

基于邻域匹配概率与类型商图的实体对齐解释方法

张晓明, 邱菁菁, 王会勇

引用本文

张晓明, 邱菁菁, 王会勇. 基于邻域匹配概率与类型商图的实体对齐解释方法[J]. 计算机科学, 2025, 52(12): 260-270.

ZHANG Xiaoming, QIU Jingjing, WANG Huiyong. [Explanation Method for Entity Alignment Based on Neighborhood Matching Probability and Type Quotient Graph](#) [J]. Computer Science, 2025, 52(12): 260-270.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

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

[基于变体注意力的关系与属性感知实体对齐](#)

Relationship and Attribute Aware Entity Alignment Based on Variant-attention
计算机科学, 2025, 52(11): 230-236. <https://doi.org/10.11896/jsjcx.240800140>

[可解释的信用风险评估模型:基于注意力机制的规则提取方法](#)

Interpretable Credit Risk Assessment Model:Rule Extraction Approach Based on AttentionMechanism
计算机科学, 2025, 52(10): 50-59. <https://doi.org/10.11896/jsjcx.250300059>

[多尺度多粒度解耦蒸馏模糊分类器及其在癫痫脑电信号检测中的应用](#)

Multi-scale Multi-granularity Decoupled Distillation Fuzzy Classifier and Its Application inEpileptic EEG Signal Detection
计算机科学, 2025, 52(9): 37-46. <https://doi.org/10.11896/jsjcx.250300096>

[基于局部梯度平滑的解释鲁棒性对抗训练方法](#)

Explanation Robustness Adversarial Training Method Based on Local Gradient Smoothing
计算机科学, 2025, 52(2): 374-379. <https://doi.org/10.11896/jsjcx.240400210>

[基于特征加权的反事实解释方法:以信贷风控场景为例](#)

Feature-weighted Counterfactual Explanation Method:A Case Study in Credit Risk Control Scenarios
计算机科学, 2024, 51(12): 259-268. <https://doi.org/10.11896/jsjcx.240300047>

基于邻域匹配概率与类型商图的实体对齐解释方法

张晓明 邱菁菁 王会勇

河北科技大学信息科学与工程学院 石家庄 050018

(zhangxiaom@hebust.edu.cn)

摘要 近年来,出现了多种基于嵌入的实体对齐方法,此类方法通过将实体和关系映射到低维向量空间,并计算这些向量表示来实现实体对齐。尽管这些方法都取得了很好的性能,但对其可解释性的研究相对较少。因此,提出了一种事后解释的实体对齐算法 PE-EA,为基于知识图嵌入的实体对齐模型的预测结果生成解释。该方法首先通过计算知识图谱中实体及其关系的连接数来评估其功能性,进而量化实体邻域结构的重要性。之后,结合实体的功能性与实体的嵌入向量计算关系的嵌入向量,据此获取关系对的匹配概率。然后,根据模型预测实体对的邻域信息,计算邻域中候选解释对的匹配概率,筛选出预测实体对的解释三元组,将它们组合成解释子图。最后,引入类型商图概念,将解释子图抽象化,压缩数据并简化解释生成过程,从而在减少候选解释数量的同时,提升解释的质量和有效性。在 5 个常用的实验数据集上,使用 fidelity 和 sparsity 两种评价指标验证了模型生成的解释有较高的准确性和简洁性。

关键词: 实体对齐; 可解释性; 邻域结构; 匹配概率; 类型商图

中图分类号 TP391

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

ZHANG Xiaoming, QIU Jingjing and WANG Huiyong

College of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

Abstract In recent years, a variety of embedding-based entity alignment methods have emerged, which achieve entity alignment by mapping entities and relations into low-dimensional vector spaces and calculating these vector representations. Although these methods have achieved good performance, there are relatively few studies on their explainability. Therefore, a post hoc explanation algorithm for entity alignment (PE-EA) is proposed to generate explanations for the prediction results of the entity alignment model based on knowledge graph embedding. The method first evaluates the functionality of entities and their relations by calculating their connection counts in the knowledge graph, and then quantifies the importance of the entity neighborhood structure. Subsequently, the entity functionality is combined with the embedding vector to calculate the embedding vector of the relation, and the matching probability of the relation pair is obtained based on it. Then, based on the neighborhood information of the predicted entity pair, the matching probability of the candidate explanation pairs in the neighborhood is calculated, and the explanation triplets of the predicted entity pair are screened out and combined into an explanation subgraph. Furthermore, the concept of type quotient graph is introduced to abstract the explanatory subgraph, compress the data and simplify the explanation generation process, thereby reducing the number of candidate explanations while enhancing the quality and effectiveness of the explanations. On five commonly used experimental datasets, the two evaluation indicators of fidelity and sparsity are used to verify that the explanations generated by the proposed model have high accuracy and simplicity.

Keywords Entity alignment, Explainability, Neighborhood structure, Probabilistic matching, Type quotient graph

1 引言

不同的图谱中识别并匹配指向现实世界中相同事物的实体^[1]。面对知识图谱的复杂拓扑结构,捕捉实体间丰富的语义关系是实现有效对齐的关键,而知识图嵌入^[2]被广泛认

实体对齐作为知识图谱融合中的一项关键技术,能够在

到稿日期:2024-11-13 返修日期:2025-03-04

基金项目:河北省自然科学基金(F2022208002);石家庄市基础研究计划(241790867A)

This work was supported by the Natural Science Foundation of Hebei Province, China (F2022208002) and Shijiazhuang Basic Research Program (241790867A).

通信作者:王会勇(wanghuiyong815@163.com)

为是处理该问题的有效方法。知识图嵌入是一种将知识图谱中的实体和关系映射到低维向量空间的方法,这些向量能够捕捉实体和关系的语义信息,同时保留图谱的结构信息。利用知识图嵌入技术进行实体对齐,是当前研究中的一个热点主题,每年都有数十种新模型被提出^[3]。当前的实体对齐解决方案大多建立在这样一个假设之上:相似的实体通常拥有相似的邻域结构。因此,这些模型利用知识图谱中的邻域结构特征信息来生成向量表示,从而实现实体之间的对齐。然而,与大多数深度学习系统一样,这种实体对齐模型并不是直接可解释的。

可解释性对于模型至关重要,它可以使模型的预测过程透明化,允许用户洞察预测背后的逻辑。近期,出现了许多在图像数据^[4]、表格数据^[5]和文本数据^[6]的深度模型中提取解释的尝试,这些框架旨在识别出对预测结果影响最大的输入特征。然而,与图像、表格和文本数据相比,知识图谱中图结构数据的复杂性和多维性为可解释性研究带来了额外的挑战。

在图卷积网络中^[7],图分类任务的可解释性研究主要集中在评估图中的节点或边对模型预测结果的贡献度上,常见的方法如采用注意力机制引入注意力权重^[8]、逐个或成组地移除图中的特征并观察模型性能的变化^[9]。而实体对齐往往涉及异构知识图谱和大规模的实体组合,图分类任务的解释方法并不适用。此外,在知识图谱补全中,解释性研究主要是通过一种基于扰动的方法^[10]来识别对预测结果影响最大的子图。该方法也不能直接转移到实体对齐任务中,因为实体对齐任务关注两个实体的结构信息,而补全通常只涉及一个知识图谱。

基于知识图嵌入的实体对齐模型虽然缺乏解释性,但可以利用模型的预测结果和实体的嵌入来获得特定上下文的解释。如图1所示,图1(a)中的实体对 (u, v) 代表由实体对齐模型生成的预测实体对,图1(b)中加粗显示的三元组表示支持这一对齐决策的关键证据。

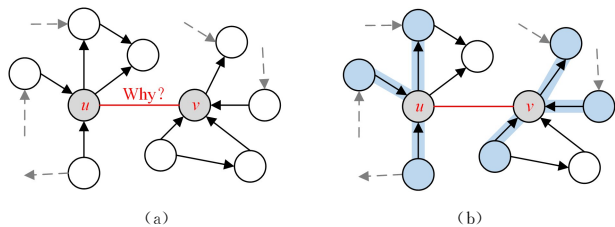


图1 实体对齐解释示例图

Fig. 1 Entity alignment explanation diagram

解释基于嵌入的实体对齐结果是一项具有挑战性的任务。首先是因为嵌入模型本身的复杂性,许多模型在进行训练时往往会使用高阶三元组和大量参数,这使得评估单个三元组对预测结果的具体贡献变得困难。其次是现有的基于图卷积网络和知识图谱补全的解释方法并不能直接适用于实体对齐任务。此外,有些实体对的生成依赖大量的三元组信息组合,会导致解释的表达不够直观。

为了应对上述挑战,本文提出了名为PE-EA的事后可解释模型。该框架同样基于:如果两个实体是相似的,那么它们周围的邻居节点在结构上也是相似的规律,在两个知识图谱

上识别并展示两个相似实体及其邻域之间的相似子图,以此来提供对模型预测的解释。同时,PE-EA模型不直接依赖实体对齐模型生成的嵌入向量来计算相似性,而是利用知识图谱结构评估实体和关系的重要性,计算邻域匹配对的概率得分,进而识别出对预测有关键影响的关系和实体对,全面地解释了模型的预测。

本文的主要贡献如下:

1) 设计了一种基于匹配概率的解释生成方法。从实体对齐模型给出的预测实体对的邻域中筛选出候选解释实体对,组合成对齐子图。计算对齐子图中关系对和候选解释实体对的匹配概率,从而确定能够支持预测结果的关键三元组,这些三元组组成解释子图。

2) 引入实体类型信息来简化和压缩解释子图。解释子图中的实体根据实体类型标签划分为一个个等价类,并以此构建类型商图,提高了解释的质量和效率。

3) 实验结果显示,PE-EA在多个知识图谱数据集上都能够很好地找到有效的解释关系三元组。同时,与基准模型相比,其在保持相似的稀疏度的前提下,还能得到更高的忠实度。

2 相关工作

2.1 基于嵌入的实体对齐

基于嵌入的方法因能有效捕捉实体在向量空间中的特征信息,已经被广泛应用于实体对齐任务。基于嵌入的实体对齐主要分为两大类:基于翻译的方法和基于图神经网络的方法。

基于翻译的实体对齐的核心思想是将跨知识图谱的实体匹配问题转换为寻找和优化实体在嵌入空间中的向量表示,以实现向量间的最优化对应关系,如MtransE^[11],BootEA^[12]等。基于图卷积网络的方法利用图卷积网络捕捉和建模知识图谱中实体的局部和全局结构特征,如GCN-Align^[13],Dual-AMN^[14]等。尽管这些知识图谱嵌入模型在预测对齐方面取得了成功,但它们在透明度和可追溯性方面的支持程度限制了研究人员开发可信模型的能力。

本文选择的实体对齐模型包含上述两类,并重点关注通过关系三元组的知识图谱结构进行学习的实体对齐模型。

2.2 实体对齐的可解释模型

为了增强实体对齐任务中模型的透明度,许多研究人员开发了各种可解释的模型。一类方法是通过简化模型结构,并运用经典的组合优化算法实现对齐。例如:SEU^[15]和MD-JEA^[16]将跨语言实体对齐问题转换为最小化成本问题,利用Sinkhorn算法高效计算实体相似度;LightEA^[17]采用非神经网络的框架和经典的标签传播算法实现对齐。这些方法在减少计算成本的同时提高了对齐过程的透明度。另一类方法是结合逻辑规则或注意力权重来实现对齐,进而提高模型的可解释性。例如:ASGEA^[18]利用ASGNN识别逻辑规则;P-NAL^[19]则采用非公理逻辑来形式化和执行实体对齐的逻辑推理;i-align^[20]模型通过分析实体对的属性和邻居的注意力权重来提供对齐解释。这些方法虽然能够为特定类型的模型

提供解释,但它们往往缺乏通用性,不能广泛适用于多种不同的模型。

2.3 事后可解释性模型

虽然基于实体对齐预测结果的可解释性未得到广泛的研究,但图卷积网络和知识图谱补全的事后解释性研究工作仍具有重要的参考意义。其中,图卷积网络普遍采用的策略是挑选与预测结果密切相关的子图,例如 GNNExplainer^[21]和 PGExplainer^[22]运用原始预测与掩码图之间的互信息作为衡量关联性的指标,并通过优化图的节点和边缘特征上的掩码来识别边缘引导的子图。SubgraphX^[23]通过结合蒙特卡罗树搜索算法和 Shapley 值量子图重要性,为图神经网络的预测提供了直观的子图级解释。知识图谱补全中最具代表性的模型包括:1)kelpie 模型^[10],其获取对齐模型预测结果的最小三元组子集作为解释,且能提供“必要解释”和“充分解释”两种互补的解释类型;2)CRIAGE^[24],其通过对抗性地添加或删除知识图中的事实来探究链接预测模型的鲁棒性和可解释性;3)KE-X^[25],其基于知识信息增益理论,通过信息熵量化解释候选项的重要性,并生成可解释的子图。这些方法都致力于识别和突出对模型预测最为关键的图结构部分。

进一步地,LIME^[26]和 SHAP^[27]是两种流行且通用的可解释人工智能框架,它们为解释模型预测提供了不同的视角。LIME 通过对预测样本的输入特征进行微调,观察这些变化对预测置信度的影响,并为每个特征分配一个相关权重,从而识别出导致特定预测结果的关键特征。SHAP 框架利用 Shapley 值,这是来自博弈论的一个概念,用于量化输入特征对模型预测结果的贡献。实体对齐是一个复杂的过程,它通常涉及多个特征信息的组合使用。ExEA^[28]模型通过分析和匹配相邻实体及其关系路径,构建出能够解释实体对齐预测的特定事实组合,为实体对齐的解释提供了新的视角。

本研究借鉴 ExEA 的解释理念,将生成实体对齐解释视为构建语义匹配子图,通过对比两个实体邻域的三元组进行匹配。与 ExEA 直接依赖实体和关系的嵌入向量不同,本文通过比较中心实体的邻域子图,对嵌入相似性进行了补充,提供了更全面的对齐证据。此外,还深入探讨了类型商图是如何在基于实体类型的解释中发挥作用的,与现有的一些解释方法相比,本文方法可以提取更简洁且高效的解释,即在保留更少的解释三元组的情况下获得更好的解释结果。

3 基本概念及问题定义

3.1 基本概念

知识图谱 $KG=(E,R,L,T)$ 是一个有向图,其中 E,R,L 分别表示实体集、关系集和类型标签集, $T\subseteq E\times R\times E$ 是关系三元组的集合。每个三元组可以表示为 (h,r,t) ,其中 $h\in E$ 和 $t\in E$ 分别表示头部实体和尾部实体, $r\in R$ 表示关系。下面给出本文中将会提到的一些相关定义。

定义 1(基于嵌入的实体对齐) 给定源知识图谱 $KG1$ 和目标知识图谱 $KG2$,实体对齐的任务是找出实体集 $E1$ 和

$E2$ 之间的匹配实体对 $M=\{(e^1,e^2)\in E1\times E2 \mid e^1\equiv e^2, E1\in KG1, E2\in KG2\}$,其中 \equiv 表示“owl:sameAs”关系。

基于嵌入的实体对齐模型通过学习实体的低维向量表示 h_{emb} 和 t_{emb} ,使得属于同一实体的向量表示在向量空间中彼此接近,而属于不同实体的向量表示则相互疏远。通常情况下,模型会提供一组种子实体对 M_{train} 作为训练数据,然后通过计算不同图谱中实体嵌入之间的相似度找出剩余的实体对 M_{res} 。

定义 2(功能性) 功能性指的是一个关系或实体在特定上下文中所具有的唯一性或排他性的程度^[29],采用功能性可以帮助确定哪些关系和实体对识别和对齐实体最为关键。关系功能性通常是指一个关系在所有三元组中出现的次数,例如,关系出生地就比关系居住地重要得多,因为一个人的一生只能有一个出生地,但可以拥有很多的居住地。对于给定的关系 $r\in R$,它的功能性 $func(r)$ 可以通过式(1)计算:

$$func(r) = \frac{|\{h \mid (h,r,t) \in T\}|}{|\{(h,t) \mid (h,r,t) \in T\}|} \quad (1)$$

其中, T 是知识图谱中的三元组集合, h 和 t 分别表示连接关系 r 的头、尾实体, $|\cdot|$ 表示集合的大小。考虑到关系具有方向性,逆关系的功能性定义如式(2)所示:

$$ifunc(r) = \frac{|\{t \mid (h,r,t) \in T\}|}{|\{(h,t) \mid (h,r,t) \in T\}|} \quad (2)$$

实体功能性可以通过计算一个实体作为关系三元组中头实体或尾实体出现的次数来得出。例如,实体身份证号码可以唯一确定人们的身份,而实体女生则代表很多人,因此实体身份证号码的功能性比实体女生的功能性高。具体来说,对于一个实体 e ,作为头实体的功能性 $func(e)$ 和作为尾实体的逆功能性 $ifunc(e)$ 可以表示为式(3)和式(4):

$$func(e) = \frac{1}{|\{(e,r,t) \mid (e,r,t) \in T\}|} \quad (3)$$

$$ifunc(e) = \frac{1}{|\{(h,r,e) \mid (h,r,e) \in T\}|} \quad (4)$$

这些功能性都能够预先确定,因为它们仅依赖知识图谱的连接性统计数据。

定义 3(对齐子图) 预测实体对 (e^1,e^2) 的对齐子图 $ASG_{(e^1,e^2)}=(M_{(e^1,e^2)},R_{(e^1,e^2)},T_{(e^1,e^2)})$ 是基于 M_{train} 和 M_{res} 构建的子图形成的。

首先,从 M_{train} 和 M_{res} 中获取预测实体对 (e^1,e^2) 的候选解释对集 $M_{(e^1,e^2)}=\{(e_m^1,e_n^2) \mid e_m^1\in N^1, e_n^2\in N^2, (e_m^1,e_n^2)\in M_{\text{train}} \vee (e_m^1,e_n^2)\in M_{\text{res}}\}$,其中 $(e_m^1,e_n^2)\in M_{(e^1,e^2)}$ 是预测实体对 (e^1,e^2) 的一个候选解释对, N^1 和 N^2 分别表示实体 e^1 和 e^2 在各自知识图谱中的一阶邻域实体集。然后提取每对候选解释对 (e_m^1,e_n^2) 与预测实体对 (e^1,e^2) 的候选三元组集 $T_{(e^1,e^2)}=\{(e^1,r^1,e_m^1)\cup(e_m^1,r^1,e^1)\cup(e^2,r^2,e_n^2)\cup(e_n^2,r^2,e^2) \mid (e^1,r^1,e_m^1)\in T1 \vee (e_m^1,r^1,e^1)\in T1, (e^2,r^2,e_n^2)\in T2 \vee (e_n^2,r^2,e^2)\in T2, (e_m^1,e_n^2)\in M_{(e^1,e^2)}, T1\in KG1, T2\in KG2\}$ 。最后从候选三元组集 $T_{(e^1,e^2)}$ 中获取候选解释关系集 $R_{(e^1,e^2)}=\{r^1\cup r^2 \mid (e^1,r^1,e_m^1)\in T_{(e^1,e^2)}, (e^2,r^2,e_n^2)\in T_{(e^1,e^2)}\}$ 构成 $ASG_{(e^1,e^2)}$ 。

3.2 问题定义

在实体对齐解释任务中,本文的目标是提供对实体对齐

模型预测结果的事后解释。事后解释意味着模型已经被训练且固定,解释方法不会修改其架构或参数。对于模型预测的实体对,本文不评估预测对的正确性或错误性,而是提供模型做出此预测的依据,即解释性证据,并将解释的输出形式表达为由关系三元组组合而成的子图。基于嵌入的实体对齐的解释定义如下:

定义4(实体对齐模型的事后解释) 给定模型预测的实体对 (e^1, e^2) ,匹配 e^1 和 e^2 邻域的三元组组合成的子图作为解释,如图2所示。首先,选取 e^1 和 e^2 各自 h 跳范围内的三元组,为了避免冗余且不相关信息, h 通常不超过2。这些三元组的并集构成候选集 $T^* = T^{e^1} \cup T^{e^2}$,通过计算候选集中关系对和候选解释实体对的匹配概率,从中找到一个最小子集 $T^{\text{exp}} \subseteq T^*$ 。即便从知识图谱中移除其他三元组 $T^{\text{del}} = T^* - T^{\text{exp}}$,模型仍能维持对 (e^1, e^2) 的预测,所选三元组 T^{exp} 形成的子图表示为解释子图 $KG^{\text{exp}}(E^{\text{exp}}, R^{\text{exp}}, T^{\text{exp}})$,其中, E^{exp} 和 R^{exp} 表示解释子图中的实体集和关系集。

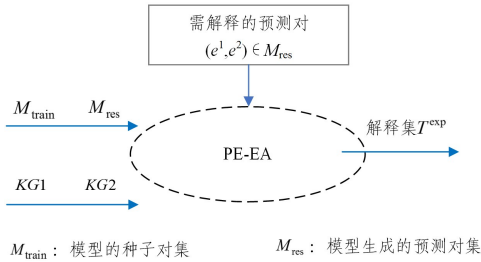


图2 问题定义图

Fig. 2 Problem definition diagram

4 本文方法

4.1 方法概述

在解释对齐模型的预测实体对时,对于每个实体,识别其在知识图谱中的邻居实体,并检查这些邻居实体是否在另一个知识图谱中有对应的对齐实体。将这些在不同图谱中有对齐的邻居实体及其关系组合为预测实体对的对齐子图。本文通过计算对齐子图中关系和实体的匹配概率,识别出支持模型进行预测的关系三元组,并依据这些三元组及其对应的概率匹配得分构建解释子图。实现方法如图3所示,PE-EA模型由3个模块组成:1)关系对的匹配概率;2)候选解释对的匹配概率;3)解释子图转换为类型商图。具体步骤如下:

步骤1 计算邻域关系对的匹配概率。利用知识图谱中每个关系所连接的头尾实体及其功能性来近似生成关系的嵌入表示,并计算关系之间的相似性,从而获取关系对的匹配概率得分。

步骤2 获取邻域候选解释对的匹配概率。基于步骤1中的关系对匹配概率和预测实体对的对齐子图获取候选解释对的匹配概率。选取匹配概率高的关系对和候选解释对组成预测实体对的解释三元组,这些三元组组成解释子图。

步骤3 采用实体类型对解释子图进行精炼。在解释子图构建完成后,提取实体的类型信息,并将具有相同类型的实体聚合为单个节点,进而基于这些节点信息构建类型商图。

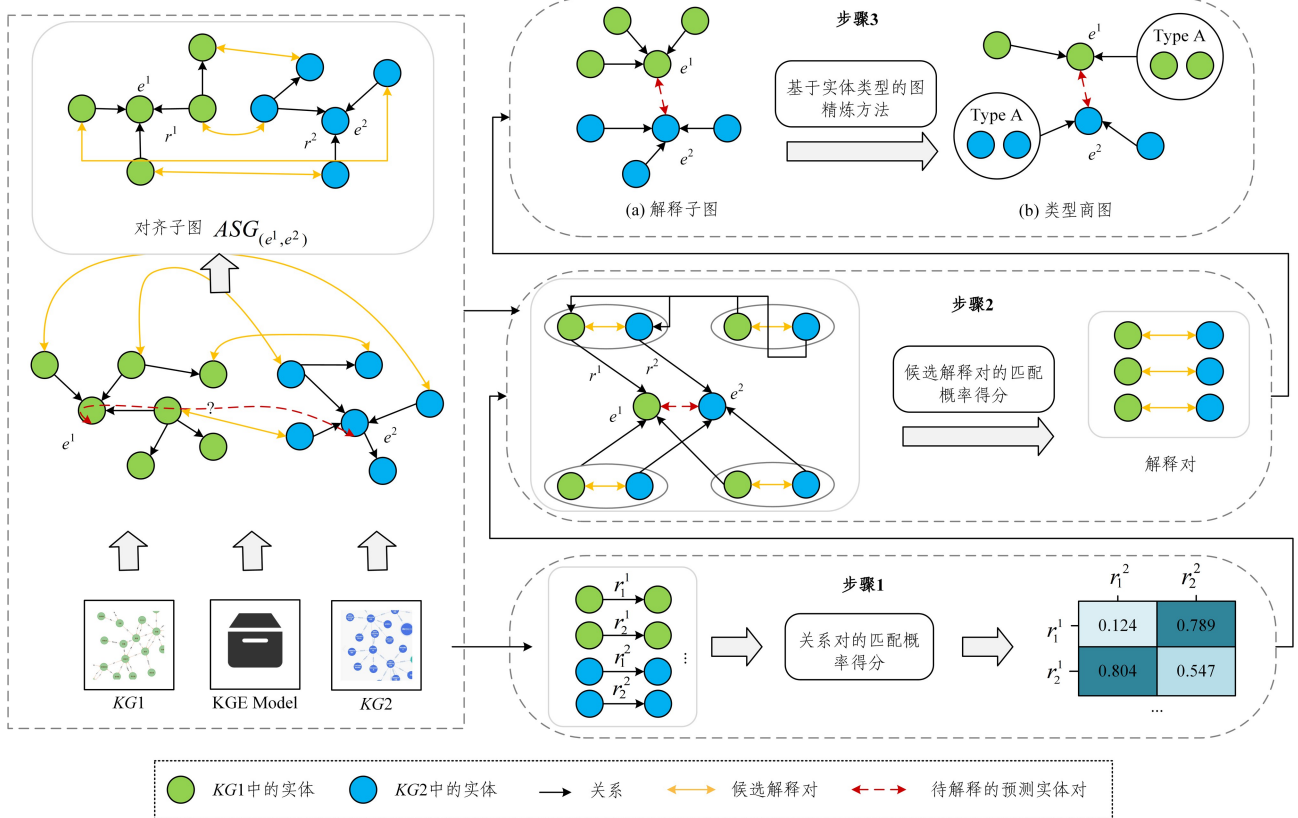


图3 PE-EA模型

Fig. 3 PE-EA model

4.2 邻域关系对的匹配概率

为了捕捉知识图谱中的关系语义,通常将关系所连接的多个头尾实体的嵌入向量结合起来,以近似表示关系本身的语义。

从 KG1 和 KG2 中依次选取关系对 (r^1, r^2) 计算其对齐概率。首先收集与这对关系相连的头实体集和尾实体集的嵌入向量。接着,结合实体功能性和实体的嵌入向量,通过加权求和的方式聚合每个关系 r 连接的头实体集和尾实体集,得到聚合的关系嵌入向量 $\mathbf{r}_{H_{emb}}$ 和 $\mathbf{r}_{T_{emb}}$,如式(5)和式(6)所示:

$$\mathbf{r}_{H_{emb}} = \sum_{h \in E_h^r} (func(h) \times \mathbf{h}_{emb}) \quad (5)$$

$$\mathbf{r}_{T_{emb}} = \sum_{t \in E_t^r} (ifunc(t) \times \mathbf{t}_{emb}) \quad (6)$$

其中, E_h^r 表示关系为 r 的头实体集; E_t^r 表示关系为 r 的尾实体集; \mathbf{h}_{emb} 和 \mathbf{t}_{emb} 表示经过一个实体对齐模型训练后得到的实体嵌入向量。

最后,计算聚合后的头实体集和尾实体集的关系向量之间的相似度,获得关系对的匹配概率得分 $match(r^1, r^2)$,如式(7)所示:

$$match(r^1, r^2) = \frac{sim(\mathbf{r}_{H_{emb}}^1, \mathbf{r}_{H_{emb}}^2) + sim(\mathbf{r}_{T_{emb}}^1, \mathbf{r}_{T_{emb}}^2)}{2} \quad (7)$$

其中, sim 表示用于计算实体向量间相似性的方法,由于不同模型可能采用不同的相似度计算策略,因此此处的 sim 作为通用表示。

4.3 候选解释对的匹配概率

在获得关系对的匹配概率后,本文进一步计算对齐子图

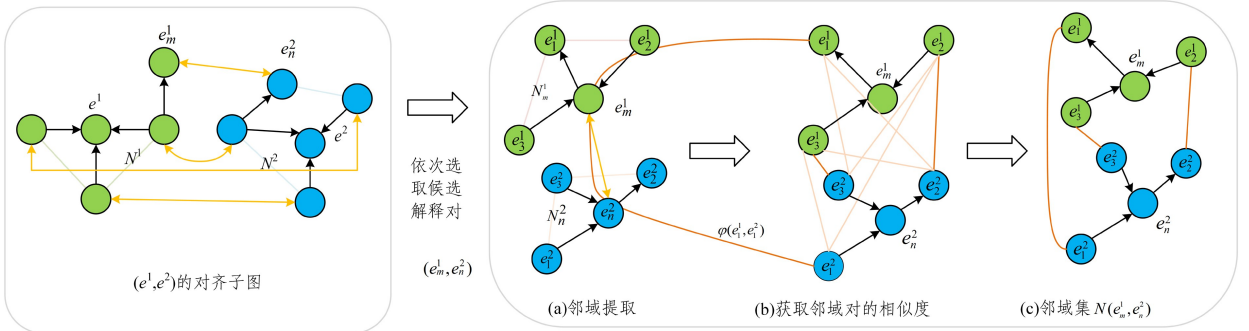


图4 获取候选解释对的邻域集

Fig. 4 Obtain the neighborhood set of candidate explanation pairs

4.3.2 计算候选解释对的匹配概率

引入关系重要性函数 $R_{imp}(e_i, e_j)$ 来归一化候选解释对的匹配概率^[29],该函数计算了相似实体对 $(e_i^1, e_j^2) \in N(e_m^1, e_n^2)$ 与 (e_m^1, e_n^2) 之间的关系功能性的乘积,如式(9)所示:

$$R_{imp}(e_i^1, e_j^2) = func(r_1^1) \times func(r_2^2) \quad (9)$$

采用类似 Jaccard 相似系数的方法,计算 (e_m^1, e_n^2) 所有相似实体对集 $N(e_m^1, e_n^2)$ 的相似度得分总和,并结合 $R_{imp}(e_i, e_j)$,获得 (e_m^1, e_n^2) 匹配概率得分 $prob(e_m^1, e_n^2)$,如式(10)所示:

$$prob(e_m^1, e_n^2) = \frac{\sum_{(e_i^1, e_j^2) \in N(e_m^1, e_n^2)} \varphi(e_i^1, e_j^2)}{\sum_{(e_i^1, e_j^2) \in N(e_m^1, e_n^2)} R_{imp}(e_i^1, e_j^2)} \quad (10)$$

在获取预测实体对 (e^1, e^2) 的所有候选解释对的匹配概率之后,筛选具有高匹配概率得分的候选解释对和关系

中候选解释对的匹配概率,以识别对预测结果具有关键影响的信息。

4.3.1 获取候选解释对的邻域信息

基于预测实体对 (e^1, e^2) 的对齐子图,从中获取预测实体对各自的实体邻域集合 N^1 与 N^2 ,邻域集中中实体 $e_m^1 \in N^1, e_n^2 \in N^2$ 。接着基于 M_{train} 和 M_{res} 依次从邻域集中选取候选解释对 (e_m^1, e_n^2) ,设 N_m^1 和 N_n^2 分别表示实体 e_m^1 和 e_n^2 在图谱 KG1 和 KG2 中的邻域集合,即 $e_i^1 \in N_m^1, e_j^2 \in N_n^2$ 。通过结合关系对的匹配概率并考虑关系的方向性和功能性来获取候选解释对的邻域信息,具体过程如图4所示。设 r_1^1 表示实体 e_m^1 和 e_i^1 之间的关系, r_2^2 表示实体 e_n^2 和 e_j^2 的关系。如果 e_m^1 是尾实体,则关系 r_1^1 的功能性为 $func(r_1^1)$,否则 r_1^1 的功能性为 $ifunc(r_1^1)$, r_2^2 的功能性表示也遵循相同的逻辑。假设关系 r_1^1 是实体 e_i^1 指向实体 e_m^1 , r_2^2 是实体 e_j^2 指向实体 e_n^2 ,定义 φ 为 e_i^1 和 e_j^2 之间的相似性,如式(8)所示:

$$\varphi(e_i^1, e_j^2) = func(r_1^1) \times func(r_2^2) \times match(r_1^1, r_2^2) \times sim(\mathbf{i}_{emb}, \mathbf{j}_{emb}) \quad (8)$$

其中, $\mathbf{i}_{emb}, \mathbf{j}_{emb}$ 分别表示 e_i^1 和 e_j^2 在一个实体对齐模型中的嵌入表示。

对于集合 N_m^1 中的每个实体 e_i^1 ,计算与 N_n^2 中所有实体的相似度得分,并选择得分最大的实体 e_j^2 作为 e_i^1 的相似实体,这些相似实体组合成 (e_m^1, e_n^2) 的新邻域匹配对集 $N(e_m^1, e_n^2) = \{(e_i^1, e_j^2) | e_i^1 \in N_m^1, j = \arg \max_{e_j^2 \in N_n^2} (\varphi(e_i^1, e_j^2))\}$ 。

对组成的关系三元组,这些三元组集组成预测实体对的解释子图。

例如,图5通过一个简单的例子阐释了如何利用对齐子图来提取解释子图,进而解释预测实体对的过程。预测实体对(云南省, Yunnan Province)的候选解释对为(昆明, Kunming)。首先,获取候选解释对的邻居信息如(春城, Spring City),并计算这些邻域对的相似性得分,如 $\varphi(\text{春城}, \text{Spring City}) = 0.832$ 。然后,聚合(昆明, Kunming)的邻居实体对和邻居关系对的相似性得分,计算 $prob(\text{昆明}, \text{Kunming})$ 作为解释的概率为 0.923。最后获取候选解释对与预测对之间的关系匹配得分为 0.891,由于匹配概率为 0.923 和 0.891,都很高,因此保留该邻域信息并将其组合成(云南省, Yunnan Province)的解释子图。

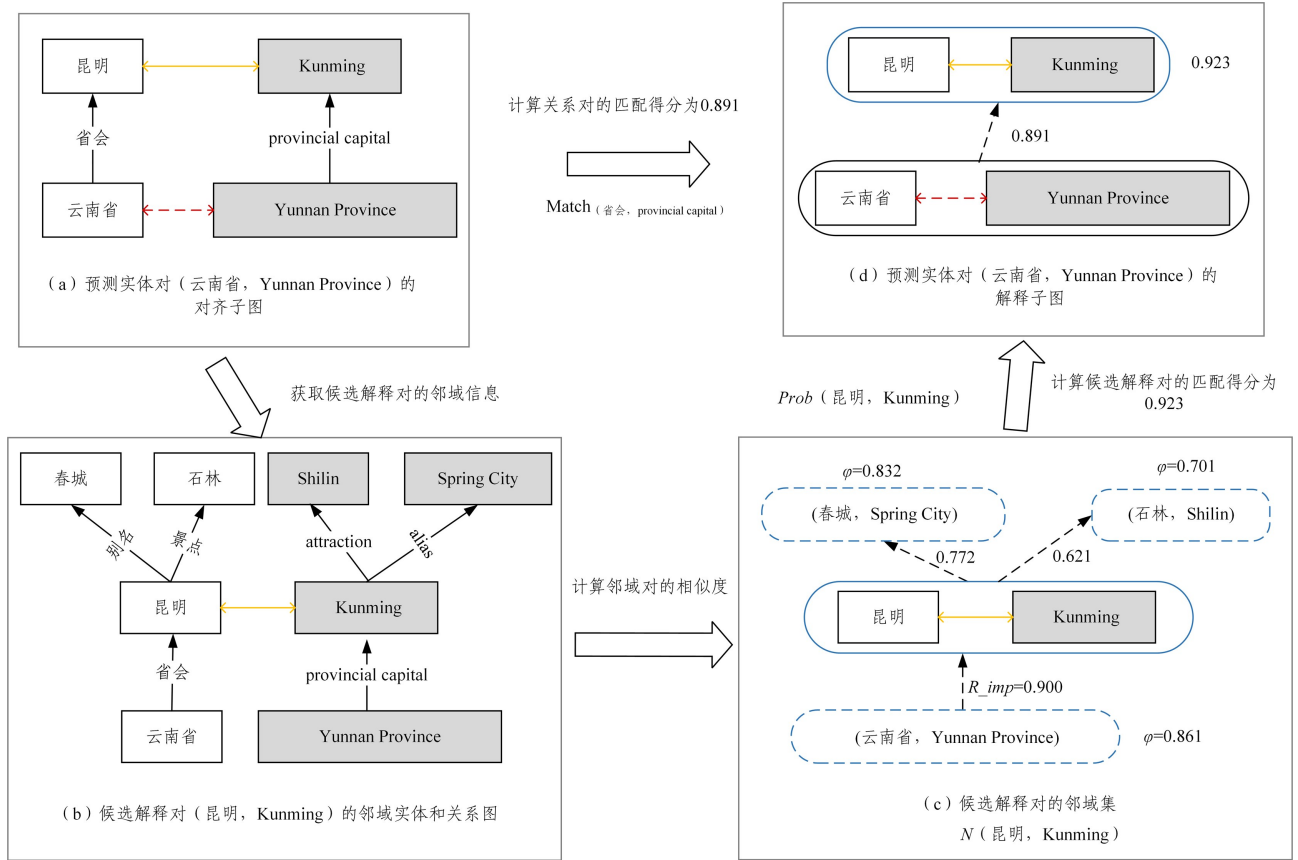


图5 预测对的解释示例图

Fig. 5 Example diagram for predicted pair explanation

4.4 基于类型的图精炼方法

解释是通过事实的组合构建的子图,而连续的事实组合可能会导致某些预测对拥有大量的解释,使得理解和分析变得复杂。为了解决这一问题,通过在解释子图中引入实体类型来有效控制解释过程中的组合爆炸问题。

4.4.1 基于实体类型的匹配过滤

在大多数情况下,实体的表示在不同类型之间差异很大,而在同一类型内差异较小。基于该发现,本文将实体类型作为一种等价关系对解释子图做进一步处理。首先,在解释子图的基础上获取实体的类型 L ,类型在本体中通常以层次树的结构进行组织,每个实体通常与一组类型关联。本文通过冲突度 $Conf[type_a, type_b]$ 来量化两个类型在本体层次结构中的语义差异,如式(11)所示:

$$Conf[type_a, type_b] = \begin{cases} 0, & \text{相交} \\ 1, & \text{互不相交} \end{cases} \quad (11)$$

其中,0 表示 a 和 b 有共同祖先或类层级关系,冲突度为 0; $Conf[type_a, type_b]=1$ 表示 a 和 b 不是同一类型的实体。

接着,比较每个解释对的实体类型是否一致。如果解释对中的两个实体的类型不一致,即存在类型冲突 $conf=1$,这些存在类型冲突的解释对将被过滤,而那些无类型冲突的解释对将被保留。

4.4.2 构建类型商图

商图^[30]是图论中的一个概念,它通过将一个图的顶点集合划分为若干个等价类,然后根据这些等价类构建一个新的

图。在商图中,每个顶点代表原图中的一个等价类,而边则表示原图中不同等价类之间的连接关系集。基于商图的理念,通过类型这一等价关系来构建类型商图。类型商图是一个抽象的图表示方法,它仅包含那些在类型层面上相匹配的实体对,从而提供一个更清晰的视角来观察和分析实体对齐的预测结果。类型商图的定义如下:

定义 5 (类型商图) $G_{type} = (E', R', T')$ 对解释子图 $KG^{exp}(E^{exp}, R^{exp}, T^{exp})$ 中的节点集 E^{exp} , 根据实体类型 $L = \{L_1, L_2, \dots, L_k\}$ 进行等价类划分。其中商图的节点集 $E' = \{\Pi_1, \Pi_2, \dots, \Pi_k\}$ 中每个节点 $\Pi_i = \{e | e \in E^{exp}, type(e) = L_i\}$ 代表一个等价类,包含一个或多个 KG^{exp} 中的实体节点。若 T^{exp} 中存在至少一条边连接等价类 Π_i 中的某个节点到预测实体对 (e^1, e^2) 中的某个节点,那么就在商图的边集 R' 中添加这条边,并在 T' 中添加由这条边、等价类 Π_i 和预测实体对中的节点组成的三元组。

类型商图的生成过程如图 6 所示,解释子图 KG^{exp} 中 $KG1$ 的实体 e_1^1, e_2^1, e_3^1 和 $KG2$ 的实体 e_1^2, e_2^2, e_3^2 的类型标签都为 Type A, 实体 e_4^1 和 e_4^2 的类型标签同为 Type B, 实体 e_5^1 和 e_5^2 的类型标签同为 Type C。将具有相同类型的实体合并,形成一个统一的类型节点,并将这个节点命名为该类型的名称,如将实体 e_1^1 和 e_1^2 聚合成名称为 Type A 的一个商图节点。可以看出,相较于解释子图,类型商图能生成更加简洁且有效的解释。

通过压缩原始解释子图,有效地减少了候选解释的数量,解决了组合爆炸问题。

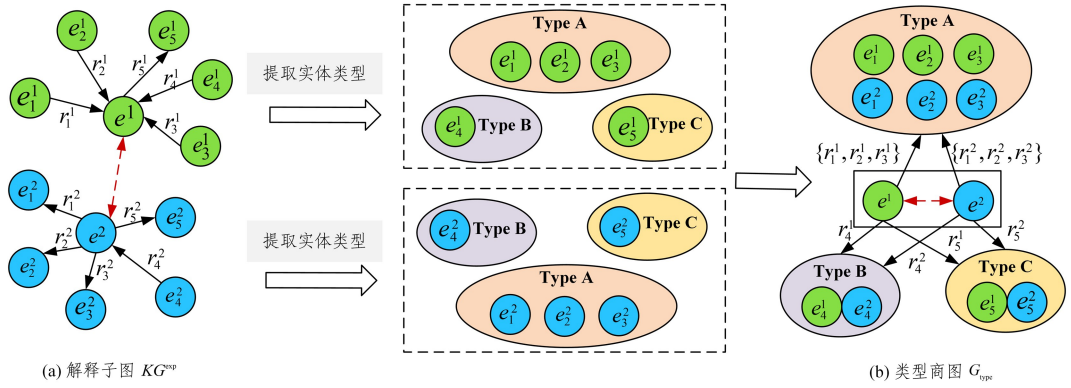


图6 解释子图通过实体类型精炼至类型商图的示意图

Fig. 6 Diagram of refining explanation subgraphs to type quotient graphs via entity types

综上所述,本文方法的伪代码如算法1所示。基于匹配概率与类型商图的实体对齐的解释方法的算法主要包括两阶段:获取关系对和候选解释对的匹配概率,将解释子图转换成类型商图。

算法1 基于匹配概率与类型商图的解释方法

输入: $KG1, KG2$, 种子对 M_{train} , 预测对 M_{res} , 匹配概率得分阈值 Θ

输出: 解释子图 KG^{exp} , 类型商图 G_{type}

/* stage 1: 构建解释子图 */

1. for each $(e^1, e^2) \in M_{res}$ do /* 获取需解释的预测对 */;

2. 初始化: $KG^{exp} \leftarrow \emptyset, G_{type} \leftarrow \emptyset$

3. 基于 M_{res} 和 M_{train} 构建对齐子图 $ASG_{(e^1, e^2)}$

4. /* 从对齐子图中获取候选解释对 */

5. for each $(e_m^1, e_n^2) \in ASG_{(e^1, e^2)}$ do

6. 新邻域集 $N(e_m^1, e_n^2) \leftarrow \emptyset$

7. $N_m^1, N_n^2 \leftarrow e_m^1, e_n^2$ 各自的邻域实体集

8. for each $e_i^1 \in N_m^1$ do

9. for each $e_j^2 \in N_n^2$ do

10. $\text{argmax}(\varphi(e_i^1, e_j^2))$

11. $N(e_m^1, e_n^2) \leftarrow \text{argmax}(\varphi(e_i^1, e_j^2))$

12. $\text{prob}(e_m^1, e_n^2) \leftarrow N(e_m^1, e_n^2)$

13. $\text{prob}(e_m^1, e_n^2) > \Theta$ then:

14. $KG^{exp} \leftarrow KG^{exp} \cup (e_m^1, e_n^2)$

15. /* stage 2: 构建类型商图 */

16. $G_{type} \leftarrow KG^{exp}$ 基于 $KG1$ 和 $KG2$ 的实体类型信息

5 实验

5.1 实验设置

5.1.1 数据集

本研究在实体对齐的3个跨语言数据集和2个异构数据集上对PE-EA模型进行实验评估。DBP15K^[31]是一个从DBpedia中提取的跨语言数据集,它包含了ZH-EN, JA-EN和FR-EN 3对语言的百科全书的结构化知识;异构数据集DBP-WD-V1(D_W)和DBP-YG-V1(D_Y)^[32]从知识图谱DBpedia, Wikidata和YAGO3中采样,每一组数据集都有15000个实体对,用于训练和评估。由于本研究只探究关系三元组的解释性,因此忽略实体的其他特征。

5.1.2 实验软硬件环境和模型参数设计

本文所有的实验都在Ubuntu 20.04系统上执行,服务器

配置为包括64核Intel Xeon Platinum 8336C处理器和2块24GB显存的NVIDIA GeForce RTX 4090显卡。实体对齐模型主要基于Zeng等^[33]提供的代码进行训练,GCN-Align模型的学习率为0.005,每一层隐藏层大小为100;AlignE采用的学习率为0.05;模型训练周期数都为1000。其他与实体对齐模型相关的超参数与目标实体对齐模型保持一致。对于PE-EA解释模型,考虑到解释的可读性,将预测实体对的邻域结构限制在两跳范围内,并将匹配概率阈值设定为0.4作为最优配置。

5.1.3 实体对齐模型

如前所述,PE-EA是一种事后解释方法,支持任何基于嵌入的实体对齐模型。为了展示其灵活性,遵循Zhang等^[32]提出的分类方法,将实体对齐的模型大致划分为两大类,即基于翻译和基于图神经网络,并选取了3个代表性模型来开展本文实验。

1) AlignE^[12]: 基于TransE方法,将实体对齐问题视为一种翻译任务,利用向量空间中的平移关系来表示知识图谱中的实体和关系。

2) GCN-Align^[13]: 使用图卷积网络来学习实体的嵌入表示。GCN-Align通过考虑图结构的局部和全局信息来提高实体对齐的准确性。

3) Dual-AMN^[14]: 一种基于图卷积网络的先进方法,使用双注意力匹配网络结合标准化硬样本挖掘技术来实现快速的实体对齐。

5.1.4 基线模型

实验的目标是验证PE-EA框架生成的解释的有效性,并深入分析在实体对齐算法中生成解释时可能遇到的挑战。在评估过程中,挑选了3种不同的解释方法作为基线模型,与提出的PE-EA模型进行对比分析。

1) EALime: 将知识图谱中的三元组作为一个独立的特征,利用LIME^[26]解释方法训练线性回归模型来实现实体对齐的可解释性。

2) EAShap: 同样将三元组视为一个特征,利用Shap^[27]中Shapley值作为每个特征的贡献度,以表示它对最终预测的贡献大小。

3) ExEA^[28]: 采用双向匹配的方法收集和表示邻近实体

与中心实体之间的关系路径,以识别语义一致的三元组,并将这些三元组组合成子图,作为实体对齐的局部解释。

5.1.5 评价指标

为了以一种客观和一致的方式评价不同解释方法的优劣,本文选择了文献[34]中提出的两个评价指标和 F 分数。

1) Fidelity: 用于衡量在只保留对模型预测至关重要的解释三元组后,模型预测准确率的变化,即计算模型在原始状态下的预测结果与只保留解释三元组的预测结果之间的差异。本文从实体对齐模型的预测中抽取了 1 000 个预测实体对,假设准确率为 1。然后从候选三元组中移除非解释性三元组,重新训练模型,统计这 1 000 对样本中仍保持对齐状态的数量,这个比例就是 Fidelity,如式(12)所示:

$$Fidelity = \frac{TP_{res}}{TP_{ori}} \quad (12)$$

其中, TP_{ori} 表示原始模型在 1 000 个预测对中正确预测的数量(假设为 1 000,即准确率为 1); TP_{res} 表示在移除非解释性三元组 T^{rel} 并重新训练模型后,在相同的 1 000 个预测实体对还能正确预测的数量。

2) Sparsity: 在生成模型解释时,一个好的解释应该只包含对预测结果影响最大的关键特征或信息,即解释应该尽可能简洁,而不是包含所有可能的特征。在本文中,稀疏度用于衡量生成的解释三元组与所有可能三元组的比例,如式(13)所示:

$$Sparsity = 1 - \frac{|T^{exp}|}{|T^*|} \quad (13)$$

其中, T^* 表示包含预测对中实体的所有三元组, T^{exp} 表示解释三元组。

3) F : $F1$ 分数是精确率和召回率的调和平均数,本文采用 $\beta=0.5$ 平衡 F 分数,提高忠实度的比例,如式(14)所示:

$$F = (1 + \beta^2) \frac{Fidelity \times Sparsity}{(\beta^2 \times Fidelity) + Sparsity} \quad (14)$$

5.2 对比实验

利用邻域匹配概率和实体类型方法来验证 PE-EA 模型

与预先训练的实体对齐模型在预测性能上的一致性。采用了 Fidelity 和 Sparsity 两个指标判断解释结果是否真实地反映了原始黑盒模型的预测效果。

一阶三元组是由中心实体和另一个实体通过直接关系连接形成的,而二阶三元组涉及中心实体通过两个中间关系和一个实体间接连接到另一个实体的更复杂的结构。实验分别使用一阶三元组和二阶三元组作为解释搜索的候选项。

5.2.1 基于跨语言数据集的一阶三元组实验

考虑一阶的原因首先是解释的可读性,当候选三元组的范围相对较大时,可能会遇到与所解释的实体对相距甚远的三元组,这使得理解它们之间的关系变得具有挑战性。其次,针对某些嵌入模型,如 AlignE 和 GCN-Align,它们只使用一阶三元组进行实体对齐。表 1 列出了 PE-EA 与其他解释模型在数据 DBP15K 中的 ZH_EN, JA_EN 和 FR_EN 数据集上的结果,通过分析可以得出以下结论: 1) 在稀疏度接近的情况下,在所有数据集的实体对齐模型上,PE-EA 的性能基本优于其他解释方法。这表明,PE-EA 可以为实体对齐任务提供高质量的解释,采用匹配概率方法和实体的类型特征可以挑选出对预测比较重要的信息。

2) ExEA 作为次优方法,有时能接近 PE-EA 的结果,这是因为 ExEA 考虑了三元组之间的相互作用,有效地识别出两个知识图谱中的相似三元组。然而,ExEA 主要关注局部解释,而 PE-EA 考虑了实体和关系的功能性,关注全局结构,能更全面地理解实体之间的关系,因此 ExEA 较 PE-EA 的效果略差。

3) 对于 EALime 和 EAShap 这两种方法,EALime 的原理是将单个三元组作为独立特征,并用线性模型来拟合实体对齐模型,而实体对齐模型通常涉及更复杂的非线性关系; EAShap 虽然考虑了特征之间的交互作用,通过 Shapley 值来分配预测贡献,但它仍然依赖局部采样,可能无法全局捕捉关键特征组合,因此效果较差。

表 1 一阶邻域下 PE-EA 与其他 3 种解释方法在跨语言数据集上的忠实度和稀疏度比较

Table 1 PE-EA vs. other methods: fidelity and sparsity in first-order neighborhood of cross-lingual datasets

方法	数据	PE-EA		ExEA		EALime		EAShap	
		Sparsity	Fidelity	Sparsity	Fidelity	Sparsity	Fidelity	Sparsity	Fidelity
AlignE	ZH_EN	0.593	0.891	0.552	0.729	0.416	0.550	0.418	0.628
	JA_EN	0.544	0.907	0.486	0.800	0.441	0.524	0.442	0.558
	FR_EN	0.599	0.943	0.545	0.874	0.575	0.596	0.575	0.547
GCN-Align	ZH_EN	0.626	0.759	0.596	0.799	0.500	0.540	0.521	0.540
	JA_EN	0.633	0.763	0.577	0.705	0.485	0.503	0.503	0.495
	FR_EN	0.637	0.759	0.602	0.740	0.599	0.439	0.614	0.418
Dual-AMN	ZH_EN	0.549	0.943	0.476	0.930	0.441	0.855	0.441	0.857
	JA_EN	0.557	0.930	0.473	0.941	0.469	0.840	0.468	0.840
	FR_EN	0.604	0.941	0.480	0.940	0.484	0.926	0.486	0.927

5.2.2 基于异构数据集的一阶三元组实验

为了证实所提出的解释模型能够适应异构图,进行了针对 D_W 和 D_Y 数据集的实验,以验证模型的有效性。此外,考虑到异构数据集中类型关系的复杂性,解释模型仅考虑概率匹配,而不涉及实体类型的解释。这一发现进一步验证了

PE-EA 模型在处理异构数据集时的有效性。然而,对比表 1 和表 2 中的数据,可以观察到表 2 中的解释指标普遍低于跨语言数据集的指标。这一现象主要归因于异构数据集的复杂性,如结构和数据来源的复杂性等,这些因素共同增加了数据处理和分析的难度。

表2 一阶邻域下 PE-EA 与其他 3 种解释方法在异构数据集上的忠实度和稀疏度比较

Table 2 PE-EA vs. other methods: fidelity and sparsity in first-order neighborhood of heterogeneous datasets

方法	数据	PE-EA(w/o Type)		ExEA		EALime		EAShap	
		Sparsity	Fidelity	Sparsity	Fidelity	Sparsity	Fidelity	Sparsity	Fidelity
AlignE	D_W	0.686	0.569	0.641	0.569	0.543	0.463	0.541	0.409
	D_Y	0.509	0.655	0.513	0.636	0.330	0.574	0.331	0.531
GCN-Align	D_W	0.729	0.407	0.706	0.388	0.623	0.370	0.642	0.369
	D_Y	0.589	0.548	0.564	0.560	0.452	0.533	0.458	0.499
Dual-AMN	D_W	0.445	0.722	0.383	0.719	0.398	0.717	0.431	0.718
	D_Y	0.274	0.786	0.329	0.779	0.198	0.784	0.230	0.784

5.2.3 基于跨语言数据集的二阶三元组实验

对于涉及 2 阶三元组作为候选三元组的实验,本文只在 Dual-AMN 上进行实验,这个决定是根据前面讨论的解释搜索的候选项做出的。从表 3 中结果可以看出,当考虑的候选三元组范围较大时,稀疏度下降的同时忠实度没有提升,性能略微下降。此外,考虑二阶邻居信息需要处理更多的数据和更复杂的图结构,会增加计算成本。因此,更适合使用 1 阶邻居信息对预测的实体对进行解释。

表3 Dual-AMN 模型中 1 阶和 2 阶邻域的结果

Table 3 Results of the 1st and 2nd order neighborhoods in the

Dual-AMN model

数据	1 hop		2 hop	
	Fidelity	Sparsity	Fidelity	Sparsity
ZH_EN	0.909	0.476	0.901	0.442
JA_EN	0.865	0.473	0.829	0.438
FR_EN	0.874	0.480	0.837	0.441

5.3 消融实验

通过对 PE-EA 的两个变体进行消融研究,以探索模型中每个组件的影响。表 4 列出了不同 PE-EA 模型变体的消融研究的性能比较结果,其中“w/o Prob”和“w/o Type”是将匹配概率和实体类型从提议的实体对齐解释框架中分离出来的两个 PE-EA 变体。

表4 AlignE 模型的实验评估结果

Table 4 Results of the experimental evaluation of AlignE

数据	评价指标	PE-EA	w/o Prob	w/o Type
ZH-EN	Sparsity	0.593	0.543	0.592
	Fidelity	0.891	0.732	0.886
JA-EN	Sparsity	0.544	0.491	0.541
	Fidelity	0.907	0.796	0.900
FR-EN	Sparsity	0.599	0.555	0.585
	Fidelity	0.950	0.814	0.927

从表 4 中可以看出,在 AlignE 模型中两种 PE-EA 变体都能获得较好的性能,并且使用整合两部分的组件在所有数据集上都表现最好。当移除匹配概率组件时,性能下降幅度较大;当移除实体类型组件时,性能下降幅度较小。这表明,匹配概率组件不论是单独使用还是结合实体类型,都能够提高解释的有效性。此外,仅采用实体类型进行解释的效果并非很突出,可能是一些预测的实体对的类型信息不充足,导致搜索空间有限,有效的类型解释很少。

5.4 匹配概率阈值分析

得分概率阈值作为一个超参数,用于从大量的知识中提取和筛选解释。在完成功能性概率计算后,选取所有候选解释对的得分大于阈值的三元组作为解释,图 7 展示了不同阈

值对解释结果的影响。随着阈值的提升,被选为解释的三元组的数量呈现递减趋势,解释的稀疏性增加,但与此同时,解释的忠实度却未能得到保证。

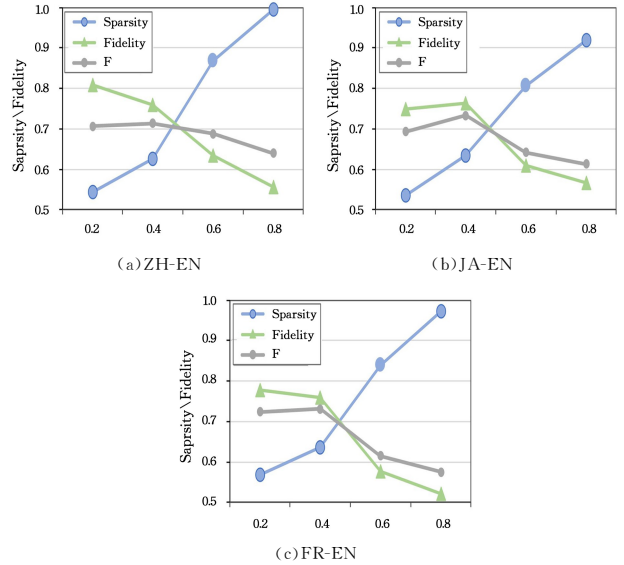


图7 GCN-align 模型中不同阈值下指标的变化情况

Fig. 7 Variation of metrics under different thresholds in the GCN-align model

从图 7 可以看出,随着阈值从 0.2 增至 0.8, Fidelity 指标呈现出递减的趋势,表明了较高阈值下,被排除的三元组实际上对模型的正确推理起到了积极的作用。阈值由 0.2 提升至 0.4 的过程中, Fidelity 的下降相对平缓,而解释的稀疏性得到了显著提升。相对地,当阈值从 0.4 提升至 0.8 时, Fidelity 的下降变得尤为显著,在此区间内解释的可靠性受损。本研究的核心目标是在确保解释的高忠实度的基础上,尽可能地采用简洁且关键的三元组进行有效解释。因此,结合指标 F 可以得出,当阈值为 0.4 时 F 得分最高,解释的简洁性得到了改善并对忠实度的负面影响较小。

5.5 模型性能分析

PE-EA 事后解释模型分为 3 个阶段:关系对匹配、实体对匹配和生成类型商图。由于实体和关系的功能性值以及关系对的相似性值是固定的,它们在关系对匹配阶段只需计算一次即可重复使用。同时,由于 PE-EA 的匹配计算阶段使得模型的运行时间略长于 EALime 等模型,但这种差异在可接受范围内,并且 PE-EA 保证了解释的准确性和质量。为了进一步分析模型的性能,图 8 展示了 PE-EA 模型在 AlignE 和 GCN-Align 上解释 1000 对预测的性能。

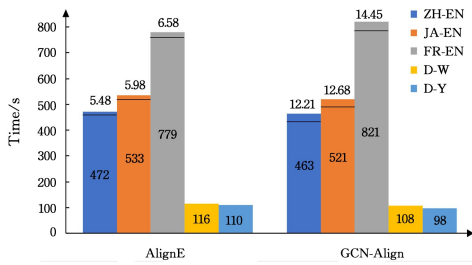


图8 PE-EA模型在解释AlignE和GCN-Align的1000对预测对所需时间

Fig. 8 PE-EA model's time in seconds needed to explain 1000 prediction pairs for AlignE and GCN-Align

跨语言数据集上的柱形图中的数值表示生成解释和类型子图的总时间,每个柱形图顶端的数值表示生成类型子图的时间;异构数据集D-W和D-Y由于实体类型的多样复杂性,

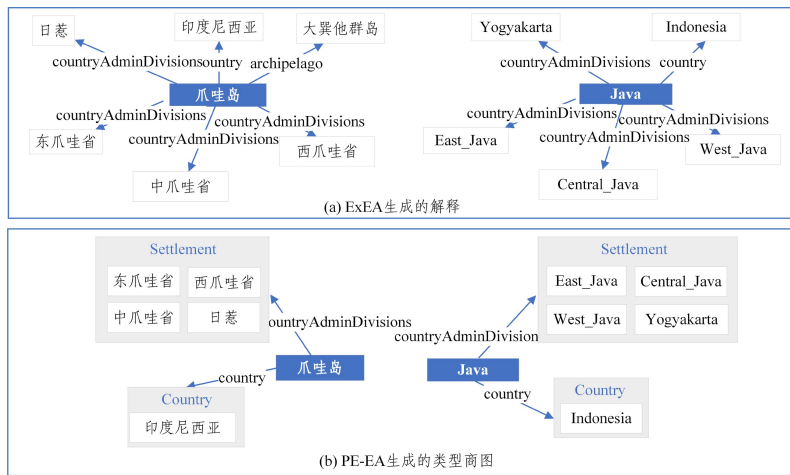


图9 类型商图的解释案例研究

Fig. 9 Explanatory case study of type quotient graph

结束语 本文提出了一种新颖的实体对齐后解释框架PE-EA,旨在通过提取关键的上下文知识生成的子图来解释预测结果。通过综合考虑实体和关系的功能性以及相应的嵌入向量,有效地捕捉了知识图谱的全局结构和局部特征,为模型的预测提供可解释的证据。此外,利用类型商图对解释子图进行语义层面的概括和聚类,改善了解释的效率和清晰度。实验结果表明,在4种对齐嵌入模型中,PE-EA与3种不同的解释方法进行对比验证时,表现最优。

然而,该解释模型仅仅考虑了实体的结构特征,忽略了实体的其他特征,如实体的属性、图像特征等。随着多模态数据的日益丰富和多模态实体对齐的深入研究,未来将尝试利用多模态特征来提升实体对齐任务的解释能力。

参考文献

[1] HUANG P,ZHANG M,ZHONG Z,et al. Representation Learning for Entity Alignment in Knowledge Graph:A Design Space Exploration[C]//2024 IEEE 40th International Conference on Data Engineering(ICDE). IEEE,2024;3462-3475.
[2] ZHANG T C,TIAN X,SUN X H,et al. Overview on Knowledge Graph Embedding Technology Research[J]. Journal of Software,2023,34(1):277-311.

仅考虑生成解释子图的运行时间。从图8可以得出,PE-EA通常能够在几百毫秒的时间就提取出一对预测对的解释;这些时间主要取决于知识图谱中数据的数量,如数据集FR-EN三元组和实体的数量最多,在2个实体对齐模型消耗的时间都是最长的,而数据集D-Y的数量最少,因此耗时也最短。

5.6 类型商图的解释案例研究

本节展示了所提出方法的一个典型解释输出示例,重点在于展示类型商图的解释的有效性。更具体地,本文在ZH-EN数据集上使用实体对齐模型AlignE进行实验,对预测实体对(爪哇岛,Java)进行解释展示。图9(a)表示由解释模型ExEA生成的解释三元组,与图9(b)相比,可以看出使用实体类型来压缩原始解释子图,更容易使用户接受和理解,并解决了生成解释时的组合爆炸问题。

[3] ZHU B,WANG R,WANG J,et al. A survey: knowledge graph entity alignment research based on graph embedding[J]. Artificial Intelligence Review,2024,57(9):1-58.
[4] VAN DER VELDEN B H M,KUIJFF H J,GILHUIJS K G A,et al. Explainable artificial intelligence(XAI) in deep learning-based medical image analysis[J]. Medical Image Analysis,2022,79:102470.
[5] CHOWDHURY T,RAHIMI R,ALLAN J. Equi-explanation maps: concise and informative global summary explanations[C]//Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency. 2022;464-472.
[6] LAMPRIDIS O,GUIDOTTI R,RUGGIERI S. Explaining sentiment classification with synthetic exemplars and counter-exemplars[C]// International Conference on Discovery Science. Cham:Springer,2020;357-373.
[7] DOU H,ZHANG L M,HAN F,et al. Survey on Convolutional Neural Network Interpretability[J]. Journal of Software,2024,35(1):159-184.
[8] FAN Y,YAO Y,JOE W C. Gen-se: Attention as explainability for node classification in dynamic graphs[C]//2021 IEEE International Conference on Data Mining(ICDM). IEEE,2021;1060-1065.

- [9] LI X, SAÚDE J. Explain graph neural networks to understand weighted graph features in node classification[C]//International Cross-Domain Conference for Machine Learning and Knowledge Extraction. Cham:Springer, 2020:57-76.
- [10] ROSSI A, FIRMANI D, MERIALDO P, et al. Explaining link prediction systems based on knowledge graph embeddings[C]//Proceedings of the 2022 International Conference on Management of Data. 2022:2062-2075.
- [11] CHEN M, TIAN Y, YANG M, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence(IJCAI). 2017:1511-1517.
- [12] SUN Z, HU W, ZHANG Q, et al. Bootstrapping entity alignment with knowledge graph embedding[C]//IJCAI. 2018.
- [13] WANG Z, LV Q, LAN X, et al. Cross-lingual knowledge graph alignment via graph convolutional networks[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:349-357.
- [14] MAO X, WANG W, WU Y, et al. Boosting the speed of entity alignment $10\times$: Dual attention matching network with normalized hard sample mining[C]//Proceedings of the Web Conference 2021. 2021:821-832.
- [15] MAO X, WANG W, WU Y, et al. From alignment to assignment: Frustratingly simple unsupervised entity alignment[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021:2843-2853.
- [16] JI G S, ZHU Y C, WANG Y, et al. Multi-depth joint entity alignment based on textual information multi-depth joint entity alignment based on textual information [J]. Applied Intelligence, 2023, 51: 8896-8912.
- [17] MAO X, WANG W, WU Y, 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. 2022: 825-838.
- [18] LUO Y, CHEN Z, GUO L, et al. ASGEA: Exploiting Logic Rules from Align-Subgraphs for Entity Alignment [J]. arXiv: 2402.11000, 2024.
- [19] XU C, CHENG J, ZHANG F. P-NAL: an Effective and Interpretable Entity Alignment Method [C] // Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2024: 13752-13772.
- [20] TRISEDYA B D, SALIM F D, CHAN J, et al. i-Align: an interpretable knowledge graph alignment model [J]. Data Mining and Knowledge Discovery, 2023, 37(6): 2494-2516.
- [21] YING Z, BOURGEOIS D, YOU J, et al. Gnnexplainer: Generating explanations for graph neural networks [C] // Proceedings of the Conference on Neural Information Processing Systems. 2019: 9244-9255.
- [22] LUO D, CHENG W, XU D, et al. Parameterized explainer for graph neural network [J]. Advances in Neural Information Processing Systems, 2020, 33: 19620-19631.
- [23] YUAN H, YU H, WANG J, et al. On explainability of graph neural networks via subgraph explorations [C] // International Conference on Machine Learning. PMLR, 2021: 12241-12252.
- [24] PEZESHKPOUR P, TIAN Y, SINGH S. Investigating robustness and interpretability of link prediction via adversarial modifications [C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 3336-3347.
- [25] ZHAO D, WAN G, ZHAN Y, et al. KE-X: Towards subgraph explanations of knowledge graph embedding based on knowledge information gain [J]. Knowledge-Based Systems, 2023, 278: 110772.
- [26] RIBEIRO M T, SINGH S, GUESTRIN C. "Why should I trust you?" Explaining the predictions of any classifier [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1135-1144.
- [27] LUNDBERG S M, LEE S I. A unified approach to interpreting model predictions [J]. Advances in Neural Information Processing Systems, 2017, 30: 4768-4777.
- [28] TIAN X, SUN Z, HU W. Generating Explanations to Understand and Repair Embedding-based Entity Alignment [C] // 2024 IEEE 40th International Conference on Data Engineering (ICDE). IEEE, 2024: 2205-2217.
- [29] XIN K, SUN Z, HUA W, et al. Informed multi-context entity alignment [C] // Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. 2022: 1197-1205.
- [30] ČEBIRIĆ Š, GOASDOUÉ F, KONDYLAŠIS H, et al. Summarizing semantic graphs: a survey [J]. The VLDB Journal, 2019, 28: 295-327.
- [31] SUN Z, HU W, LI C. Cross-lingual entity alignment via joint attribute-preserving embedding [C] // International Semantic Web Conference. Cham: Springer, 2017: 628-644.
- [32] ZHANG R, TRISEDYA B D, LI M, et al. A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning [J]. The VLDB Journal, 2022, 31(5): 1143-1168.
- [33] ZENG K, LI C, HOU L, et al. A comprehensive survey of entity alignment for knowledge graphs [J]. AI Open, 2021, 2: 1-13.
- [34] POPE P E, KOLOURI S, ROSTAMI M, et al. Explainability methods for graph convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10772-10781.



ZHANG Xiaoming, born in 1975, Ph.D., professor, is a member of CCF (No. 30537M). His main research interests include knowledge graph, semantic computing and multi-modal knowledge fusion.



WANG Huiyong, born in 1980, Ph.D., associate professor, is a member of CCF (No. 42025M). His main research interests include knowledge graph and computer vision.