知的困难实体重放,以进一步增强模型对模糊实体细节的鲁棒性。总的来说,本文所选的这篇文章的主要贡献为:

- 提出现有 MMEA 研究中知识图谱级别模态融合策略的局限性,并提出使用更具适应性的实体级别模态融合策略,即元模态融合动态预测模态之间的权重。
- 实验证明 MEAformer 能够在多种训练场景下实现最先进的性能,包括监督、半监督、无监督。

而本复现工作主要为:

- 在原有的数据集上,利用规则推理将数据集中的三元组扩充至原来的 1.7 倍,极大的扩充了训练数据。
- 将规则损失函数引入原文,并对实验结果进行了详尽的分析。

2 相关工作

2.1 实体对齐

实体对齐 (EA) 旨在发现不同知识图谱中的等价实体,以促进知识融合。早期的 EA 系统利用逻辑推理和词汇匹配等技术进行实体映射构建,严重依赖于 (特别设计的) 启发式方法 [14,20]。近期基于嵌入的 EA 方法通过学习一个嵌入空间来表示那些待对齐的知识图谱,其中相似的实体位置接近,而不相似的实体则相隔较远,从而缓解了异构性问题 [23]。具体来说,我们将它们分为两类 [31]:

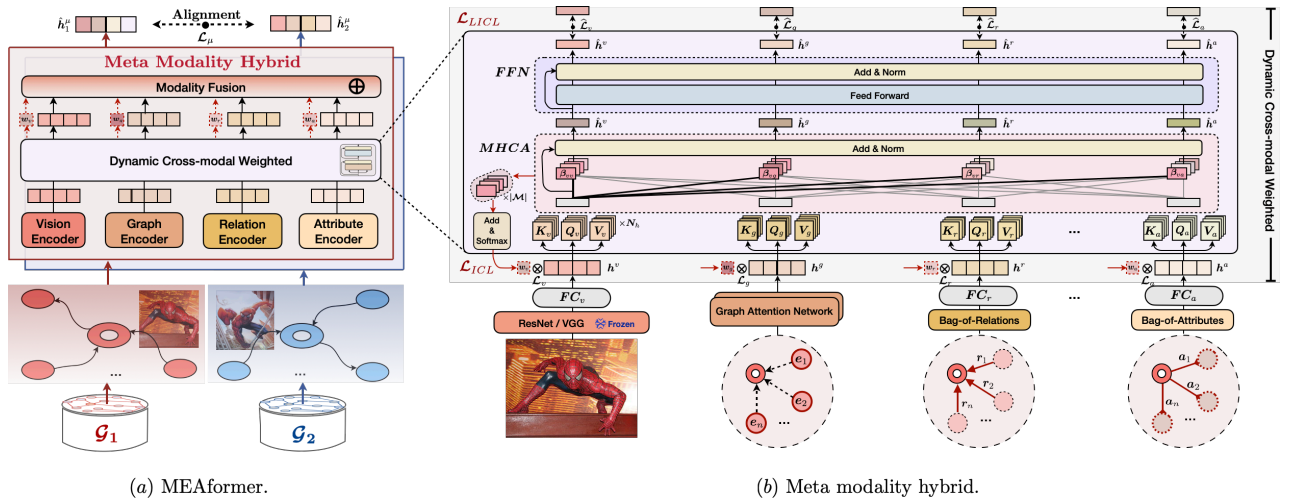
(1) 基于图神经网络 (GNNs) 的 *EA* 方法主要利用图神经网络 (GNNs) 如 GCN [15] 和 GAT [27] 进行邻域实体特征聚合。

(2) 基于翻译的 *EA* 方法采用 TransE [1] 等基于翻译的知识图谱嵌入方法, 试图捕获关系三元组中的实体结构信息。

通常, 这些方法会在部分预对齐的实体对 (即种子对齐) 上应用对齐目标 (例如, 嵌入余弦相似度), 以校准知识图谱的语义空间。先前的工作表明, 可以通过额外的策略增强实体对齐的进展, 例如: 参数共享 [33] (即, 在不同知识图谱间共享种子对齐的实体嵌入); 在多个异构知识图谱之间显式链接种子对齐 [33]; 迭代学习 [24] (即, 迭代地将实体对标记为伪种子监督); 属性值编码 [25]; 集体稳定匹配用于相互依赖的对齐决策 [30]; 或通过本体模式引导实体对齐 [28]。

2.2 多模态实体对齐

自从被引入作为多模态知识图谱 (MMKG) 构建任务以来, 将视觉模态纳入知识图谱中的实体对齐已逐渐受到关注, 随着近年来多模态学习的发展。[3] 融合了模态的知识表示, 然后最小化对齐实体的整体嵌入之间的距离。[17] 应用可学习的注意力权重方案, 以给每个模态不同的重视程度。[4] 整合视觉特征以指导关系特征学习, 同时为对齐分配权重有价值的属性。同时, [4] 通过对比学习进一步增强了模态内学习, 并应用 KL 散度在联合和单模态嵌入的输出分布之间减少模态差异。然而, 所有这些方法都忽略了每个实体的动态模态间效应。在现实世界的实体对齐场景中, 这是不可忽视的, 因为知识图谱中 (尤其是多模态知识图谱) 不可避免地存在错误和噪声, 例如那些包含无法识别的图像的知识图谱。此外, 模态内特征差异 (例如, 节点度数) 和模态间源偏好 (例如, 模态缺失、不平衡或模糊现象) 在知识图谱中也很常见。而 MEAformer, 它提供了一个完全动态的元模态混合策略来解决上述问题。



(a) MEAformer.

(b) Meta modality hybrid.

图 2. (a) MEAformer 的整体框架和 (b) 元模态混合 (MMH) 的实现细节。

3 原论文核心方法

这里这里采用原文的定义, 即定义一个多模态知识图谱 (MMKG) 为一个五元组, 即 $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{A}, \mathcal{V}, \mathcal{T}\}$. $\mathcal{E}, \mathcal{R}, \mathcal{A}$ 和 \mathcal{V} 分别表示实体集、关系集、属性集和图像集。 $\mathcal{T} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ 是关系三元组的集合。给定两个 MMKGs $\mathcal{G}_1 = \{\mathcal{E}_1, \mathcal{R}_1, \mathcal{A}_1, \mathcal{V}_1, \mathcal{T}_1\}$ 和 $\mathcal{G}_2 = \{\mathcal{E}_2, \mathcal{R}_2, \mathcal{A}_2, \mathcal{V}_2, \mathcal{T}_2\}$,

多模态实体对齐 (MMEA) 旨在识别每对实体 (e_i^1, e_i^2) , 其中 $e_i^1 \in \mathcal{E}_1$, $e_i^2 \in \mathcal{E}_2$ 且 e_i^1 和 e_i^2 对应于同一个现实世界实体 e_i 。为了清晰起见, 在上下文中省略了表示实体来源知识图谱的上标符号, 除非在陈述或公式中明确需要。每个实体都与多个属性相关联, 并且有 0 或 1 张图像。一组预对齐的实体对被提供, 它们根据给定的种子对齐比例 (R_{sa}) 成比例地分为训练集 (种子对齐 \mathcal{S}) 和测试集 \mathcal{S}_{te} 。此外, 用 \mathcal{M} 表示可用的模态集。

3.1 多模态知识嵌入

本小节详细说明了如何将给定 MMKGs 中每个实体的模态 m 嵌入到低维向量 h^m 中。

3.1.1 图邻域结构嵌入

设 $x_i^g \in \mathbb{R}^d$ 表示实体 e_i 的随机初始化图嵌入, 其中 d 是预设的隐藏维度。这里采用一个 (具有两个注意力头和两层的) 图注意力网络 (GAT) [27] 来捕获 \mathcal{G} 的结构信息, 并配备一个对角权重矩阵 $W_g \in \mathbb{R}^{d \times d}$ 用于线性变换:

$$h_i^g = GAT(W_g, M_g; x_i^g), \quad (1)$$

其中 M_g 表示图邻接矩阵。

3.1.2 关系、属性、视觉和表面嵌入。

为了避免在 GNN 网络中混合关系和属性表示带来的信息污染 [17], 这里应用由 $W_m \in \mathbb{R}^{d_m \times d}$ 参数化的独立全连接层来转换这些特征 x^m :

$$h_i^m = FC_m(W_m, x_i^m), \quad m \in \{r, a, v, s\}, \quad (2)$$

其中 r, a, v, s 分别表示关系、属性、视觉和表面 (即实体名称) 模态。 $x_i^m \in \mathbb{R}^{d_m}$ 是实体 e_i 对应模态 m 的输入特征。这里遵循 Yang 等人的方法 [29], 使用词袋特征来表示关系 (x^r) 和属性 (x^a), 其类型作为最小单位。而对于视觉模态, 这里采用预训练的视觉模型作为编码器 (Enc_v), 以获取实体 e_i 的每个可用图像 v_i 的视觉嵌入 x_i^v , 其中最终层输出在 logits 之前作为图像特征。整体模型图如图2所示, 这里省略了表面编码器以提高清晰度。

3.2 元模态混合

本小节介绍了元模态混合 (MMH) 模块, 该模块实现了实体的动态模态融合。

3.2.1 动态跨模态加权 (DCMW)

DCMW 旨在为每个模态动态生成实体级别的元权重。受到原始变换器的启发 [26], 这里在 DCMW 中引入了两种类型的子层: 多头跨模态注意力 (MHCA) 块和全连接前馈网络 (FFN)。具体来说, MHCA 在其注意力函数中操作 N_h 个并行头。第 i 个头由模态共享矩阵 $W_q^{(i)}, W_k^{(i)}, W_v^{(i)} \in \mathbb{R}^{d \times d_h}$ 参数化, 将多模态输入 h^m 转换为模态感知查询 $Q_m^{(i)}$, 键 $K_m^{(i)}$ 和值 $V_m^{(i)}$ 在 \mathbb{R}^{d_h} 中 ($d_h = d/N_h$):

$$Q_m^{(i)}, K_m^{(i)}, V_m^{(i)} = h^m W_q^{(i)}, h^m W_k^{(i)}, h^m W_v^{(i)}. \quad (3)$$

对于模态 m 的特征，其输出为：

$$\text{MHCA}(h^m) = \bigoplus_{i=1}^{N_h} \text{head } i^m \cdot W_o, \quad (4)$$

$$\text{head}_i^m = \sum_{j \in \mathcal{M}} \beta_{mj}^{(i)} V_j^{(i)}, \quad (5)$$

其中 $W_o \in \mathbb{R}^{d \times d}$ 且 \bigoplus 表示连接操作。每个头中实体模态 m 和 j 之间的注意力权重 (β_{mj}) 如下公式化：

$$\beta_{mj} = \frac{\exp(Q_m^\top K_j / \sqrt{d_h})}{\sum_{n \in \mathcal{M}} \exp(Q_m^\top K_n / \sqrt{d_h})}, \quad (6)$$

其中 $d_h = d/N_h$ 。此外，使用层归一化 (LN) 和残差连接 (RC) 来稳定训练：

$$\hat{h}^m = \text{LayerNorm}(\text{MHCA}(h^m) + h^m). \quad (7)$$

FFN 由两个线性变换层组成，中间有一个 ReLU 激活函数，之后应用 LN&RC，如下所示：

$$\text{FFN}(\hat{h}^m) = \text{ReLU}(\hat{h}^m W_1 + b_1) W_2 + b_2, \quad (8)$$

$$\hat{h}^m \leftarrow \text{LayerNorm}(\text{FFN}(\hat{h}^m) + \hat{h}^m), \quad (9)$$

其中 $W_1 \in \mathbb{R}^{d \times d_{in}}$ 且 $W_2 \in \mathbb{R}^{d_{in} \times d}$ 。这里注意到，在这里的 MEAformer 中，FFN 是一个可选组件，因为减少参数有时可以减轻过拟合，尤其是在数据不太复杂的情况下（例如，属性类型较少）。这里定义每个模态 m 的输出元权重 w_m 为：

$$w_m = \frac{\exp(\sum_{j \in \mathcal{M}} \sum_{i=0}^{N_h} \beta_{mj}^{(i)} / \sqrt{|\mathcal{M}| \times N_h})}{\sum_{k \in \mathcal{M}} \exp(\sum_{j \in \mathcal{M}} \sum_{i=0}^{N_h} \beta_{kj}^{(i)} / \sqrt{|\mathcal{M}| \times N_h})}, \quad (10)$$

它捕获了关键的模态间接口信息，并适应性地调整模型对不同模态的跨模态对齐偏好。

3.2.2 模态融合

设 w_m^i 为实体 e_i 对模态 m 的元权重；这里定义联合嵌入为：

$$h_i^\mu = \bigoplus_{m \in \mathcal{M}} [w_m^i h_i^m], \quad h_i^\xi = \bigoplus_{m \in \mathcal{M}} [w_m^i \hat{h}_i^m], \quad (11)$$

其中 h_i^μ 和 h_i^ξ 分别定义为早期和晚期融合嵌入。作为先前 EA 工作中模态融合的常见范式 [4, 16, 17]，连接操作可以用来防止模态多样性的模糊和过度平滑。这里根据实验选择 h_i^μ 作为最终的实体表示用于评估，并推测晚期融合嵌入中的模态特异性在变换器层 [32] 中被持续削弱，这减少了实体之间的区别。

3.3 模态自适应对比学习以及损失函数

本小节介绍了模态自适应对比学习，以充分利用有限的种子对齐 \mathcal{S} 挖掘每个模态中包含的信息。具体来说，这里对种子对齐 \mathcal{S} 进行负样本对齐的破坏，遵循 EA 研究中使用的 1 对 1 对齐假设 [24]。对于 \mathcal{S} 中的每个实体对 (e_i^1, e_i^2) ，这里定义 $\mathcal{N}_i^{ng} = \{e_j^1 | \forall e_j^1 \in \mathcal{E}_1, j \neq i\} \cup \{e_j^2 | \forall e_j^2 \in \mathcal{E}_2, j \neq i\}$ 作为其负实体集。此外，这里采用批次内负采样策略 [5]，以提高效率，将 \mathcal{N}_i^{ng} 的采样范围限制在小批量内。对齐概率定义如下：

$$p_m(e_i^1, e_i^2) = \frac{\gamma_m(e_i^1, e_i^2)}{\gamma_m(e_i^1, e_i^2) + \sum_{e_j \in \mathcal{N}_i^{ng}} \gamma_m(e_i^1, e_j)}, \quad (12)$$

其中 $\gamma_m(e_i, e_j) = \exp(h_i^{m\top} h_j^m / \tau)$, τ 是温度超参数。考虑到方程(12)中反映的实体对的对齐方向, 这里为每个模态 (m) 定义双向对齐目标为:

$$\mathcal{L}_m = -\log(p_m(e_i^1, e_i^2) + p_m(e_i^2, e_i^1)) / 2. \quad (13)$$

此外, 这里引入模态内对比损失 \mathcal{L}_{ICL} , 鼓励模型明确对齐多模态特征 h^m [16]。同时, 这里涉及晚期模态内对比损失 \mathcal{L}_{LICL} , 通过跨模态注意力知识转移实现模态间的相互信息补充:

$\mathcal{L}_{ICL} = \sum_{m \in \mathcal{M}} \mathcal{L}_m$, $\mathcal{L}_{LICL} = \sum_{m \in \mathcal{M}} \hat{\mathcal{L}}_m$, (14) 其中 $\hat{\mathcal{L}}_m$ 是 \mathcal{L}_m 的变体, $\hat{\gamma}_m(e_i, e_j) = \exp(\hat{h}_i^{m\top} \hat{h}_j^m / \tau)$ 。最后, 这里训练网络以最小化总体损失:

$$\mathcal{L} = \mathcal{L}_\mu + \mathcal{L}_{ICL} + \mathcal{L}_{LICL}, \quad (15)$$

其中 \mathcal{L}_μ 基于联合早期融合嵌入 h^μ , 遵循方程(15), $m = \mu$ 。

4 复现细节

逻辑规则在知识获取和推理中得到广泛研究, 通常基于马尔可夫逻辑网络 [19], 最近, 学术界对结合逻辑规则和嵌入模型越来越感兴趣。比如 [11] 尝试通过整数线性规划或马尔可夫逻辑网络利用规则来完善嵌入模型所做的预测。因此在这里我们将其引入来增加实体的背景信息, 让模型更好的把实体对齐。实验结果表明, 在有些数据集上效果有一定的提升。

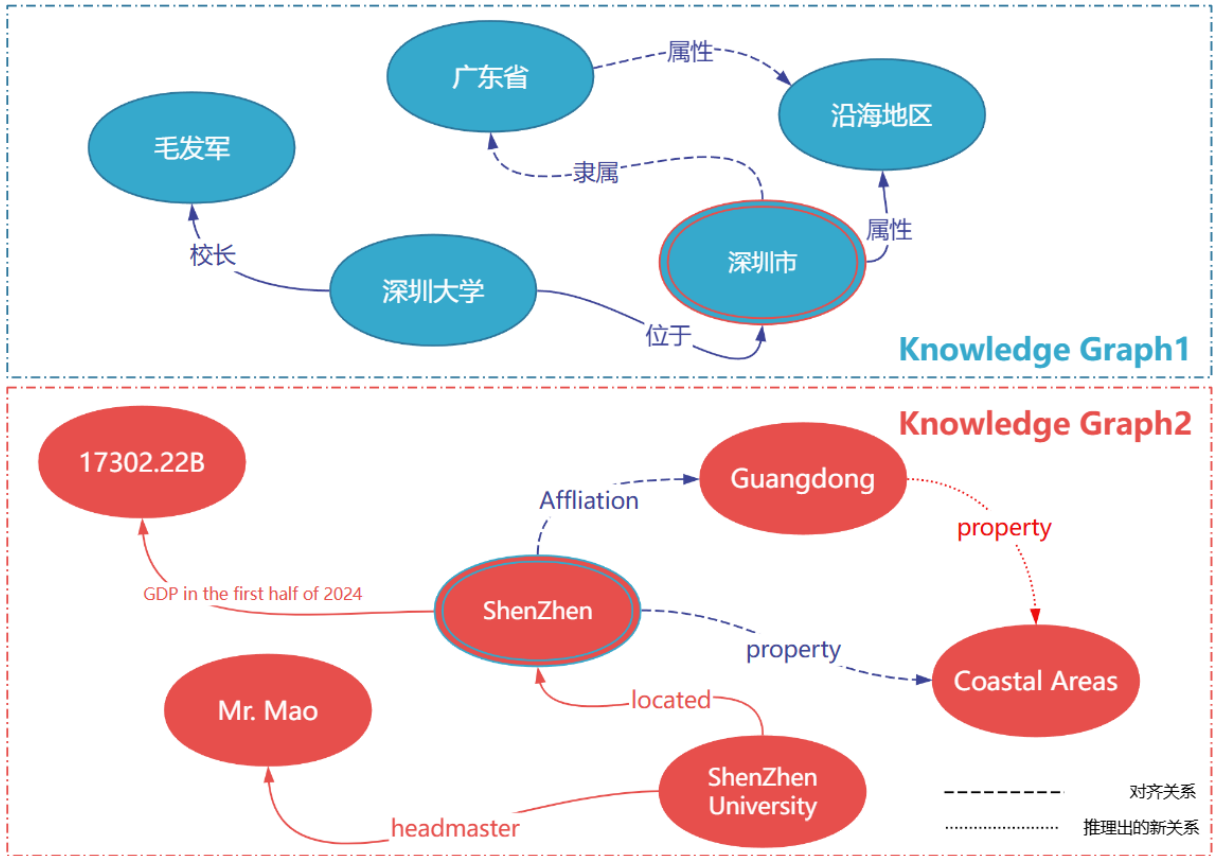


图 3. 规则推理示意图, 其虚线代表从 KG1 中推理的来的规则运用到 KG2, 双线的椭圆代表对齐的实体

4.1 规则推理前置知识

我们用三元组关系来定义规则，并且可以从知识图谱中推导出来，例如，以 $\forall x, y \in E : (x, r_s, y) \Rightarrow (x, r_c, y)$ 的形式，表示如果两个实体通过 r_s 相关联，那么它们也可能通过 r_c 相关联。箭头的左侧被定义为前提，右侧为结论。我们用 $k = (r_c | r_{s1}, \dots, r_{sp})$ 来表示规则，它包含一个或多个 $|p|$ 前提和仅一个结论，最后我们用 $\mathcal{K} = \{k\}$ 定义为规则的集合。一个规则必然会有一个或多个实例，对于规则 k ，用 $g(k) = (t_c | t_{s1}, \dots, t_{sp})$ 来表示它的一个实例，其中 t_c 表示的是结论三元组，而 t_{si} 表示的是前提三元组。这些三元组满足： $t_{s1} \wedge \dots \wedge t_{sp} \Rightarrow t_c$ ，其中 \wedge 是逻辑与，类似于“和”。其他组合包括逻辑或 \vee （类似于“或”）和逻辑非 \neg （类似于“非”）。例如，给定规则 $\text{bornIn}(x, y) \wedge \text{cityOf}(y, z) \Rightarrow \text{nationality}(x, z)$ ，我们在知识图谱中对其进行实例化，得到： $\text{bornIn}(\text{Obama}, \text{Hawaii}) \wedge \text{cityOf}(\text{Hawaii}, \text{United States}) \Rightarrow \text{nationality}(\text{Obama}, \text{United States})$ 。我们用 $G(k) = \{g(k)\}$ 来表示规则 k 的所有实例。

4.2 尝试一：扩充数据集

我们首先利用规则挖掘系统 AMIE+ [10]，从大规模知识图谱中高效地找到 Horn 规则，例如 $\text{marryTo}(x, y) \wedge \text{liveIn}(x, z) \Rightarrow \text{liveIn}(y, z)$ 。其源代码可在线获取¹。给定两个知识图谱 G 和 G' ，我们首先分别挖掘规则，并获得两组规则知识 \mathcal{K} 和 \mathcal{K}' 。这些规则知识由于知识图谱是为了满足不同应用或语言的需求而构建的，因此它们之间存在很大差异。虽然它们可以分别用于完善各自的知识图谱，但我们进一步通过 **知识不变性假设** 将这两组规则相互转换：知识具有普遍性，无论是在哪种语言或领域中。

4.2.1 规则迁移

给定对齐关系 \mathcal{A}_r^s 和一个规则 $k \in \mathcal{K}$ ，如果存在 $(r_c, r'_c), (r_{si}, r'_{si}) \in \mathcal{A}_r^s, i = 1, \dots, p$ ，我们将规则 $k = (r_c | r_{s1}, \dots, r_{sp})$ 中涉及的所有关系替换为其对应关系。因此，我们得到这样的规则 $k' = (r'_c | r'_{s1}, \dots, r'_{sp})$ ，并在 $k' \notin \mathcal{K}'$ 的情况下将其添加到 $\tilde{\mathcal{K}}' = \mathcal{K}' \cup k'$ 。可以在实验中找到转移规则的实际例子。注意，如果找不到对齐关系 $\mathcal{A}_r^s = \emptyset$ ，则可能没有转移规则。

4.2.2 图谱三元组的扩充

以知识图谱 G 为例，给定一个规则 $k \in \mathcal{K}$ ，我们收集其前提三元组可以在知识图谱中找到但结论三元组不在其中的实例： $\mathcal{G}(k) = \{g(k) | t_{s1}, \dots, t_{sp} \in T, t_c \notin T\}$ 。因此，我们将所有结论三元组添加到知识图谱中 $\tilde{G} = G \cup t_c, t_c \in \mathcal{G}(k)$ 。类似地，我们可以将知识图谱 G' 完善为 \tilde{G}' 。如图 3 所示，我们从信息丰富的 KG1 中获得规则 隶属 $(x, y) \wedge$ 属性 $(y, z) \Rightarrow$ 属性 (x, z) ，然后根据对齐关系 *province* 和 属性 将其转移到 KG1 中。因此，在 KG2 中，我们找到合适的三元组 $\text{Affiliation}(\text{ShenZhen}, \text{Guangdong}) \wedge \text{property}(\text{ShenZhen}, \text{Coastal Areas})$ ，因此获得新的三元组 $\text{property}(\text{Guangdong}, \text{Coastal Areas})$ 即 $(\text{Guangdong}, \text{property}, \text{Coastal Areas})$ 。

¹<https://www.mpi-inf.mpg.de/>

4.2.3 最后的训练

这里我们以法语和英文的数据集为例子,在上一步的推理扩充之后获得了新三元组 152,680 而原有的三元组数量为 221,720。最后我们将所获得的新三元组注入到数据集中,用于训练。这里我们单单的对数据集路扩充,没有对原方法做修改,对原方法的修改方法见 4.3。

4.3 尝试二: 引入规则推理损失函数

由于规则的正确不好评估,这里我们采用 t -范数模糊逻辑来计算规则的正确性 [12],该逻辑通过基于特定 t 范数的逻辑连接词将复杂公式的真值定义为其成分真值的组合。从前面章节的描述中我们知道规则也是若干个三元组组成,因此这里的损失函数也是通过三元组计算,相关的定义如下:以知识图谱 G 为例,遵循 [2],我们定义损失函数如下:

$$\begin{aligned} \mathcal{L}_r = & \sum_{g^+ \in \mathcal{G}(\mathcal{K})} \sum_{g^- \in \mathcal{G}^-(\mathcal{K})} [\gamma_r - I(g^+) + I(g^-)]_+ \\ & + \sum_{t^+ \in T} \sum_{t^- \in T^-} [\gamma_r - I(t^+) + I(t^-)]_+ \end{aligned} \quad (16)$$

其中 g 是规则具体化 $g(k)$ 的简写, $\mathcal{G}(\mathcal{K})$ 和 T 表示所有规则具体化和所有三元组。 $\mathcal{G}^-(\mathcal{K})$ 和 T^- 是通过最近邻采样 [21] 替换其中一个实体得到的负样本集。 $I(\cdot)$ 是三元组 t 的真值函数:

$$I(t) = 1 - \frac{1}{3\sqrt{d}} \|\mathbf{e}_i + \mathbf{r}_{ij} - \mathbf{e}_j\|_2 \quad (17)$$

或者对于具体化 $g = (t_c | t_{s1}, \dots, t_{sp})$, 它是递归计算的:

$$\begin{aligned} I(t_s) &= I(t_{s1} \wedge t_{s2}) = I(t_{s1}) \cdot I(t_{s2}) \\ I(t_s \Rightarrow t_c) &= I(t_s) \cdot I(t_c) - I(t_s) + 1 \end{aligned} \quad (18)$$

其中 d 是嵌入的大小。类似地,我们为知识图谱 G' 获得损失 \mathcal{L}'_r 。因此,总体损失函数为源方法损失 (15) 加上该损失,最终得到如下损失函数如下:

$$\mathcal{L} = \mathcal{L}_\mu + \mathcal{L}_{ICL} + \mathcal{L}_{LICL} + \mathcal{L}'_r + \mathcal{L}_r \quad (19)$$

5 实验设置与实验结果分析

除非另有说明,实验的所有超参数都是和原文一致,由于实验室设备限制,这里只跑了一个数据集即 DBP15K 中的 $fr-en$ 以及 $zh-en$, 而我的尝试实验都是在 $fr-en$ 一个数据集上进行,详细的介绍将会在 5.1 中介绍。

5.1 实验设置

5.1.1 数据集

我们的主要数据集为 DBP15K [22], 在原有数据集中包含三个从 DBpedia 多语言版本构建的数据集,包括 $DBP15K_{ZH-EN}$, $DBP15K_{JA-EN}$ 和 $DBP15K_{FR-EN}$ 。每个数据集包

含大约 400K 个三元组和 15K 对预对齐的实体对，而这里我们只用前两和作为我们的实验内容，其中选择部分作为种子对齐，选择的数量占中数量的比例为 $R_{sa} \in \{0.2, 0.5, 0.8\}$ 。其中实体附有图像，对于没有关联图像的实体，我们使用正态分布生成随机向量作为它们的视觉特征，该正态分布由其他可用图像的均值和标准差参数化。

5.1.2 实现细节

(1) 所有网络的隐藏层维度为 300。总的训练周期设置为 500，可选的迭代训练策略再应用 500 周期，遵循 [16]。采用了包括余弦预热计划（15% 步骤用于学习率预热）、早停和梯度累积在内的训练策略。使用 AdamW 优化器 ($\beta_1 = 0.9$, $\beta_2 = 0.999$)，固定批次大小为 3500。(2) 为了展示模型的稳定性，遵循 [3, 16]，在 DBP15K 上将视觉编码器 Enc_v 设置为 ResNet-152 [13] (3) 遵循 [29]，选择词袋 (BoW) 对关系 (x^r) 和属性 (x^a) 进行编码，作为固定长度（例如， $d_r = 1000$ ）的向量。我们使用预训练的 300-d GloVe 向量以及字符二元组进行表面表示，在应用机器翻译处理实体名称后，如 [18] 所述。总体来说我的所有实验设置都是和原文一样，实验都是在 Intel(R) Xeon(R) W-2225 CPU @ 4.10GHZ 上进行。

5.2 实验结果

5.2.1 尝试一：实验结果分析

我们的所有实验设置均与原文一致即没用表面信息、迭代学习以及难样本重放策略，在实验后，发现我们的简单复现基本和原文一致，而我们的两个尝试中差别比较大。其中尝试一为仅仅扩充数据集中的关系三元组，且都为规则推理得来详情见 4.2，从实验的结果中 1 我们可以看到该尝试的结果相比原论文减少了 7 个百分点，经过初步的分析我们认为：

- 新推出来的三元组都是在原有的实体上的，并没有新增任何的实体，这就会使得在计算后的嵌入关系过于复杂，最终难以区分。
- 不同的知识图谱之间的实体对并不会完全一致，这样就会导致在一个图谱中确实的实体引起的关系缺少引入一定的噪声。

以上的两点问题我们初步的解决方案为，用跟完善的推理体系，将所有的关系都推理出来，并且寻找或者研发出新的方法来补全缺失的实体，或者删除奇异值的实体，最终再用规则的损失函数，当然这留在我们的未来工作当中。

表 1. 浮现实验结果对比表，加粗代表最好的结果。

Models	三元组数量	$FBDB15K_{fr-en}$		
		H@1	H@10	MRR
原方法	221720	.770	.961	.841
复现版	221720	.769	.962	.839
原方法 + 扩充数据集	374400	.711	.951	.810
原方法 + 规则损失函数	221720	.768	.959	.837

5.2.2 尝试二：实验结果分析

在经过表1中的结果对比，我们发现我们的尝试二比尝试一有很大的提升，但是与原文相比还是减少一定的百分比，经过分析，我们初步认为原因和尝试一类似但是又有不同，相同之处在于都是关系影响了最终的实体嵌入，让模型很难学习到一个较好的相对距离。而不同之处在于尝试一是通过直接增加了关系数量来影响实验结果，而尝试二则是通过用损失函数将规则纳入，在最后的对比学习上引入噪声，即规则的损失会将关系不对称的实体的嵌入拉开即便他们是应该对齐，从而最终影响了最终的结果。

6 总结与展望

本次的复现工作虽然没有对原方法有一个质的提升，但是做了很多的尝试。通过简单的复现到发现新的方法，再到实验，从中获得很多。通过本次复现不仅让自己代码功底加深，还让自己更加理解该方向的基本原理。未来的工作也会继续解答在第5节中提出的问题以及进行相应的实验，即利用的规则主要基于关系三元组，但是由于图谱之间的实体并不是一一对应，因此引入的规则或者是新增三元组会在最后的实体嵌入学习中引入噪声，即把原本应该对齐的实体嵌入距离拉开，导致结果下降，而未来的解决方向可以是可以对实体的关系进行裁剪或者新增预测实体。此外，视觉模态的缺失是一个普遍的挑战。同时它对结果的影响较大。因此这也可以是未来的工作方向。

参考文献

- [1] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.
- [2] Yixin Cao, Zhiyuan Liu, Chengjiang Li, Juanzi Li, and Tat-Seng Chua. Multi-channel graph neural network for entity alignment. *arXiv preprint arXiv:1908.09898*, 2019.
- [3] Liyi Chen, Zhi Li, Yijun Wang, Tong Xu, Zhefeng Wang, and Enhong Chen. MMEA: entity alignment for multi-modal knowledge graph. In *KSEM (1)*, volume 12274 of *Lecture Notes in Computer Science*, pages 134–147. Springer, 2020.
- [4] Liyi Chen, Zhi Li, Tong Xu, Han Wu, Zhefeng Wang, Nicholas Jing Yuan, and Enhong Chen. Multi-modal siamese network for entity alignment. In *KDD*, pages 118–126. ACM, 2022.
- [5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 2020.
- [6] Zhuo Chen, Jiaoyan Chen, Yuxia Geng, Jeff Z. Pan, Zonggang Yuan, and Huajun Chen. Zero-shot visual question answering using knowledge graph. In *ISWC*, volume 12922 of *Lecture Notes in Computer Science*, pages 146–162. Springer, 2021.

- [7] Zhuo Chen, Yufeng Huang, Jiaoyan Chen, Yuxia Geng, Yin Fang, Jeff Z. Pan, Ningyu Zhang, and Wen Zhang. Lako: Knowledge-driven visual question answering via late knowledge-to-text injection. In *IJCKG*, pages 20–29. ACM, 2022.
- [8] Yin Fang, Qiang Zhang, Haihong Yang, Xiang Zhuang, Shumin Deng, Wen Zhang, Ming Qin, Zhuo Chen, Xiaohui Fan, and Huajun Chen. Molecular contrastive learning with chemical element knowledge graph. In *AAAI*, pages 3968–3976. AAAI Press, 2022.
- [9] Yin Fang, Qiang Zhang, Ningyu Zhang, Zhuo Chen, Xiang Zhuang, Xin Shao, Xiaohui Fan, and Huajun Chen. Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nature Machine Intelligence*, pages 1–12, 2023.
- [10] Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M Suchanek. Fast rule mining in ontological knowledge bases with amie +. *The VLDB Journal*, 24(6):707–730, 2015.
- [11] Shu Guo, Quan Wang, Bin Wang, Lihong Wang, and Li Guo. Semantically smooth knowledge graph embedding. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 84–94, 2015.
- [12] Petr Hájek. *Metamathematics of fuzzy logic*, volume 4. Springer Science & Business Media, 2013.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778. IEEE Computer Society, 2016.
- [14] Ernesto Jiménez-Ruiz and Bernardo Cuenca Grau. Logmap: Logic-based and scalable ontology matching. In *ISWC (1)*, volume 7031 of *Lecture Notes in Computer Science*, pages 273–288. Springer, 2011.
- [15] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR (Poster)*. OpenReview.net, 2017.
- [16] Zhenxi Lin, Ziheng Zhang, Meng Wang, Yinghui Shi, Xian Wu, and Yefeng Zheng. Multi-modal contrastive representation learning for entity alignment. In *COLING*, pages 2572–2584. International Committee on Computational Linguistics, 2022.
- [17] Fangyu Liu, Muhao Chen, Dan Roth, and Nigel Collier. Visual pivoting for (unsupervised) entity alignment. In *AAAI*, pages 4257–4266. AAAI Press, 2021.
- [18] Xin Mao, Wenting Wang, Yuanbin Wu, and Man Lan. From alignment to assignment: Frustratingly simple unsupervised entity alignment. In *EMNLP (1)*, pages 2843–2853. Association for Computational Linguistics, 2021.
- [19] Matthew Richardson and Pedro Domingos. Markov logic networks. *Machine learning*, 62:107–136, 2006.

- [20] Fabian M. Suchanek, Serge Abiteboul, and Pierre Senellart. PARIS: probabilistic alignment of relations, instances, and schema. *Proc. VLDB Endow.*, 5(3):157–168, 2011.
- [21] Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21–25, 2017, Proceedings, Part I 16*, pages 628–644. Springer, 2017.
- [22] Zequn Sun, Wei Hu, and Chengkai Li. Cross-lingual entity alignment via joint attribute-preserving embedding. In *ISWC (1)*, volume 10587 of *Lecture Notes in Computer Science*, pages 628–644. Springer, 2017.
- [23] Zequn Sun, Wei Hu, Chengming Wang, Yuxin Wang, and Yuzhong Qu. Revisiting embedding-based entity alignment: A robust and adaptive method. *IEEE Trans. Knowl. Data Eng.*, 35(8):8461–8475, 2023.
- [24] Zequn Sun, Wei Hu, Qingheng Zhang, and Yuzhong Qu. Bootstrapping entity alignment with knowledge graph embedding. In *IJCAI*, pages 4396–4402, 2018.
- [25] Bayu Distiawan Trisedya, Jianzhong Qi, and Rui Zhang. Entity alignment between knowledge graphs using attribute embeddings. In *AAAI*, pages 297–304. AAAI Press, 2019.
- [26] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, pages 5998–6008, 2017.
- [27] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *ICLR (Poster)*. OpenReview.net, 2018.
- [28] Yuejia Xiang, Ziheng Zhang, Jiaoyan Chen, Xi Chen, Zhenxi Lin, and Yefeng Zheng. Ontoea: Ontology-guided entity alignment via joint knowledge graph embedding. In *ACL/IJCNLP (Findings)*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 1117–1128. Association for Computational Linguistics, 2021.
- [29] Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. Aligning cross-lingual entities with multi-aspect information. In *EMNLP/IJCNLP (1)*, pages 4430–4440. Association for Computational Linguistics, 2019.
- [30] Weixin Zeng, Xiang Zhao, Jiuyang Tang, and Xuemin Lin. Collective entity alignment via adaptive features. In *ICDE*, pages 1870–1873. IEEE, 2020.
- [31] Rui Zhang, Bayu Distiawan Trisedya, Miao Li, Yong Jiang, and Jianzhong Qi. A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning. *VLDB J.*, 31(5):1143–1168, 2022.

- [32] Daquan Zhou, Bingyi Kang, Xiaojie Jin, Linjie Yang, Xiao Chen Lian, Qibin Hou, and Jiashi Feng. Deepvit: Towards deeper vision transformer. *CoRR*, abs/2103.11886, 2021.
- [33] Hao Zhu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Iterative entity alignment via joint knowledge embeddings. In *IJCAI*, pages 4258–4264. ijcai.org, 2017.