

文章编号:1671-6833(2024)04-0053-09

基于关系学习的小样本知识图谱补全模型

李卫军, 顾建来, 张新勇, 高庾潇, 刘锦彤

(北方民族大学 计算机科学与工程学院, 宁夏 银川 750021)

摘要: 在小样本知识图谱中, 实体对之间的关系表示复杂多样。然而, 现有的小样本知识图谱补全方法普遍存在关系学习能力不足和忽略实体上下文语义的问题。为解决这些问题, 提出了一种基于关系学习的小样本知识图谱补全模型 FRLC。首先, 在聚合高阶邻域实体信息的过程中引入了门控机制, 这一步骤旨在丰富中心实体表达的同时减少噪声对邻居的不良影响。其次, 在关系学习阶段充分利用参考集实体对之间的相关性, 实现更加准确的关系表示。最后, 在 Transformer 学习器中, 引入了 LSTM 结构进一步学习实体和关系的上下文语义信息, 用于预测新的事实知识。为了验证 FRLC 的有效性, 在公开的 NELL-One 和 Wiki-One 数据集上将 FRLC 与 6 个小样本知识图谱补全模型和 5 个传统模型的 5-shot 链接预测进行了对比实验, 结果表明: FRLC 在 MRR 、 $Hits@10$ 、 $Hits@5$ 和 $Hits@1$ 这 4 个指标上都有所提升, 证明了模型的有效性。

关键词: 知识图谱补全; 小样本关系; 邻域聚合; 链接预测

中图分类号: TP391; TP183

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2024.01.016

知识图谱(knowledge graph, KG)是 Google 公司于 2012 年提出的用于提高搜索引擎速度的概念^[1]。其中包括了丰富的社会多元关系数据, 并广泛地运用在人类社会的各种生产任务当中。现有的知识图谱如 Freebase^[2]、YAGO^[3]等, 都包括了大量由事实所构成的三元组, 而这种三元组一般以(头实体, 关系, 尾实体)的形式表示, 即 (h, r, t) 。在工作生产中, 这种以图结构表示的知识在诸多下游应用中具有重要作用, 包括语义搜索^[4]、智能问答^[5]以及个性化推荐^[6]等领域。然而, 知识图谱仍然存在事实不完整的问题, 为了解决这个问题, 有必要通过知识图谱补全(knowledge graph completion, KGC)来自动推导和填充缺失的事实, 进一步提高知识图谱的利用价值。

近年来, 相关研究者针对 KGC 任务提出了许多基于知识图嵌入技术^[7]的知识图谱补全模型, 在一定程度上解决了知识图谱实体和关系不完整的问题。但实际背景下, 传统知识图谱补全模型不能很好地解决不常见的实体和关系的知识图谱补全问题。相关研究者提出多种小样本的知识图谱补全

(few-shot knowledge graph completion, FKGC)方法。在 FKGC 任务上, 基线模型都取得了很好的结果, 但 these 方法还存在一些不足。

1 相关工作

知识图谱补全任务是指通过自动推理和学习的方式填补知识图谱中缺失的实体、关系和属性等信息。目前, 基于知识图谱嵌入的补全方法主要有如下 3 类: 基于翻译模型的方法、基于语义匹配的方法以及基于神经网络的方法。基于翻译模型的方法是将关系看作实体对之间的翻译操作, 并通过距离评分函数来判断实体和关系之间是否存在关联。如 TransE^[8]是一个经典的翻译模型, 将关系视为头实体到尾实体的翻译过程, 从而有效地应对大规模的知识图谱补全任务。基于语义匹配的方法是使用基于相似性的评分函数来挖掘实体和关系之间的潜在语义, 以解决知识图谱补全任务。如 DISTMULT^[9]模型将关系矩阵限制为对角阵简化模型, 但也导致其不能处理非对称关系。ComplEx^[10]模型使用复向量来表示实体和关系, 能够有效地建模多种二元关

收稿日期: 2023-08-05; 修订日期: 2023-10-09

基金项目: 中央高校基本科研业务费专项资金资助项目(2021JCYJ12); 国家自然科学基金资助项目(62066038, 61962001); 宁夏自然科学基金资助项目(2021AAC03215)

作者简介: 李卫军(1979—), 男, 陕西渭南人, 北方民族大学讲师, 博士, 主要从事本体论的构建与再利用、知识图谱构建、深度学习研究, E-mail: lwj@nmu.edu.cn。

引用本文: 李卫军, 顾建来, 张新勇, 等. 基于关系学习的小样本知识图谱补全模型[J]. 郑州大学学报(工学版), 2024, 45(4): 53-61. (LI W J, GU J L, ZHANG X Y, et al. Relation learning completion model for few-shot knowledge graphs [J]. Journal of Zhengzhou University (Engineering Science), 2024, 45(4): 53-61.)

系。相较于基于翻译模型的方法,基于语义匹配的方法能够更好地捕捉实体间的隐含语义。基于神经网络的方法利用神经网络强大的学习能力和表达能力进行建模。如 GATs^[11]通过聚合知识图谱中的路径信息和图结构数据来捕捉信息,使用图注意力网络的注意力机制为不同邻居节点分配不同权重来达到捕捉最关键的邻居节点信息的目的。

传统的知识图谱补全方法需要大量的训练实例来提高模型准确率,当知识图谱无法为这些新增事实提供足够的训练实例时,模型的推理结果会受到极大的限制。因此,相关研究人员开始探究在只有少量实例的情况下进行知识图谱补全任务。

在 FKGC 任务中,由 Xiong 等^[12]提出的 GMatching 模型首次定义了小样本知识图谱补全问题。使用基于度量学习的方法,通过邻域中的一条邻居结构获得嵌入表征,使用 LSTM 网络将嵌入表征与目标匹配进而得到相似度分数来度量查询三元组与参考集的相似性。然而,GMatching 对邻居信息并未进行区分,引入了噪声问题。FSRL^[13]可以有效地从异构图结构中捕获知识,聚合少量样本的表示,并利用异构邻居解码器给邻域信息分配不同的权重。但上述 2 种模型并未充分考虑知识图谱中的实体和关系具有动态属性的这一特点。FAAN^[14]考虑了实体和关系的动态属性,通过注意力机制捕获不同任务中发生变化的动态特征,以提高其细粒度语义表示。另一方面,MetaR^[15]与上述利用邻域信息增强实体嵌入的方法不同,其采用了包括梯度元和关系元的元学习框架进行 FKGC 任务,有效地提升了模型的学习能力。GANA^[16]对 MetaR 进行了改进,通过采用全局-局部框架有效地过滤了邻域中的噪声信息,并且优化了在知识图谱中涉及一对多、多对一和多对多等复杂关系问题的表现。

在 FKGC 任务中,上述模型都取得了很好的结果,但这些方法还存在一些不足:①FAAN^[14]使用注意力机制聚合邻域信息时,并没有很好地将高阶邻域实体的有效信息利用起来,同时针对不同邻居信息的重要性带来的噪声问题没有妥善解决;②FSRL^[13]简单地使用循环自编码器聚合少量的参考集,但随着训练次数的增加会导致关系嵌入对实体的依赖,造成关系过拟合的问题,从而降低了模型的泛化性;③FSRL 在进行匹配查询时,未充分考虑 TransE^[8]模型的平移性对匹配精度造成的影响;④以往的方法并未充分考虑上下文语义对实体对表示的影响,从而降低了链接预测的准确性。

针对上述问题,本文提出一种基于关系学习的

小样本知识图谱补全模型(few-shot relation learning completion, FRLC),该模型通过学习实体和小样本关系的细粒度语义表示,达到小样本知识图谱补全的目的。

2 模型设计与实现

2.1 问题描述

知识图 \mathcal{G} 由多种事实组成,每个事实可以用三元组集合 $\mathcal{G} = \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$ 来表示,其中 \mathcal{E} 表示实体集合, \mathcal{R} 表示关系集合。本文研究的是给定头实体 h 和关系 r 预测尾实体 t 。

基于上述定义,小样本知识图谱补全任务可定义为给定一个任务关系 r 和它的参考集 $S_r = \{(h_i, t_i) \mid (h_i, r, t_i) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$, 根据知识图 \mathcal{G} 和参考集 S_r 提供的信息,并利用查询三元组 $(h_j, r, ?)$ 中的头实体 h_j 和任务关系 r 的链接,预测尾实体 t_j 。其中 $|S_r| = K$, 通常 K 值很小,故称为小样本知识图谱补全^[12]。因为小样本知识图谱补全是一个特定的关系任务,所以在知识图 $\mathcal{G} = \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$ 中,关系 \mathcal{R} 分为 \mathcal{R}_{br} , \mathcal{R}_{task} 。其中 \mathcal{R}_{br} 为背景知识图 \mathcal{G}' 中的关系集,背景知识图 \mathcal{G}' 为知识图 \mathcal{G} 移除所有任务关系的子图; \mathcal{R}_{task} 为任务关系集,其中包括 \mathcal{R}_{train} , $\mathcal{R}_{validation}$, \mathcal{R}_{test} , 分别在 FKGC 任务的元训练(meta-train)、元验证(meta-validation)以及元测试(meat-test)3 个阶段使用。

图 1 给出了一个 5-shot KGC 任务示例。在小样本知识图谱补全任务中,查询三元组 (Windows, ProducedBy, ?) 的参考集有 5 个关联三元组,利用参考集为查询三元组 (Windows, ProducedBy, ?) 匹配正确的尾实体。即真实尾实体 Microsoft 的排名要高于其他尾实体。通过少量的参考信息预测新的事实,就是 FKGC 任务的核心。

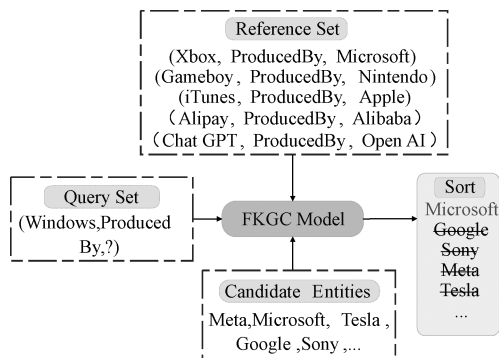


图 1 5-shot KGC 示例

Figure 1 An example of 5-shot KGC task

本文提出的 FRLC 模型是通过利用背景知识图谱以及知识图谱结构信息来训练学习,并实现小样

本关系的尾实体预测。FRLC 的整体框架如图 2 所示,其中包括:基于门控机制的高阶邻域实体编码器,使实体自适应地聚合邻域信息;关系表示编码器,利用参考集实体对的邻居关系信息,达到降低关系对于实体嵌入依赖性的目的,提高模型泛化性;Transformer 学习器,将 LSTM 单元与 Transformer 模块结合,进一步学习任务关系表示;匹配处理计算,利用 Transformer 学习器输出的关系语义嵌入,并计算与查询关系的相似性,预测新的三元组。

2.2 基于门控机制的高阶邻域实体编码器

在知识图谱中,实体的异构邻居对实体的表示有不同的影响,根据任务关系 r 自适应聚合邻域信息能够提高中心实体的表达质量。但这种方法没有将高阶邻域的实体信息聚合到中心实体,并且忽略了有用邻居和无用邻居对中心实体的编码影响。基于此,本文设计了一个基于门控机制的高阶邻域实

体编码器,拓展了 FAAN^[14] 模型的“自适应邻域编码器”,在一阶邻域的基础上拓展到高阶邻域,并且加入了门控机制。在提高实体的表达性的同时减少噪声邻居对中心实体编码更新的影响。基于门控机制的高阶邻域实体编码器的主要结构如图 3 所示。

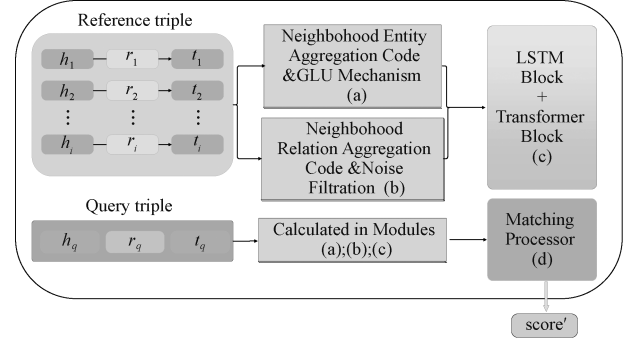


图 2 FRLC 模型的整体框架

Figure 2 Overview of the FRLC framework

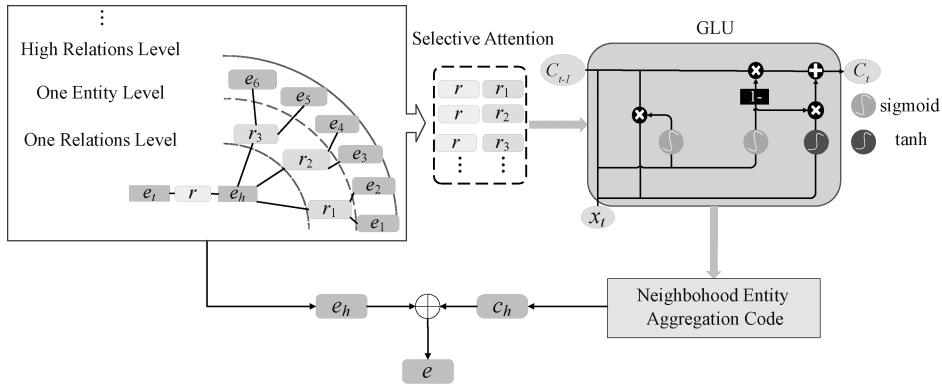


图 3 基于门控机制的高阶邻域实体编码器

Figure 3 High order neighborhood entity encoder based on gating mechanism

给定任务三元组 (h, r, t) , 假设实体 h 为目标实体,通过高阶邻域实体编码器更新它的编码。定义实体 h 的高阶邻域为 $N_h^o = \{r_i^o, t_i^o \mid (h, r_i^o, t_i^o) \in \mathcal{G}'\}$, 其中 \mathcal{G}' 表示背景知识图; r_i^o 和 t_i^o 表示实体 h 的第 i 个高阶邻居关系以及对应的尾实体。为了量化实体 h 的特征,首先使用度量函数 $\varphi^o(\cdot)$ 计算参考关系 r 与实体 h 的相邻高阶关系 r_i^o 的相似性:

$$\varphi^o(r, r_i^o) = r^T (W r_i^o + b). \quad (1)$$

式中: r 为任务关系 r 的初始特征; r_i^o 为高阶相邻关系 $r_i^o \in N_h^o$ 的预训练嵌入; W 和 b 分别为权重和偏置参数; $\varphi^o(\cdot)$ 为双线性点积函数。

根据高阶邻域编码的相似性得分 $\varphi^o(\cdot)$, 得到较高分数的实体 t_i^o 赋予较大注意力得分。为了进一步提高实体 h 的表达质量,降低噪声邻居的影响,为更重要的邻居分配更高的权重。在计算邻居权重的过程引入“门控机制”自适应计算权重 α_i :

$$\alpha_i = \text{softmax}(G(h^o, e_i^o)); \quad (2)$$

$$G(h^o, e_i^o) = \sum_{o=1}^n (h^o \cdot e_i^o) \cdot r_i^o. \quad (3)$$

式中:门控机制函数 $G(\cdot)$ 通过计算中心实体和邻居实体的内积; n 为中心实体 h^o 与邻居实体 e_i^o 最大相邻阶; o 代表第 o 阶。 $G(\cdot)$ 将 h^o 和 e_i^o 作为输入,输出表示邻域与中心实体相关性的度量值, $\text{softmax}(\cdot)$ 将输入转换为邻居的概率分布。

为了更好地捕捉特定任务关系,引入了一个可学习的任务关系 R , 用于更新任务关系 r_i^o :

$$r_i'^o = R \cdot r_i^o. \quad (4)$$

式中: $r_i'^o$ 为更新后的任务关系。由此,可以获得中心实体 h 的邻域编码 c_h :

$$c_h = \text{GLU}\left(\sum_{o=1}^n \text{softmax}(G(h^o, e_i^o)) \cdot e_i^o \cdot r_i'^o\right). \quad (5)$$

式中: $\text{GLU}(\cdot)$ 为门控线性单元激活函数。最后,利用邻域编码器的输出和中心实体的初始特征自适应更新实体表示:

$$e = \sigma(w_1 \cdot c_h + w_2 \cdot e_h). \quad (6)$$

式中: $\sigma(\cdot)$ 为 sigmoid 激活函数; h 为中心实体的初始化特征; w_1, w_2 为可学习参数; e 为中心实体 h 更新后的实体嵌入。此实体更新步骤应用于参考集和查询集的所有实体。本文提出的高阶邻域实体编码器考虑了高阶邻域实体的邻居所蕴含的信息,并引入门控机制,针对特定任务关系的噪声邻居进行过滤。本方法将高阶邻域的范围设置为 3, 因为过低的范围将丢失大量的隐藏信息,但过高的范围将降低模型的性能。

2.3 关系表示编码器

通过上述步骤,对参考集和查询集的实体进行了编码,以获得它们的表示 h 和 t 。将参考集的原型关系 r_s 用 h_s 和 t_s 相连接起来的方式表达:

$$r_s \approx h_s + t_s. \quad (7)$$

然而,大部分的 FKGC 方法都只是简单地将实体对的平均嵌入作为关系的表示,并未充分利用参考集中的实体对之间的相关性。使用参考集中实体对的邻居关系能够丰富当前关系的语义表达,降低关系对于实体嵌入的依赖性,提高泛化性。关系表示编码器的主要结构如图 4 所示。

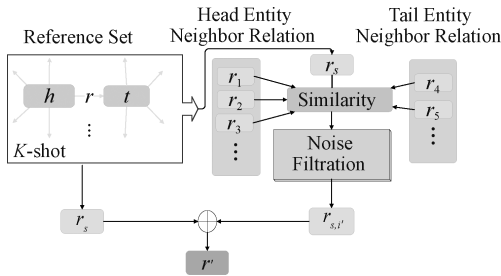


图 4 关系表示编码器

Figure 4 Relationship representation encoder

首先,为了表示参考集中第 i 个实体对 (h_i, e_i) 的头实体 h_i 和尾实体 e_i 的邻居关系,将参考集实体对邻居关系定义为 $R'_r = S_r \{r | (h_i, r, ?), (e_i, r, ?) \in \mathcal{G}\}$ 。将邻居关系集合中的关系 r 表示为

$$E_{r_{s,i}} = \{E_r | r \in S_r\}. \quad (8)$$

式中: $E_{r_{s,i}}$ 为集合 R'_r 中第 i 个邻居关系 r_s 的表示。在丰富当前关系的语义表达的同时,降低噪声的影响。使用特征向量的点积相似度计算 $E_{r_{s,i}}$ 与参考集原型关系 r_s 的相似性:

$$E_{\text{sim}}(i) = r_s \cdot E_{r_{s,i}}^T. \quad (9)$$

式中: $E_{\text{sim}}(i)$ 表示第 i 个邻居关系 r_s 的相似性分数,本文只保留相似分数最高的 T 个邻居关系。为此,使用噪声过滤的方式,过滤掉相似分数较低的,方法为

$$E'_{\text{sim}}(i) = \begin{cases} 1, & \text{当 } E_{\text{sim}}(i) \in \text{Top}(E_{\text{sim}}(i), T); \\ 0, & \text{其他。} \end{cases} \quad (10)$$

式中:函数 $\text{Top}(\cdot)$ 返回的是 $E_{\text{sim}}(i)$ 中相似性得分最高的 T 个(本文 T 设置为 3)邻居关系; $E'_{\text{sim}}(i)$ 表示噪声过滤后的邻居关系。接下来,根据噪声过滤后的邻居关系 $E'_{\text{sim}}(i)$ 进行编码:

$$r'_{s,i} = \sum W_{r_{s,i}} \cdot E'_{\text{sim}}(i); \quad (11)$$

$$W_{r_{s,i}} = \frac{\exp(\lambda \cdot \sigma(E'_{\text{sim}}(i)))}{\sum_{r_{s,i} \in R'_r} \exp(\lambda \cdot \sigma(E'_{\text{sim}}(i^T)))}. \quad (12)$$

式中: $r'_{s,i}$ 为融合了参考集实体对中噪声最小的邻居关系的表示; $W_{r_{s,i}}$ 为编码过程中的权重; $E'_{\text{sim}}(i^T)$ 为参考集实体对中每个邻居关系的表示; λ 为一个学习参数; $\sigma(\cdot)$ 为激活函数。最后,将参考集的原型关系 r_s 和 $r'_{s,i}$ 融合:

$$r' = \sigma(w_1 \cdot r_s + w_2 \cdot r'_{s,i}). \quad (13)$$

式中: $\sigma(\cdot)$ 为激活函数; w_1, w_2 均为可学习的参数。

2.4 Transformer 学习器

在 FKGC 任务中,核心目标是最后输出的关系表示 r 包含尽可能多的语义信息,从而提高小样本参考集中不同实体对的细粒度语义。本文使用具有强大学习能力的 Transformer 模型^[17] 作为学习器继续来学习三元组的关系表示。为了获得更准确的关系表示,FRLC 模型将高阶邻域实体编码器产生的实体嵌入和关系表示编码器生成的任务关系嵌入作为 Transformer 学习器的输入。这样的设计有助于进一步优化关系表示。受到 R-TLM^[18] 和 FRLN^[19] 的启发,本文采用了简化后的 R-TLM 模块来进行学习器的优化,如图 5 所示。

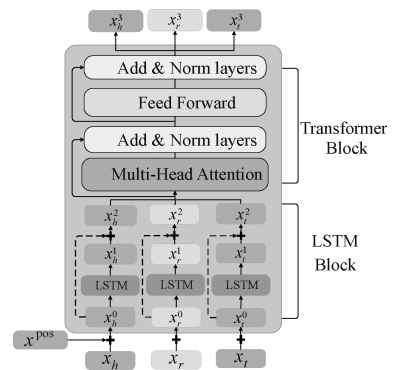


图 5 Transformer 学习器

Figure 5 Transformer learning framework

首先,将任务关系 r 和对应的实体对 (h, t) 表示为一个序列 $X = \{x_1, x_2, x_3\}$, 其中 x_1 和 x_3 分别表示头实体和尾实体, x_2 表示任务关系。定义 Transformer

的输入为 z_i^k , 对于序列 X 中的元素 x_i 表示如下:

$$z_i^k = x_i^{\text{emb}} + x_i^{\text{pos}}. \quad (14)$$

式中: x_i^{emb} 为元素的嵌入; x_i^{pos} 为位置嵌入。实体嵌入 x_1^{emb} 和 x_3^{emb} 是高阶邻域实体编码器更新后的实体表示, 关系嵌入 x_2^{emb} 来自关系表示编码器。将 z_i^0 输入到 LSTM 模块中, 得到隐藏状态 z_i^1 作为输出。接下来, 将 z_i^0 和 z_i^1 进行残差连接, 作为 Transformer 模块的输入, 进而生成学习器的输出 z_i^3 。具体计算步骤如下:

$$z_i^1 = \text{LSTM}(z_i^0); \quad (15)$$

$$z_i^2 = \text{Fusion}(z_i^0 + z_i^1); \quad (16)$$

$$z_i^3 = \text{TransformerBlock}(z_i^2). \quad (17)$$

将学习器的输出 z_i^3 作为任务三元组 (h, r, t) 的最终关系表示 $z(h, r, t)$ 。至此, 对于每一个小样本关系 r 和它的任务三元组都有对应的最终关系表示。

2.5 匹配处理计算

经过 Transformer 学习器的再学习, 参考集和查询集中的每个实体都得到了对应的关系表示。FR-LC 使用基于度量的方法计算查询三元组与参考集的语义相似性, 并将相似性分数最高的三元组作为模型预测结果。然而, 在匹配计算过程中, 每个参考集都蕴含着不同的信息, 针对不同的查询三元组应当动态学习参考关系 r 的表示。FR-LC 为每个查询实体对生成一个动态的查询关系表示 r_m :

$$r_m = \sum_{S_{r_i} \in S_r} \lambda_i \cdot S_{r_i}; \quad (18)$$

$$\lambda_i = \frac{\text{softmax}(\cos(S_{r_i}, q_r)) \cdot \exp(\cos(S_{r_i}, q_r))}{\sum_{S_{r_j} \in S_r} \exp(\cos(S_{r_j}, q_r))}. \quad (19)$$

式中: λ_i 为各个参考集关系表示的 $\text{softmax}(\cdot)$ 注意力权重; S_{r_i} 为参考集的第 i 个实体对经过 Transformer 学习器得到的关系表示; q_r 为查询实体对经过 Transformer 学习器得到的关系表示; $\cos(\cdot)$ 为余弦相似度; j 为参考集的大小。接下来, 计算动态的 r_m 和 q_r 的语义相似性分数:

$$\text{score}(r_m, q_r) = r_m \cdot q_r. \quad (20)$$

当 $\text{score}(\cdot)$ 的值增加时, 表示查询实体对 (h, t) 在参考关系 r 下与小样本参考集之间的语义相似性更高, 增加了查询三元组成立的可能性; 反之, $\text{score}(\cdot)$ 值较小时, 语义相似性减弱, 可能性降低。然而, 在应用平移模型 (如 TransE^[8]) 时, 还要充分考虑平移性对于 $h + r \approx t$ 的期望值的影响。根据式 (17), (z_h^3, z_r^3, z_t^3) 为查询实体对 (h, t_q) 经过 Trans-

former 学习器后的输出序列, 定义平移得分 s 为

$$s(h_q, t_q) = \|z_h^3 + z_r^3 - z_t^3\|_2. \quad (21)$$

式中: z 的 L_2 范数表示为 $\|z_i^3\|_2$; s 表示 $z_h^3 + z_r^3$ 与 z_t^3 之间的距离。因此, s 越小表示查询三元组成立的可能性就越大。综合考虑 $\text{score}(\cdot)$ 和 $s(\cdot)$ 的值可以得到待查询实体对 (h_q, t_q) 的最终匹配得分 $\text{score}'(\cdot)$:

$$\text{score}'(h_q, t_q) = \text{score}(r_m, q_r) - \delta s(h_q, t_q). \quad (22)$$

式中: δ 为调整因子。对所有的待查三元组进行计算, 得出最终的匹配得分 $\text{score}'(\cdot)$ 并进行排序。

2.6 模型训练

本文遵循其他 FKGC 的模型训练设置^[12]。给定一个任务关系 $r \in \mathcal{R}_{\text{train}}$ 和它的相关三元组, 在三元组集合中随机抽取 K 个三元组构建参考集 $S_r = \{(h_i, t_i) \mid (h_i, r, t_i) \in \mathcal{G}\}_i$, 并从剩余的三元组中随机抽取 B (batch size) 个三元组作为正查询集 $Q_r = \{(h_q, t_q) \mid (h_q, r, t_q) \in \mathcal{G}\}$ 。因为知识图谱本身不存在负三元组, 所以通过替换查询集 Q_r 中的三元组尾实体构建对应的负查询集 $Q_r^- = \{(h_q, t_q^-) \mid (h_q, r, t_q) \in \mathcal{G}, (h_q, r, t_q^-) \notin \mathcal{G}\}$ 。本文采用合页损失函数 (hinge loss) 进行训练:

$$\mathcal{L} = \sum_r \sum_{(h_q, t_q) \in Q_r} \sum_{(h_q, t_q^-) \in Q_r^-} [\gamma + \text{score}'(h_q, t_q^-) - \text{score}'(h_q, t_q)]^+. \quad (23)$$

式中: $[\cdot]^+ = \max(0, x)$ 为标准合页损失函数; γ 为 margin 超参数, 用来分隔正查询三元组和负查询三元组。最后, 本文采用基于批量抽样的元训练策略^[13], 在最小化损失函数 \mathcal{L} 的同时, 进行模型参数集的优化。为了避免过拟合, 还对模型参数集进行了 L_2 正则化。在优化过程中, 采用了 Adam^[20] 优化器进行模型的参数优化。该模型伪代码如算法 1 所示。

算法 1 FR-LC 的训练过程。

输入: 训练任务集 $T_{\text{meta-training}}$, TransE 知识图谱嵌入向量, 矩阵模型初始化参数 θ , 参考样本大小 K ;

输出: 保存模型的优化参数 W_c, λ_c, θ 。

- ① For epoch in 1 to M do
- ② 对 $T_{\text{meta-training}}$ 中的任务进行打乱
- ③ For T_r in $T_{\text{meta-training}}$ do
- ④ 抽取 K 个 G 中关系 r 的实体对构建小样本参考集 S_r
- ⑤ For k in K do
- ⑥ 增强头实体和尾实体的嵌入向量表示, 并更新小样本关系的表示
- ⑦ End For

- ⑧通过 Transformer 学习器处理三元组
- ⑨ $Q_r = \text{Sample}(r, G) - S_r$ //构建正三元组查询集
- ⑩ $Q_r^- = \text{Pollute}(Q_r)$ //污染正三元组的尾实体获得负三元组
- ⑪计算 Q_r 与 Q_r^- 中每个实体对与 S_r 的匹配得分
- ⑫计算损失函数 \mathcal{L}
- ⑬使用 Adam 优化器更新参数集
- ⑭End For
- ⑮End For

3 实验

本文在实验中使用 NELL-One 和 Wiki-One 这

2 个数据集进行链接预测实验,并与多个基准模型进行比较,最后基于实验结果评估 FRLC 模型的性能。

3.1 数据集与评价指标

本文使用的 2 个基准数据集 NELL-One 和 Wiki-One 都是小样本知识图谱补全任务的常用数据集。在这 2 个数据集中,将关联三元组的数量限定在 50 到 500 之间的关系定义为小样本关系(任务关系 $\mathcal{R}_{\text{task}}$),而与这些三元组相关的其他关系则构成了背景知识图 \mathcal{G}' 。按照 FKGC 任务的设定,根据比例分别将 NELL-One 和 Wiki-One 中的任务关系进行划分,划分比例为 51/5/11 和 133/16/34,分别对应训练关系/验证关系/测试关系($\mathcal{R}_{\text{train}} / \mathcal{R}_{\text{validation}} / \mathcal{R}_{\text{test}}$)。数据集的详细统计如表 1 所示。

表 1 数据集的统计信息
Table 1 Statistics of dataset

数据集	数量						
	实体	关系	三元组	任务关系	训练关系	验证关系	测试关系
NELL-One	68 545	58	181 109	67	51	5	11
Wiki-One	4 838 244	22	5 859 240	183	133	16	34

3.2 基准模型与参数设置

为了评估 FRLC 模型的有效性,本文选用传统知识图谱嵌入模型和 FKGC 模型这两类基准模型进行比较。选取传统知识图谱嵌入方法 TransE^[8], DISTMULT^[9], ComplEx^[10], Simple^[21] 和 RotatE^[22] 5 个模型作为对照模型。对比的 FKGC 模型主要包括 GMatching^[12], FSRL^[13], FAAN^[14], MetaR^[15] 和 GANA^[16]。其中, GMatching 模型采用 GMatching (MaxP)作为对比。它提出了一个邻居编码器和匹配处理器,并通过最大池化(max pooling)的方式完成小样本推理任务。MetaR 模型分为 MetaR (Pre-train)和 MetaR (In-train)这 2 种情况,其中 MetaR (Pre-train)是只使用背景知识图谱进行实体嵌入训练,而 MetaR (In-train)将背景知识图谱和原始训练集中的三元组进行抽样,并添加到模型训练过程当中。在对比实验中,传统知识图谱嵌入模型和小样本知识图谱补全模型按照各方法的最优参数设置,分别进行 5 次 5-shot 知识图谱补全实验,将 5 次实验结果取均值作为最终结果。

FRLC 模型使用 PyTorch 框架实现,并在单张 NVIDIA GeForce RTX 4090 GPU 上进行实验。在预训练嵌入模型的选择上采用 TransE 模型,并针对 NELL-One 和 Wiki-One 数据集进行了训练。具体的参数设置如下:实体和关系嵌入的维度 d 分别为 100 和 50;批处理大小 $batch\ size$ 为 128;初始学习率

lr 分别为 5×10^{-5} 和 6×10^{-5} ;最大邻居数量 M 固定为 150; γ 超参数设置为 5;使用 Adam 作为优化器。在训练过程的前 10^4 步,模型逐渐增加学习率,之后线性下降。每 10^4 训练步进行一次模型验证,最大训练步数设置为 3×10^5 。在模型验证过程中,取 MRR 最高的模型参数作为 FRLC 的最优训练结果。

3.3 性能比较

所有模型在 NELL-One 和 Wiki-One 数据集上的 5-shot 链接预测结果如表 2 所示。从表 2 中可以看出,与传统的知识图谱嵌入方法相比,FRLC 在 2 个数据集上均取得了较好的成绩。实验结果表明,FRLC 能够有效地预测小样本关系中的缺失实体。在 2 个数据集上,FRLC 模型在 4 个评价指标上均优于基准模型中的最佳结果。相比于 NELL-One 数据集上最佳表现的 MetaR (In-train)模型,FRLC 模型分别将 MRR 、 $Hits@10$ 、 $Hits@5$ 和 $Hits@1$ 指标提高了 2.9 百分点、1.9 百分点、3.1 百分点和 4.3 百分点。在 Wiki-One 数据集上,相较于其他次优模型,性能分别提高了 3.3 百分点、4.3 百分点、3.4 百分点和 3.2 百分点。值得注意的是 MetaR 仅有 Pre-train 或 In-train 中的一种设置在单个数据集上表现良好。这表明本文提出的 FRLC 模型对于不同的数据集有一定的泛化性,可以利用知识图谱中实体的上下文语义和结构信息提高小样本知识图谱补全的性能。

表 2 模型在 NELL-One 和 Wiki-One 上的 5-shot 链接预测结果

Table 2 5-shot link prediction results on NELL-One and Wiki-One

模型类别	模型	NELL-One				Wiki-One			
		<i>MRR</i>	<i>Hits@ 10</i>	<i>Hits@ 5</i>	<i>Hits@ 1</i>	<i>MRR</i>	<i>Hits@ 10</i>	<i>Hits@ 5</i>	<i>Hits@ 1</i>
传统模型	TransE	0. 174	0. 313	0. 231	0. 101	0. 133	0. 187	0. 157	0. 100
	DISTMULT	0. 200	0. 311	0. 251	0. 137	0. 071	0. 151	0. 099	0. 024
	ComplEx	0. 184	0. 297	0. 229	0. 118	0. 080	0. 181	0. 122	0. 032
	SimplE	0. 158	0. 285	0. 226	0. 097	0. 093	0. 180	0. 128	0. 043
	RotatE	0. 176	0. 329	0. 247	0. 101	0. 049	0. 090	0. 064	0. 026
小样本模型	GMatching(MaxP)	0. 176	0. 294	0. 233	0. 113	0. 263	0. 387	0. 337	0. 197
	MetaR(Pre-train)	0. 162	0. 282	0. 233	0. 101	0. 320	0. 443	0. 397	0. 262
	MetaR(In-train)	0. 308	0. 502	0. 423	0. 210	0. 229	0. 323	0. 289	0. 197
	FSRL	0. 269	0. 482	0. 369	0. 178	0. 221	0. 269	0. 183	0. 163
	FAAN	0. 265	0. 416	0. 347	0. 187	0. 314	0. 451	0. 384	0. 245
	GANa	0. 296	0. 497	0. 412	0. 194	0. 324	0. 437	0. 375	0. 264
	FRLC	0. 337	0. 521	0. 454	0. 253	0. 357	0. 494	0. 431	0. 296

注:加粗数字为最优结果,划线数字为次优结果,下同。

3.4 消融实验

FRLC 模型的框架主要由 3 个关键部分组成:基于门控机制的高阶邻域实体编码器、关系表示编码器和 Transformer 学习器。为了体现每个部分对 FRLC 的整体效果影响,本文在 Wiki-One 数据集上进行 5-shot 链接预测消融实验。表 3 给出了模型在移除其中一个部分后的对比实验结果。为了研究高阶邻域实体编码器中门控机制的有效性,首先,在 A1_a 中仅对一阶邻域实体进行编码输出;而在 A1_b 中移除了门控机制,并使用邻域实体嵌入的均值代替了 c_h ;其次,为了探究关系表示编码器的有效性,在 A2_a 中简单地将参考集实体对的平均嵌入作为关系的表示;最后,针对 Transformer 学习器的有效性,在 A3_a 中移除了 LSTM 模块,而在 A3_b 中则取消了 Transformer 模块的使用。

表 3 FRLC 在 Wiki-One 数据集上的 5-shot 链接预测消融实验

Table 3 Ablation study of FRLC with 5-shot settings on Wiki-One

消融模型	5-shot 链接预测			
	<i>MRR</i>	<i>Hits@ 10</i>	<i>Hits@ 5</i>	<i>Hits@ 1</i>
A1_a	0. 314	0. 443	0. 386	0. 258
A1_b	0. 336	0. 469	0. 395	0. 272
A2_a	0. 343	0. 483	0. 425	0. 267
A3_a	0. 331	0. 453	0. 383	0. 279
A3_b	0. 301	0. 432	0. 371	0. 264
FRLC	0. 357	0. 494	0. 431	0. 296

由表 3 可以看出,完整的 FRLC 模型在推理效果上均优于其各个变体。这说明基于门控机制的高阶邻域实体编码器可以有效利用中心实体的邻域信

息,并通过门控机制过滤掉噪声邻居的影响,提高了中心实体的语义表达;通过引入参考集中实体对的邻居关系表示,FRLC 可以提高关系嵌入的质量,有助于提高关系预测准确度,并降低对关系对实体的依赖性,提高模型的泛化能力;结合 LSTM 网络和 Transformer 模块的组合结构在 Transformer 学习器中表现出优势,相较于各自独立使用时的效果,LSTM 网络提升了细粒度的上下文语义表达,再和 Transformer 模块结合使用可以进一步提高链接预测的准确性。

3.5 参考集大小影响分析

为了体现参考集中实例数量 (few-shot size) 对模型性能的影响,本文在 Wiki-One 数据集上进行了对比实验,如图 6 所示。由图 6 可知,随着参考集实例数量 K 的增加,FRLC 在 *MRR* 和 *Hits@ 1* 上的数值逐步上升,所有模型的预测精度总体在提高,但在提高一定程度后趋于平缓又下降。这表明在预测新三元组时,参考集的大小是影响链接预测准确性的一个关键因素。

在以 *MRR* 为评价指标的实验中,FRLC 在参考集实例数量 K 为 6 时预测精度达到最高,其他部分模型也出现了类似的情况。在以 *Hits@ 1* 为评价指标的实验中,FRLC 在参考集实例数量 K 从 1 到 6 的条件下都优于其他模型。这验证了 FRLC 模型在小样本条件下的效果。随着参考集数量的增加,查询三元组能够查询更多的参考信息,链接预测准确度得以提升。但当参考集数量过多后,出现了预测精度处于平缓甚至下降的现象。这表明当有更多的参考信息时,模型在学习关系表示的过程中学习了

更多的无用信息,引入了不必要的噪声从而导致预测精度下降;其次,三元组的实体有了更多的解释和

意义,在进行关系学习的过程中复杂程度显著提升,这也造成了预测精度的降低。

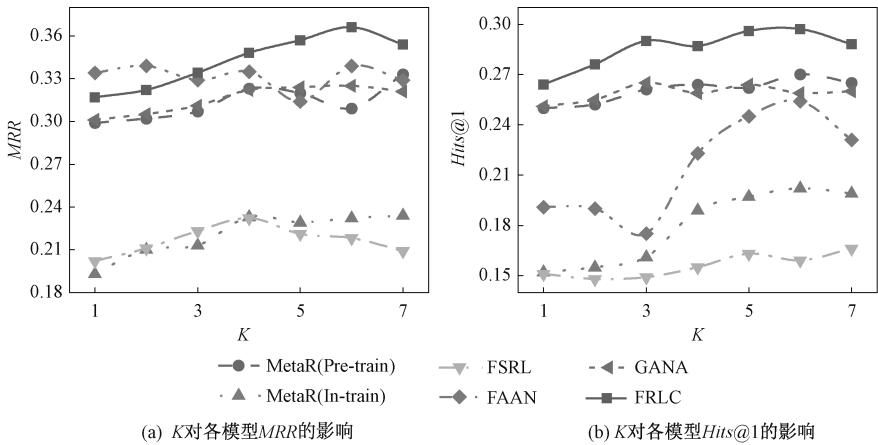


图 6 在 Wiki-One 上参考集尺寸对各模型性能影响
Figure 6 The impact of reference set size on the performance of various models on Wiki-One

4 结论

本文针对小样本知识图谱补全任务提出了基于关系学习的小样本知识图谱补全模型 FRLC。聚合高阶领域实体信息过程中引入门控机制,过滤邻域实体的噪声并进行聚合,提高实体编码质量;在学习关系表示过程中,FRLC 对参考集中实体对邻居关系所蕴含的信息进行学习,这提高了关系嵌入质量并降低对实体的依赖,提升模型泛化性;最后,在 Transformer 学习器中加入 LSTM 网络,进一步提高小样本关系的质量。实验结果表明,FRLC 在链接预测精度上优于基线 FKGC 模型。但知识图谱是动态变化的,模型的链接预测精度是否能够得到保障,这是一个具有挑战的问题。在未来的工作中,将通过引入外部知识的手段增强实体和关系的表示,例如使用关系和实体的文本描述来提高知识图谱的表征能力。

参考文献:

[1] CHEN X J, JIA S B, XIANG Y. A review: knowledge reasoning over knowledge graph[J]. Expert Systems with Applications, 2020, 141: 112948.

[2] 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: ACM, 2008: 1247-1250.

[3] SUCHANEK F M, KASNECI G, WEIKUM G. YAGO: a core of semantic knowledge[C] // Proceedings of the 16th International Conference on World Wide Web. New York: ACM, 2007: 697-706.

[4] BERANT J, CHOU A, FROSTIG R, et al. Semantic

parsing on freebase from question-answer pairs [C] // 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2013: 1533-1544.

[5] 左敏, 徐泽龙, 张青川, 等. 基于双维度中文语义分析的食品领域知识库问答[J]. 郑州大学学报(工学版), 2020, 41(3): 8-13.

ZUO M, XU Z L, ZHANG Q C, et al. A question answering model of food domain knowledge bases with two-dimension Chinese semantic analysis [J]. Journal of Zhengzhou University (Engineering Science), 2020, 41(3): 8-13.

[6] ZHANG F Z, YUAN N J, LIAN D F, et al. Collaborative knowledge base embedding for recommender systems [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 353-362.

[7] CHAMI I, WOLF A, JUAN D C, et al. Low-dimensional hyperbolic knowledge graph embeddings [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 6901-6914.

[8] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data [C] // Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 2787-2795.

[9] YANG B S, YIH W T, HE X D, et al. Embedding entities and relations for learning and inference in knowledge bases[EB/OL]. (2015-08-29) [2023-07-08]. <https://arxiv.org/abs/1412.6575>.

[10] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction [C] // Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York: ACM, 2016: 2071-2080.

- [11] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2018-02-04) [2023-07-08]. <https://arxiv.org/abs/1710.10903>.
- [12] XIONG W H, YU M, CHANG S Y, et al. One-shot relational learning for knowledge graphs [EB/OL]. (2018-08-27) [2023-07-08]. <https://arxiv.org/abs/1808.09040>.
- [13] ZHANG C X, YAO H X, HUANG C, et al. Few-shot knowledge graph completion [EB/OL]. (2019-11-26) [2023-07-08]. <https://arxiv.org/abs/1911.11298>.
- [14] SHENG J W, GUO S, CHEN Z Y, et al. Adaptive attentional network for few-shot knowledge graph completion [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2020: 1681-1691.
- [15] CHEN M Y, ZHANG W, ZHANG W, et al. Meta relational learning for few-shot link prediction in knowledge graphs [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2019: 4217-4226.
- [16] NIU G L, LI Y, TANG C G, et al. Relational learning with gated and attentive neighbor aggregator for few-shot knowledge graph completion [C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 213-222.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
- [18] SUN G, ZHANG C, WOODLAND P C. Transformer language models with LSTM-based cross-utterance information representation [C]//2021 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2021: 7363-7367.
- [19] 冉丈杰, 孙林夫, 邹益胜, 等. 基于关系学习网络的小样本知识图谱补全模型 [J]. 计算机工程, 2023, 49(9): 52-59.
- RAN Z J, SUN L F, ZOU Y S, et al. Few-shot knowledge graph completion model based on relation learning network [J]. Computer Engineering, 2023, 49(9): 52-59.
- [20] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2023-07-08]. <https://arxiv.org/abs/1412.6980>.
- [21] KAZEMI S M, POOLE D. Simple embedding for link prediction in knowledge graphs [EB/OL]. (2018-10-26) [2023-07-08]. <https://arxiv.org/abs/1802.04868>.
- [22] SUN Z Q, DENG Z H, NIE J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space [EB/OL]. (2019-02-26) [2023-07-08]. <https://arxiv.org/abs/1902.10197>.

Relation Learning Completion Model for Few-shot Knowledge Graphs

LI Weijun, GU Jianlai, ZHANG Xinyong, GAO Yuxiao, LIU Jintong

(School of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China)

Abstract: In few-shot knowledge graphs, the representation of relationships between entity pairs was diverse and complex. However, existing few-shot knowledge graph completion methods commonly suffered from insufficient relational learning capabilities and the neglect of contextual semantics associated with entities. To address these challenges, a novel approach called the few-shot relation learning completion model (FRLC) was proposed. Firstly, during the process of aggregating high-order neighborhood entity information, a gating mechanism was introduced to mitigate the adverse effects of noise on neighbors while enriching the representation of central entities. Secondly, in the phase of relation representation learning, the correlations among entity pairs in a reference set were leveraged to obtain more accurate relationship representations. Lastly, within the Transformer-based learning framework, an LSTM structure was incorporated to further capture contextual semantic information of entities and relationships, which was used for predicting new factual knowledge. To validate the effectiveness of FRLC, comparative experiments were conducted on the publicly available NELL-One and Wiki-One datasets, in which FRLC was compared with six few-shot knowledge graph completion models and five traditional models for 5-shot link prediction. The experimental results showed improvements in FRLC across four metrics: *MRR*, *Hits@10*, *Hits@5*, and *Hits@1*, demonstrating the model's effectiveness.

Keywords: knowledge graph completion; few-shot relation; neighborhood aggregation; link prediction