



计算机科学

COMPUTER SCIENCE

基于背景结构感知的小样本知识图谱补全

张静, 潘景豪, 姜文超

引用本文

张静, 潘景豪, 姜文超. 基于背景结构感知的小样本知识图谱补全[J]. 计算机科学, 2026, 53(2): 331-341.

ZHANG Jing, PAN Jinghao, JIANG Wenchao. [Background Structure-aware Few-shot Knowledge Graph Completion](#) [J]. Computer Science, 2026, 53(2): 331-341.

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)

[深度融合句法和语义特征的情感三元组片段级抽取方法](#)

Method for Span-level Sentiment Triplet Extraction by Deeply Integrating Syntactic and Semantic Features

计算机科学, 2026, 53(2): 322-330. <https://doi.org/10.11896/jsjcx.250100061>

[语义引导的红外与可见光图像混合交叉特征融合方法](#)

Semantic-guided Hybrid Cross-feature Fusion Method for Infrared and Visible Light Images

计算机科学, 2026, 53(2): 253-263. <https://doi.org/10.11896/jsjcx.250100123>

[基于注意力机制的音频驱动数字人脸视频生成方法](#)

Attention-based Audio-driven Digital Face Video Generation Method

计算机科学, 2026, 53(2): 245-252. <https://doi.org/10.11896/jsjcx.241200067>

[双支特征融合的带约束的多损失视频异常检测](#)

Constrained Multi-loss Video Anomaly Detection with Dual-branch Feature Fusion

计算机科学, 2026, 53(2): 236-244. <https://doi.org/10.11896/jsjcx.250300103>

[基于时频域注意力的时间序列异常检测模型](#)

Time-Frequency Attention Based Model for Time Series Anomaly Detection

计算机科学, 2026, 53(2): 161-169. <https://doi.org/10.11896/jsjcx.241200106>

基于背景结构感知的小样本知识图谱补全

张静 潘景豪 姜文超

广东工业大学计算机学院 广州 510006

(844840296@qq.com)

摘要 小样本知识图谱补全旨在通过少量参考数据预测知识图谱中长尾关系的未知事实。如何在数据稀疏条件下高效编码实体和关系特征并构建有效的三元组评分函数对补全效果影响显著。现有的小样本知识图谱补全模型忽略了实体上下文背景结构信息对实体编码和评分函数的影响,导致关系表示学习能力不足。针对上述问题,提出了一种基于背景结构感知的小样本知识图谱补全模型(BSA)。首先,设计了一种实体对上下文背景结构信息交互指标,通过衡量邻居实体在结构上的影响,指导模型将注意力集中在与中心实体结构更相似的邻居节点,以减少噪声邻居的不良影响。其次,在关系表示学习阶段,引入背景知识图谱中语义和结构相似的关系信息进一步增强目标关系的嵌入表示。最后,在评分函数中引入头尾实体对的上下文信息交互指标,提升模型对复杂关系的推理能力。实验结果表明,与当前主流方法相比,BSA模型在 NELL-One 数据集测试中, MRR, Hit@5 和 Hit@1 评价指标分别提高了 0.4 个百分点,0.8 个百分点和 0.5 个百分点。在 Wiki-One 数据集测试中, MRR, Hit@10 和 Hit@5 指标分别提高了 1.9 个百分点,2.2 个百分点和 2.2 个百分点,充分证明了 BSA 模型的有效性。

关键词: 小样本知识图谱补全;背景结构感知;表示学习;注意力机制

中图分类号 TP391.1

Background Structure-aware Few-shot Knowledge Graph Completion

ZHANG Jing, PAN Jinghao and JIANG Wenchao

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China

Abstract Few-shot knowledge graph completion aims to predict unseen facts in long-tail relationships within knowledge graphs using only a small number of reference data. The key challenge of this task lies in how to efficiently encode entity and relation features under conditions of data scarcity, and to construct an effective triplet scoring function. Existing few-shot knowledge graph completion models generally overlook the impact of entity pair contextual information on both entity encoding and the scoring function, while also suffering from insufficient relation representation learning. To address these issues, this paper proposes a background-structure-aware few-shot knowledge graph completion model—BSA. Firstly, it designs a metric for entity pair contextual interaction, which guides the model to focus attention on neighbor nodes that are structurally similar to the central entity by measuring the structural influence of neighboring entities, thereby reducing the negative impact of noisy neighbors. Secondly, during the relation representation learning phase, it incorporates background relation information from the knowledge graph that is semantically and structurally similar to the target relation to enhance its embedding representation. Finally, it introduces a contextual interaction metric for the head-tail entity pair in the triplet scoring function to improve the model's reasoning capability for complex relations. Experimental results show that, compared to the best results from baseline models, the BSA model improves MRR, Hit@5, and Hit@1 by 0.4 percentage points, 0.8 percentage points, and 0.5 percentage points on the NELL-One dataset, respectively, and improves MRR, Hit@10, and Hit@5 by 1.9 percentage points, 2.2 percentage points, and 2.2 percentage points on the Wiki-One dataset, respectively, demonstrating the effectiveness and feasibility of the proposed method.

Keywords Few-shot knowledge graph completion, Background structure-aware, Representation learning, Attention mechanism

1 引言

知识图谱是由谷歌提出的一种用图来表示知识结构化的语

义的知识库^[1]。知识图谱由大量三元组构成,知识以(头实体,关系,尾实体)三元组形式作为图结构的一部分存入数据库,使得原本分散在不同体系的知识能以图结构形式联系在

到稿日期:2025-01-16 返修日期:2025-06-13

基金项目:国家自然科学基金重点项目(62237001);广东省自然科学基金(2024A1515011502);珠海市科技计划(2320004002758)

This work was supported by the National Natural Science Foundation of China for Key Program (62237001), Natural Science Foundation of Guangdong Province, China (2024A1515011502) and Zhuhai Science and Technology Plan (2320004002758).

通信作者:姜文超(jiangwenchao@gdut.edu.cn)

一起。异构知识结构化的特性让知识图谱在现实生活中得到广泛应用。知识图谱数据如 YAGO^[2], Freebase^[3], DBpedia^[4] 被广泛用于智能问答^[5]、推荐系统^[6]、信息检索^[7] 等领域。然而,由于知识抽取方法的局限性和知识库动态更新特性,知识图谱普遍存在数据稀疏以及数据不完整问题^[8]。数据稀疏问题的产生主要是由于信息表达带有偏向性。信息的数量和质量与实体的关注度呈正相关,不同类型实体的关注度不同,关注度越高的实体,包含的数据越多。研究表明,真实的知识图谱数据大多呈现长尾分布^[9],例如在数据集 Wikidata 中,只有 10% 的关系拥有超过 10 个三元组实例。为了提升知识图谱的数据质量,知识图谱补全 (Knowledge Graph Completion, KGC) 已成为一个重要方向。

目前知识图谱补全的主流方法为应用知识图谱嵌入技术,代表模型有 TransE^[10], TransH^[11], DistMult^[12], ComplEx^[13] 等。这类模型的核心思想是,在同一个低维向量空间中表示出知识图谱的实体以及关系的嵌入向量,从而基于嵌入向量度量出三元组的合理性^[10]。然而,基于知识图谱嵌入的方法在训练时,每条关系需要参考大量的同类型三元组来学习查询三元组的相关嵌入^[14]。此外,在社交媒体、推荐系统等实际应用过程中,知识图谱的动态更新会使得新关系通常只有少量的三元组实例^[15],大部分关系只包含少量的参考三元组,导致传统的知识图谱补全模型无法起到很好的效果。

为了解决数据稀疏以及不完整问题导致的知识图谱补全效果不佳的问题,小样本条件下的知识图谱补全 (Few-shot Knowledge Graph Completion, FKGC) 问题成为新的研究热点。FKGC 的任务是:在给定 K 条任务关系 r 参考三元组的条件下,预测出查询三元组中缺失的实体或关系信息^[16]。现有模型通常分为基于度量学习的方法和基于优化的元学习方法两大类。两者的关键区别在于:度量学习方法通过建立统一的度量空间来直接计算三元组之间的相似度,进而进行尾实体预测^[17];优化方法通过微调支持集来调整模型,使其适应特定任务的关系,并在此基础上对查询三元组进行尾实体预测^[18]。

尽管现有模型在 FKGC 任务上取得了良好的效果,但也存在一些不足。1) 噪声邻居问题。研究表明,许多真实的图数据中大部分邻居实体与中心实体无关^[19]。传统的邻域编码器在聚合邻域信息时,聚合无关邻居会引入大量无关信息甚至噪音信息,从而降低模型的补全效果^[20]。2) 任务关系 r 的嵌入表示问题。传统方法依赖参考集中的已知实体对学习任务关系的嵌入表示,这可能导致目标三元组的关系嵌入过度依赖实体,降低了模型的有效性。3) 这些方法忽略了实体对结构相似度对实体聚合和评分函数的影响,这可能进一步影响三元组预测的准确性。

为了解决上述问题,提出一个基于背景结构感知的小样本知识图谱补全模型 (Background Structure-Aware Few-shot Knowledge Graph Completion, BSA)。针对实体噪声邻居问题,BSA 设计了一种实体对上下文信息交互指标,通过在结构上衡量邻居实体对于中心实体的影响,指导模型将注意力集中在与中心实体结构更相似的邻居节点上,降低了噪声邻居的影响;同时,BSA 模型改进了传统的任务关系学习方法,

引入背景知识图谱中语义和结构上相似的背景关系信息以增强目标关系的嵌入表示,进一步提高了模型链接预测的准确性;最后,BSA 模型在三元组的打分函数中引入了头尾实体对的结构相似性评估,在综合考虑语义相似性和结构相似性的基础上提升了查询三元组得分的准确性和模型对复杂关系的推理能力。实验结果表明,BSA 模型在 5-shot 条件下的表现显著优于基准模型 SAIA,并在多项指标上超越了最新模型 FCC。具体而言,在 NELL-One 数据集上,BSA 模型在 MRR, Hit@10, Hit@5 和 Hit@1 指标上较 SAIA 模型分别提高了 3.2 个百分点, 3.4 个百分点, 3.4 个百分点和 2.5 个百分点;在 Wiki-One 数据集上,较 SAIA 模型分别提高了 0.7 个百分点, 0.6 个百分点, 0.6 个百分点和 0.8 个百分点。实验结果验证了 BSA 模型的有效性。

2 相关工作

现有的小样本知识图谱补全模型主要有基于度量学习的方法和基于优化的元学习方法、基于其他模型的方法 3 类。

1) 基于度量学习的方法。这类方法主要采用编码器-解码器结构,模型通过邻域编码器从实体对的邻域中学习实体的语义特征,获得实体对的嵌入表示,再通过解码器评估查询实体对对于参考实体对的相似性分数来作为模型的评价标准。GMatching^[16] 首次针对 one-shot 场景研究小样本知识图谱补全问题,提出使用邻域编码器学习查询实体对的实体嵌入,并通过 LSTM 核心的匹配网络学习查询实体对和参考集的相似性。FSRL^[21] 进一步拓展到了 few-shot 场景,在邻域编码器中引入了注意力机制,通过有选择地学习邻域实体的信息,提高了实体嵌入的质量。FAAN^[17] 提出了不同关系对相邻实体聚合具有不同影响,创新性地应用动态注意力机制改进了邻域编码器,并引入 Transformer 改进了传统的匹配网络算法。CIAN^[22] 提出需要关注实体对间的交互信息,并设计了任务感知模块和实体对感知模块来学习实体间的交互注意力,为每个实体提取与任务关系相关的邻居关系语义信息。Wang 等^[23] 提出了一个基于语义和不确定性信息度量的度量框架,将实体和关系的固有不确定性纳入不确定性嵌入过程中。FCC^[24] 针对稀疏邻居问题设计了一个双分支特征提取器用于捕获实体的互补和完整表示,不同分支的特征提取器关注实体特征的不同方面。同时,FCC 方法还设计了一个结构化关系学习模块,用于捕获实体间的结构化交互表示。

2) 基于优化的元学习方法。这类方法受到了小样本学习领域中元学习方法的启发,专注于设计学习关系元信息的元学习器,以提高模型的学习能力。MetaR^[18] 首次应用元学习方法解决小样本知识图谱补全问题,它定义了关系元和梯度元两种元信息,并设计了关系元学习和嵌入学习两个模块,提高了模型的学习能力。GANA^[25] 参考了度量学习的方法,在 MetaR 的基础上增加了一组邻域编码器,通过引入门控和注意力机制提高了实体嵌入的质量。同时,GANA 深入研究了知识图谱中 1-N, N-1 以及 N-N 等多种复杂关系,并设计了一个建模的元学习方法。GL-GFKG^[26] 提出通过任务感知注意机制捕获基于关系路径的远程邻居特征和局部邻居特征,然后结合自适应高斯混合模型更好地表示完成任务。

随后,HiRE^[27]提出捕获层次关系信息,即实体级、三元组级和上下文级,并使用对比损失将三元组拉近其实际上下文,最后基于 TransD 的元评分函数来捕获实体和关系的多样性。NP-FKGC^[28]则提出通过一种归一化的基于流的神经过程来预测缺失的事实,同时估计不确定性,进一步引入一种基于注意力的关系建模方法来提高模型性能。

3)除了上述两种主流方法外,还有研究者正在拓展其他方法。P-INT^[29]设计了一种实体对的路径交互评估算法,利用从头实体到尾实体的路径,使用路径中的关系集合表示实体对,再计算参考实体对和查询实体对之间的路径交互。TransAM^[30]设计了一种 Transformer 匹配模块,通过将查询实体对附加到序列化的参考实体序列中,并利用 Transformer 捕获三元组实体内部和三元组实体之间的交互来计算概率。Liang 等^[31]提出一个用于学习少样本关系的多视图框架,除了 KGC 中基于三元组的结构视图,还添加了文本视图和常识视图,用来获得更丰富的语义表示,以及对复杂关系构建高质量的负采样。Informix-FKGC^[32]创新性地邻域编码器和匹配网络中引入了胶囊网络和神经张量网络。首先基于单跳邻居选择策略对实体的单跳邻居、属性和文字描述等多方面信息进行编码;然后利用胶囊网络对支持集进行聚合得到少镜头关系嵌入;最后使用神经张量网络对查询集和支持集进行匹配。此外还有基于图神经网络的方法,代表模型有 R-GCN^[33],KBGAT^[34]等。这类模型采用了编码器-解码器的结构,通过图神经网络强大的学习和表达能力对知识图谱进行建模,实现对知识图谱缺失信息的预测。R-GCN 使用图卷积操作对实体局部邻居信息进行聚合的方法在 FKGC 领域得到了广泛应用,是目前邻域编码器的主流方法之一。

上述方法在 FKGC 任务上都取得了良好的效果,但也存在一些问题。对于噪声邻居问题,该领域的主流方法^[17,22,24]是通过调整目标关系与邻接关系的交互注意力来降低噪声邻居的影响,忽略了实体对结构相似度对邻域编码的影响。

同时,在关系嵌入的学习方法问题上,传统方法大多依赖实体对的交互去学习对应关系的嵌入表示,这会导致目标三元组的关系嵌入对实体过度依赖,降低了模型的有效性。

3 模型的设计与实现

3.1 问题描述

由大量三元组构成的知识图谱可表示为 $G = \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$,其中 \mathcal{E} 和 \mathcal{R} 分别表示知识图谱的实体集与关系集。在给定三元组部分实体与关系的基础上,知识图谱补全任务的目的是预测三元组缺失部分的信息。本文的工作集中在预测 $(h, r, ?)$ 中缺失的尾实体。

遵循 FKGC 任务的常规设定^[16,21],将知识图谱 G 中的关系 \mathcal{R} 分为背景关系集 \mathcal{R}_{bg} 和任务关系集 \mathcal{R}_{task} ,并将知识图谱移除所有任务关系后的子集定义为背景知识图谱 G' 。任务关系集 \mathcal{R}_{task} 分为 \mathcal{R}_{train} , \mathcal{R}_{test} 和 \mathcal{R}_{valid} 3 部分,分别用于元训练、元测试和元验证 3 个阶段。遵循小样本学习的设置,在元训练阶段,对于每一条任务关系 $r \in \mathcal{R}_{train}$,将其关联三元组随机划分为参考集 $S_r = \{(h_i, t_i)\}_{i=1}^K$ 和查询集 $Q_r = \{(h_j, t_j, C_{h_j, r})\}_{j=1}^B$ 。其中, K 和 B 表示参考集和查询集包含的实体对个数, t_j 是查询三元组 $(h_j, r, ?)$ 的真实尾实体,而 $C_{h_j, r}$ 是其对应候选尾实体集,每一个查询三元组的候选尾实体都基于实体类型约束构建。对于查询三元组 $(h_j, r, ?)$,模型根据知识图谱 G 和参考集 S_r 提供的信息对真实尾实体 t_j 和候选尾实体 $t_j' \in C_{h_j, r}$ 进行排序,使得真实尾实体 t_j 的排序高于所有候选尾实体。

构造 BSA 模型的目的是解决小样本知识图谱补全 (FKGC) 中数据稀疏性导致的噪声邻居干扰和关系表示不足问题。具体而言,现有方法在聚合邻域信息时难以区分无关实体,导致编码质量下降;同时,目标关系的嵌入过度依赖有限参考实例,忽视了背景知识图谱中语义和结构相似的背景关系。如图 1 所示,BSA 模型通过以下模块实现目标。

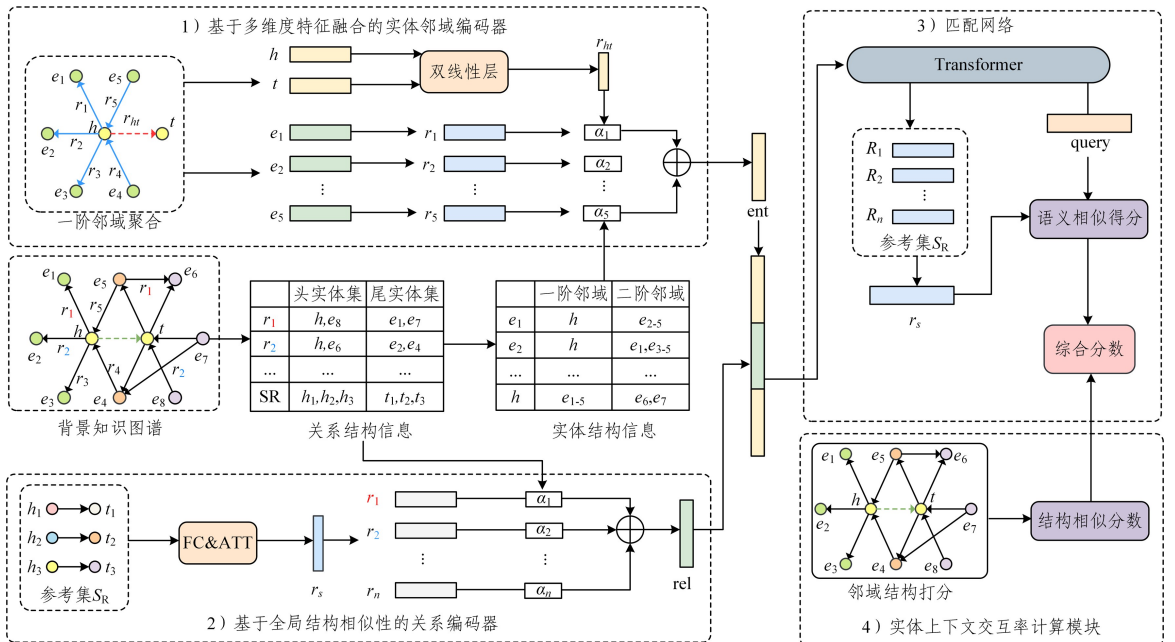


图 1 BSA 模型的整体框架

Fig. 1 Overall framework of BSA

1) 基于多维度特征融合的实体邻域编码器。在传统实体编码器的基础上引入实体对上下文交互率(E-CIR),用于衡量中心实体与邻居实体的结构相似性。模型优先聚合高相似性邻居的语义信息,抑制噪声干扰。

2) 基于全局结构相似性的关系编码器。从背景知识图谱中筛选与目标关系语义和结构双相似的关联关系,加权融合其嵌入表示,增强目标关系的泛化能力。

3) 匹配网络。结合 Transformer 编码语义关联与 E-CIR 的结构相似性评估,联合优化查询三元组的语义匹配度和结构一致性。

4) 实体上下文交互率计算模块。该模块从邻域结构信息的 3 个组成部分综合计算实体对的交互率,为模块 1 和模块 3 提供结构感知基础。

3.2 实体对邻域结构打分模块

知识图谱中,节点的邻域信息通常包含语义信息和结构信息。文献[35]表明,FKGC 任务中只使用 K 条参考三元组的措施,限制了模型使用细粒度的关系语义信息,链接预测的效果更加依赖于实体对粗粒度的位置信息。基于此,SAIA^[36]证明了在实体对度量函数中引入邻域结构的相似性分数有助于提高链接预测的准确性,并提出实体对公共交互率(CIR)用于衡量实体对在二阶范围内的交互程度。SLAN^[37]首次在实体的邻域聚合中引入图上下文信息,证明了其对知识图谱补全的有效性,并且把图上下文信息定义为邻居上下文和关系上下文两部分。SLAN 指出实体的语义不仅由其邻居实体决定,还受其关联关系的影响。不同的实体若共享相同关系,那么即使它们的实体邻居不重叠,也可能因关系语义的相似性而具有潜在关联。

基于此,本文设计了一种实体对上下文信息交互率(E-CIR),在 SAIA 的实体对公共交互率计算模块基础上,增加了实体对关系上下文相似度的计算,以更加全面地衡量实体对在二阶范围内的交互程度。在少样本场景下,传统结构指标(如 CIR)因邻居稀疏性而易失效,E-CIR 通过语义相似性补充稀疏结构,即使实体邻居不重叠,仍可基于关系语义相似性识别关联性,从而在数据稀缺时提升链接预测的鲁棒性。

3.2.1 实体上下文信息交互率

给定一个关于任务关系 r 的目标三元组 (h, r, t) ,实体对 (h, t) 的邻域定义为 $\{h, t, \mathcal{N}_h, \mathcal{N}_t\}$,其中,实体 h 和 t 的邻域分别表示为 $\mathcal{N}_h = \{(r_i, e_i) \mid (h, r_i, e_i) \in \mathcal{G}'\}$ 和 $\mathcal{N}_t = \{(r_j, e_j) \mid (t, r_j, e_j) \in \mathcal{G}'\}$, \mathcal{G}' 表示背景知识图谱,需分别计算实体 h 和 t 对于另一实体邻域的交互指数。这里将邻域交互指数分为实体上下文、关系上下文和邻居上下文 3 部分,从邻域结构信息的 3 个组成部分综合考虑实体对 (h, t) 的交互程度。

以实体 h 为例,首先,对于每一个 $e_{h_i} \in \mathcal{N}_h$,获取 e_{h_i} 的邻域结构信息 $\{\mathcal{N}_e(e_{h_i}), \mathcal{N}_r(e_{h_i}), \mathcal{N}_n(e_{h_i})\}$ 。其中, $\mathcal{N}_e(e_{h_i})$ 表示实体 e_{h_i} 的一阶邻居实体集合; $\mathcal{N}_r(e_{h_i})$ 则表示实体 e_{h_i} 的一阶邻居关系集合; $\mathcal{N}_n(e_{h_i})$ 将实体和关系结合,表示实体 e_{h_i} 的一阶邻居集合。实体对的结构信息表现为离散的实体和关系集合,使用 Jaccard 相似度可直接衡量集合的重叠度,无需进行向量化处理。实体 e_{h_i} 的邻域结构与实体 t 的邻域结构的 Jaccard 相似度计算如式(1)一式(3)所示:

$$S_e(e_{h_i}, t) = \frac{|\mathcal{N}_e(e_{h_i}) \cap \mathcal{N}_e(t)|}{|\mathcal{N}_e(e_{h_i}) \cup \mathcal{N}_e(t)|} \quad (1)$$

$$S_r(e_{h_i}, t) = \frac{|\mathcal{N}_r(e_{h_i}) \cap \mathcal{N}_r(t)|}{|\mathcal{N}_r(e_{h_i}) \cup \mathcal{N}_r(t)|} \quad (2)$$

$$S_n(e_{h_i}, t) = \frac{|\mathcal{N}_n(e_{h_i}) \cap \mathcal{N}_n(t)|}{|\mathcal{N}_n(e_{h_i}) \cup \mathcal{N}_n(t)|} \quad (3)$$

$S_e(e_{h_i}, t)$ 和 $S_r(e_{h_i}, t)$ 从实体和关系两个角度衡量目标实体 h 的一阶邻居 e_{h_i} 对于实体 t 的相似度; $S_n(e_{h_i}, t)$ 表示基础的上下文相似度。最后,为了增强 $S_e(e_{h_i}, t)$ 和 $S_r(e_{h_i}, t)$ 的表达能力,基于信息论的原理获得实体对 (e_{h_i}, t) 的上下文相似度,计算式如式(4)所示:

$$S_c(e_{h_i}, t) = \frac{S_n(e_{h_i}, t) + \alpha \cdot S_e(e_{h_i}, t) + \beta \cdot S_r(e_{h_i}, t)}{1 + \alpha + \beta} \quad (4)$$

其中, α 和 β 分别表示 $S_e(e_{h_i}, t)$ 和 $S_r(e_{h_i}, t)$ 对于 $S_n(e_{h_i}, t)$ 的权重。本文方法的权重设计借鉴了熵归一化思想,通过将实体和关系的局部信息量(邻居数量的对数)归一化为全局信息量的比例,实现自适应的权重分配。采用对数变换是为了抑制极端值的影响,使得邻居数量在不同数量级时仍能保持平衡的贡献比例。当实体数量 n_e 较大时,实体权重 α 升高,反之亦然。

$$\alpha = \frac{\lg(n_e)}{\lg(n_e \cdot n_r)} \quad (5)$$

$$\beta = \frac{\lg(n_r)}{\lg(n_e \cdot n_r)} \quad (6)$$

最后,运用度量公式 $S_c(e_{h_i}, t)$ 计算出实体 h 对于实体 t 的交互率 $\Gamma(h, t)$,如式(7)和式(8)所示:

$$\Gamma(h, t) = \sum_{e_i \in \mathcal{N}_h} S_c(e_i, t) \quad (7)$$

$$E_CIR(h, t) = \Gamma(h, t) + \Gamma(t, h) \quad (8)$$

3.2.2 搜索效率分析

E-CIR 模块的效率问题主要体现在其时间复杂度上。实体对上下文信息交互率的计算需要对知识图谱中所有实体对的邻居上下文进行相似度比较^[37]。假设知识图谱包含 n_e 个实体,每个实体的邻居上下文中平均包含 d_1 个邻居,其中又包含 d_2 个实体和 d_3 个关系,则上下文相似度(见式(1)一式(4))的时间复杂度为 $O(n_e^2 \cdot (d_1 + d_2 + d_3))$ 。进一步对头尾实体的邻居相似度进行加权聚合(见式(7)一式(8)),时间复杂度为 $O(n_e^2 \cdot (d_1 + d_2 + d_3) \cdot M)$,其中 M 表示每个实体的最大邻居数量。

这一复杂度在大型知识图谱(如 Wiki-One 包含 483 万个实体)中难以直接应用。为优化效率,本文提出基于预计算的字典存储策略,训练开始前将频繁访问的 E-CIR 值存储为字典,训练时通过读取字典获得实体对的 E-CIR 值。本质上,E-CIR 值是结构相似性的度量方法,其依赖知识图谱的拓扑结构而非训练参数。该方法可将训练时的动态计算开销 $O(n_e^2 \cdot (d_1 + d_2 + d_3) \cdot M)$ 转换为静态存储查询 $O(1)$,符合“以空间换时间”的经典优化原则。

3.3 基于多维度特征融合的实体邻域编码器

在 FKGC 任务中,三元组打分函数和关系学习模块都依赖于优质的实体表征,如何从邻域中更好地学习到实体表征是 FKGC 任务的重要课题。现有的 FKGC 研究^[17,22,24]都倾向于从实体的一阶邻域中学习实体的嵌入表示。然而,这样

做一方面会过于依赖嵌入模型学习到的实体嵌入表示,另一方面也忽视了实体邻域拓扑结构信息在筛除噪声节点方面的作用,而这一信息在知识图谱嵌入、知识图谱对齐等领域已被证明具有重要价值^[37]。文献[19-20]表明,在实体的一阶邻域中有很大一部分节点与中心节点无关,这类噪声节点在聚合过程中也会被分配一定程度的注意力,最终影响实体的整体编码过程。引入实体邻域丰富的结构信息可以使邻域聚合器的注意力集中在结构相似的邻居实体上,减轻噪声邻居对实体编码过程的影响。基于此,设计了一个基于多维度特征融合的实体邻域编码器(ESNE),在SAIA模型^[36]的基础上,引入实体上下文信息交互率学习更优质的实体表征。

给定一组实体对 (h, t) ,邻域编码器通过聚合各个实体的邻域信息来更新中心实体的嵌入表示。假设实体 h 为目标实体,则将实体 h 的邻域信息定义为 $\mathcal{N}_h = \{(r_i, e_i) | (h, r_i, e_i) \in \mathcal{G}'\}$ 。其中, (r_i, e_i) 表示实体 h 的一阶邻居实体 e_i 及其对应关系 r_i 。对于实体对 (h, t) ,首先定义了一个双线性函数来对任务关系 r 进行初步建模。考虑到关系和反向关系通常具有不同含义,将任务关系 r 分为了正向关系 r_{ht} 和反向关系 r_{th} ,如式(9)所示:

$$[r_{ht} \parallel r_{th}] = \text{Bilinear}(h, t) \quad (9)$$

其中, \parallel 表示拼接方法, h 和 t 表示实体 h 和 t 的预训练嵌入表示。将正向关系 r_{ht} 和反向关系 r_{th} 分别用于头实体 h 和尾实体 t 的补全过程,然后通过一个度量函数 $\varphi(\cdot)$ 评估任务关系 r 与邻域关系 $r_i \in \mathcal{N}_h$ 的相似性,如式(10)所示:

$$\varphi(r_{ht}, r_i) = r_{ht}^T \mathbf{W}_2 r_i + \mathbf{b} \quad (10)$$

其中, r_{ht} 为任务关系 r 的初始嵌入表示, r_i 为邻域关系 r_i 的预训练嵌入, $\mathbf{W}_2 \in \mathbb{R}^{d \times d}$ 和 $\mathbf{b} \in \mathbb{R}$ 分别表示可训练的权重矩阵和偏置向量。为了提高邻域编码器的编码质量,降低噪声邻居对编码过程的影响,在度量函数中引入3.2.1节设计的实体对上下文信息交互率,并计算中心实体 h 对邻居实体 e_i 的初始注意力分数 f_i ,如式(11)和式(12)所示:

$$\text{Score}(h, e_i) = \varphi(r_{ht}, r_i) \cdot (1 + S(h, e_i)) \quad (11)$$

$$f_i = \frac{\exp(\text{Score}(h, e_i))}{\sum_{r_j \in \mathcal{N}_h} \exp(\text{Score}(h, e_j))} \quad (12)$$

其中, $S(h, e_i)$ 表示中心实体 h 与邻居实体 e_i 的结构相似分数。由式(11)可知, $S(h, e_i)$ 越大,中心实体对于邻居节点的相似性分数 $\text{Score}(h, e_i)$ 越高,中心实体 h 分配的注意力分数就越大,邻域编码器的注意力集中于邻域中结构相似的实体上。

为了进一步降低噪声节点的影响,在计算注意力权重 ω_i 的过程中引入了“收缩理解范围”的机制。通过引入关系相似性度量函数,限制了邻域编码器的关注范围,如式(13)和式(14)所示:

$$\omega_i = \frac{f_i \cdot \exp(-\lambda_2 \cdot d(r_i, r_{ht}))}{\sum_{r_j \in \mathcal{N}_h} f_j \cdot \exp(-\lambda_2 \cdot d(r_j, r_{ht}))} \quad (13)$$

$$d(r, r_i) = \sqrt{\sum_{k=1}^m (r_k - r_{ik})^2} \quad (14)$$

其中, $d(r_i, r)$ 表示邻居关系 r_i 和目标关系 r 的相异度量函数, λ_2 是控制 $d(r_i, r)$ 比重的超参数。相异度 $d(r_i, r)$ 值越大,则对应实体的注意力权重 ω_i 越低。由此,可获得中心实体 h 的邻域嵌入 c_h 。邻域编码器通过融合邻域嵌入 c_h 和中心实体

h 的初始嵌入 h 获得更新后的中心实体嵌入 e_h ,如式(15)和式(16)所示:

$$c_h = \sum_{e_i \in \mathcal{N}_h} \omega_i \cdot e_i \quad (15)$$

$$e_h = \sigma(\mathbf{W}_1 \cdot c_h + \mathbf{W}_2 \cdot h) \quad (16)$$

其中, $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d \times d}$ 表示可训练的权重矩阵, h 和 e_i 表示中心实体和邻居实体的初始嵌入, $\sigma(\cdot)$ 表示激活函数。对尾实体 t 进行相同的操作,可获得更新后的实体对嵌入表示 (e_h, e_t) 。

3.4 基于全局结构相似性的关系编码器

对于FKGC任务,任务关系在三元组的打分模块中发挥着重要作用,如何获得关系的优质嵌入表示是FKGC的重要任务。过去的研究^[17,18,22,24,28]倾向于使用更新后的实体对嵌入来学习关系的语义表示,这种做法依赖于邻域编码器学习到的实体嵌入,容易造成过拟合,同时忽略了背景知识图谱中丰富的背景关系对于任务关系 r 的补全作用。基于此,提出了一个基于全局结构相似性的关系编码器(Global Structural Similarity based Relation Encoder, GSRE),从背景知识图谱中抽取语义和结构上相似的背景关系补全任务关系 r 。

对于任务关系 r ,首先使用更新后的实体对嵌入表示 (e_h, e_t) 进行建模。遵循嵌入模型TransE^[11]的关系定义,这里将关系 r 定义为实体对 (h, t) 的翻译操作,并应用注意力机制聚合参考集中各实体对的关系嵌入表示,获得参考集任务关系的语义表示 r_s ,如式(17)和式(18)所示:

$$e_r = e_t - e_h \quad (17)$$

$$r_s = \sum_{j=1}^{|K|} \text{ATT}(\text{FC}(e_{r_j})) \cdot e_{r_j}, j=1, 2, \dots, |K| \quad (18)$$

其中, e_h 和 e_t 表示参考集各目标三元组更新后的实体对嵌入表示; $|K|$ 表示参考集的大小是 K ;FC和ATT分别表示使用激活函数Relu和softmax的全连接层。然后,通过特征向量点积计算背景知识图谱关系集 $\mathcal{R}_{bg} = \{r_{b_1}, r_{b_2}, \dots, r_{b_{n-1}}, r_{b_n}\} \in \mathbb{R}^{n \times d}$ 中各关系 r_{b_i} 对参考集任务关系的语义相似性,如式(19)所示:

$$S_{\text{sem}}(r_{b_i}, r) = r_{b_i} \cdot r_s \quad (19)$$

其中, r_{b_i} 表示背景关系的初始预嵌入, $S_{\text{sem}}(r_{b_i}, r)$ 表示各背景关系对于目标关系的语义相似度。同时,为了进一步提高关系编码质量,综合考虑了各背景关系与任务关系的结构相似性。这里,将关系间的结构相似度定义为关系对应头尾实体集合的结构相似度之和。头实体集合和尾实体集合需要分开处理,这里以头实体集合为例。对于任一背景关系 $r_{b_i} \in \mathcal{R}_{bg}$,首先构造其头实体集合 $\mathcal{H}(r_{b_i})$ 并统计各实体的出现次数,取出现次数最多的 k 个实体作为头实体集合的代表实体集 $\mathcal{H}_k(r_{b_i})$,如式(20)所示:

$$\mathcal{H}_k(r_{b_i}), N(\mathcal{H}_k(r_{b_i})) = \text{Top}(\mathcal{H}(r_{b_i}), k) \quad (20)$$

其中, $\text{Top}(\cdot)$ 函数返回出现次数最多的 k 个实体集合 $\mathcal{H}_k(r_{b_i})$ 和对应的出现次数集合 $N(\mathcal{H}_k(r_{b_i}))$ 。对于所有 $e_i \in \mathcal{H}_k(r_{b_i})$,构造其一阶邻居实体集合 $\mathcal{N}_h(\mathcal{H}_k(r_{b_i}))$ 作为该关系的背景结构信息。同时,给定大小为 $|K|$ 的参考集 S_r ,分别构造其头尾实体集的一阶邻居实体集合 $\mathcal{N}_h(S_r)$ 和 $\mathcal{N}_t(S_r)$ 。对于背景关系 r_{b_i} 和任务关系 r ,计算其头尾实体集一阶邻域的Jaccard相似度之和,可得两关系的结构相似度 $S_{\text{str}}(r_{b_i}, r)$,

r), 如式(21)和式(22)所示:

$$\tau_h(r_{b_i}, r) = \frac{|\mathcal{N}_h(\mathcal{H}_k(r_{b_i})) \cap \mathcal{N}_h(S_r)|}{|\mathcal{N}_h(\mathcal{H}_k(r_{b_i})) \cup \mathcal{N}_h(S_r)|} \quad (21)$$

$$S_{\text{stu}}(r_{b_i}, r) = \tau_h(r_{b_i}, r) + \tau_r(r_{b_i}, r) \quad (22)$$

通过计算各背景关系和任务关系 r 的语义相似度 $S_{\text{sem}}(r_{b_i}, r)$ 和结构相似度 $S_{\text{stu}}(r_{b_i}, r)$ 之和, 可获得各背景关系对于任务关系 r 的综合相似度。为了减少噪声背景关系对于任务关系补全过程的影响, 保留综合相似度最高的 t 条背景关系, 如式(23)和式(24)所示:

$$S_{\text{sim}}(r_{b_i}, r) = S_{\text{sem}}(r_{b_i}, r) + S_{\text{stu}}(r_{b_i}, r) \quad (23)$$

$$\mathcal{R}_{\text{select}} = \text{Top}(\{S_{\text{sim}}(r_{b_i}, r)\}, t), r_{b_i} \in \mathcal{R}_{\text{bg}} \quad (24)$$

其中, $\mathcal{R}_{\text{select}}$ 表示与任务关系 r 最相似的 t 条背景关系集合, 此处取超参数 $t=1$ 。然后, 取各背景关系的综合相似度 $S_{\text{sim}}(r_{b_i}, r)$ 作为初始注意力分数, 并应用注意力机制对 $\mathcal{R}_{\text{select}}$ 中的关系做加权聚合, 获得任务关系 r 的嵌入表示, 如式(25)和式(26)所示:

$$\alpha(i) = \frac{\exp(S_{\text{sim}}(r_{b_i}, r))}{\sum_{j=1}^{|\mathcal{R}_{\text{select}}|} \exp(S_{\text{sim}}(r_{b_j}, r))} \quad (25)$$

$$r_o = \sum_{j=1}^{|\mathcal{R}_{\text{select}}|} \alpha(i) \cdot r_{b_j}, r_{b_j} \in \mathcal{R}_{\text{select}} \quad (26)$$

其中, r_{b_i} 表示 $\mathcal{R}_{\text{select}}$ 中各背景关系的初始化嵌入表示。

3.5 匹配网络

本文采用基于度量学习(Metric-based)的方法来计算查询三元组与参考集的相似性, 并取相似性最高的查询三元组作为模型的预测结果。受文献[17,36]启发, 本文设计了一个Transformer模块学习各三元组的关系表示。给定任务关系 r 的一个实体对 (h, t) , 首先使用3.3节的邻域编码器和3.4节的关系编码器, 对三元组的头尾实体及其对应关系进行补全; 然后, 按照三元组 (h, r, t) 的形式, 将实体对和对应关系补全后的嵌入表示结合位置编码拼接为序列 $X = \{x_1, x_2, x_3\}$, 并通过Transformer模块进行编码。

$$x_i^o = x_i^{\text{emb}} + x_i^{\text{pos}} \quad (27)$$

$$x_i^l = \text{Transformer}(x_i^{l-1}), l=1, 2, \dots, L \quad (28)$$

其中, x_i^{emb} 表示三元组 (h, r, t) 各位置补全后的嵌入表示, x_i^{pos} 表示该位置的位置嵌入, L 表示Transformer模块堆叠的层数。具体的模块细节可参考文献[17,36]。最后取位置为2的输出向量作为该三元组的关系代表向量, 如式(29)所示:

$$x(h, r, t) = x_2^l \quad (29)$$

考虑到参考集实例对不同查询三元组可能有不同的贡献度, 设计了一个原型网络用于为每一个查询三元组生成对应的参考集的原型关系表示 r_p , 如式(30)和式(31)所示:

$$r_p = \sum_{s_r \in S_r} c_i \cdot s_r \quad (30)$$

$$c_i = \frac{\exp(\cos(s_r, q_r))}{\sum_{s_r \in S_r} \exp(\cos(s_r, q_r))} \quad (31)$$

其中, s_r 和 q_r 分别表示经过Transformer模块获得的第 i 个参考集三元组的关系嵌入表示以及查询三元组的关系嵌入表示; $\cos(\cdot)$ 表示余弦相似度; r_p 表示针对查询关系嵌入表示 q_r 构建的参考集原型关系嵌入表示。之后, 可计算出查询 q_r 与 r_p 的语义相似性分数, 如式(32)所示:

$$\text{score}(r_p, q_r) = r_p \cdot q_r \quad (32)$$

对于查询三元组 (h, r, t) , 上述公式可以从语义层面评估该查询实体对 (h, t) 的正确性。然而, 式(32)并没有考虑查询实体对在结构层面的相似性。直观上理解, 结构相似的实体对间更有可能形成新链接, 综合考虑实体对的语义信息和结构信息有助于提高实体和关系编码的质量, 同时有助于提升链接预测的准确性。因此, 本文在评分函数部分引入实体对上下文信息交互指标, 综合考虑查询实体对 (h, t) 语义和结构上的相似性, 获得最终的相似性分数, 如式(33)所示:

$$\text{score}(r_p, q_r) = \text{score}(r_p, q_r)' \times (1 + \lambda \cdot \text{CIR}(h, t)) \quad (33)$$

其中, 超参数 λ 控制结构信息对于综合相似分数的贡献度。从公式层面可知, 查询实体对的结构相似性越大, 则综合相似性分数越高。当查询任务出现多组语义相似的查询三元组时, 引入结构信息有助于提升链接预测任务的预测精度。

3.6 模型训练

遵循其他FKGC模型^[17,36]的训练设置, 对于任务关系 $r \in \mathcal{R}_{\text{train}}$, BSA模型在三元组集合中随机抽取 K 个三元组构建参考集 S_r , 再在剩余三元组中构建正查询集 $\mathcal{Q}_r = \{(h_q, t_q) | (h_q, r, t_q) \in \mathcal{G}\}$, 并通过随机替换正查询集 \mathcal{Q}_r 的尾实体获得负查询集 $\mathcal{Q}_r^- = \{(h_q, t_q^-) | (h_q, r, t_q) \in \mathcal{G}, (h_q, r, t_q^-) \notin \mathcal{G}\}$ 。通过合页损失函数优化BSA模型, 如式(34)所示:

$$r_p = \sum_r \sum_{h_q, t_q \in \mathcal{Q}_r} \sum_{h_q, t_q \in \mathcal{Q}_r^-} [\gamma + \text{score}(h_q, t_q^-) - \text{score}(h_q, t_q)]^+ \quad (34)$$

其中, $[\cdot]^+ = \max(0, x)$ 表示标准的合页损失函数, γ 表示间隔距离。最后, 本文采用基于批量采样的元训练方法, 并使用Adam优化器优化模型的参数。

4 实验及结果分析

4.1 数据集与评价指标

为了验证BSA模型的预测效果, 本文在NELL-One和Wiki-One两个数据集上与其他基准模型进行对比。遵循已有工作的设置, 将关联三元组数量大于50且小于500的关系定义为小样本关系, 其中, 数据集NELL-One和Wiki-One的任务关系分别有67个和183个。采用与文献[16]相同的训练集/测试集/验证集分割方法, 关系的数量分别是51/5/11和133/16/34。数据集分割方法如表1所列。

表1 数据集统计信息

Table 1 Statistical information of datasets

Dataset	Ent	Rel	Triples	$\mathcal{R}_{\text{task}}$	$\mathcal{R}_{\text{train}}$	$\mathcal{R}_{\text{valid}}$	$\mathcal{R}_{\text{test}}$
NELL-One	68545	358	181109	67	51	5	11
Wiki-One	4838244	822	5859240	183	133	16	34

采用两个广泛使用的评价指标MRR和Hits@N与传统模型进行直接对比。MRR表示正确尾实体的平均倒数排名, Hits@N表示top-N范围内的正确尾实体命中率, N设置为1, 5, 10。

4.2 参数设置

本文的模块代码实现采用FAAN^[17]相同架构。遵循已有工作的设置, 实体和关系的初始嵌入使用TransE模型的预嵌入, NELL-One和Wiki-One的嵌入向量维度分别为100和50, Transformer模块的层数和多头注意力机制的头数分

别为 3 和 4 以及 4 和 8;在构建每轮训练的参考集和查询集时,模型将实体的最大邻居数量 M 设置为 30,间隔设置为 10,邻居数量大于 30 时模型将进行随机采样。同时,模型使用小批量梯度下降算法,优化器选择 Adam 优化器,批处理大小设为 128,学习率设置为 6×10^{-5} 。在模型训练过程中,保存模型验证步骤的最高 MRR 值,并将此次的模型参数作为 BSA 的最终训练结果。

4.3 基准模型设置

为了评估 BSA 模型的有效性,选用传统知识图谱嵌入方法、小样本知识图谱补全方法两类基准模型的主流模型进行比较。传统知识图谱嵌入方法中,选取 3 个具有代表性的主流模型作为基线模型,分别为 TransE, DistMult, ComplEx。小样本知识图谱补全方法包含基于度量学习和基于元学习两种技术路线,本文的技术路线属于度量学习的方法,选择的基线模型包括 GMatching, FSRL, FAAN, SAIA, FCC。其中, FAAN 是该技术路线的代表性模型,它提出使用动态注意力

机制来学习直接邻居关于目标关系的权重,大幅提高了实体嵌入表示的质量。SAIA 是 BSA 模型的直接基础模型,它在 FAAN 架构上进一步引入了选择性注意力机制以增强实体编码质量,并通过融合背景关系补全任务关系的语义信息,设计了实体对公共交互率(CIR),从而显著提升了打分函数对复杂关系的表达能力。BSA 在 SAIA 的架构上提出了 ES-NE, GSRE 和 E-CIR 模块,引入了背景知识图谱的结构信息以克服 SAIA 在稀疏关系和噪声邻域场景下的局限性。FCC 模型是度量学习路线的最新研究成果,它针对稀疏邻居问题设计了一个双分支特征提取器用于捕获实体的完整表示。基于元学习方法的技术路线,本文选择的基准模型有 MetaR 和 GANA。其中,MetaR 模型根据是否预先训练实体嵌入表示,分为 Pre-train 和 In-train 两种情况。

4.4 性能比较

NELL-One 和 Wiki-One 数据集上的模型链接预测结果分别如表 2 和表 3 所列。

表 2 模型在 5-shot 条件下的链接预测结果

Table 2 Link prediction results of the model under 5-shot conditions

Model (5-shot)	Nell-One				Wiki-One			
	MRR	Hit@10	Hits@5	Hit@1	MRR	Hit@10	Hits@5	Hit@1
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
GMatching	0.176	0.294	0.233	0.113	0.263	0.387	0.337	0.197
FSRL	0.153	0.319	0.212	0.073	0.158	0.287	0.206	0.097
MetaR(Pre-train)	0.209	0.355	0.280	0.141	0.342	0.463	0.395	0.281
MetaR(In-train)	0.261	0.437	0.350	0.168	0.221	0.302	0.264	0.178
FAAN	0.279	0.428	0.364	0.200	0.341	0.463	0.395	0.281
GANA	0.344	0.517	0.437	0.246	0.351	0.446	0.407	0.299
SAIA	0.346	0.510	0.442	0.262	<u>0.358</u>	<u>0.488</u>	<u>0.426</u>	0.289
FCC	<u>0.374</u>	0.566	<u>0.468</u>	0.282	0.346	0.472	0.410	0.272
BSA(ours)	0.378	0.544	0.476	0.287	0.365	0.494	0.432	<u>0.297</u>

注:加粗表示最优结果,下划线表示次优结果。

表 3 模型在 3-shot 条件下的链接预测结果

Table 3 Link prediction results of the model under 3-shot conditions

Model (3-shot)	Nell-One				Wiki-One			
	MRR	Hit@10	Hits@5	Hit@1	MRR	Hit@10	Hits@5	Hit@1
TransE	0.193	0.320	0.256	0.193	0.111	0.176	0.134	0.069
DistMult	0.231	0.375	0.306	0.164	0.112	0.195	0.156	0.069
ComplEx	0.185	0.273	0.223	0.129	0.106	0.145	0.117	0.085
GMatching	0.279	0.464	0.370	0.198	0.171	0.324	0.235	0.095
FSRL	0.318	0.507	0.433	0.211	0.241	0.406	0.327	0.155
GANA	0.322	0.510	0.432	0.225	0.331	0.435	0.389	0.283
FCC	<u>0.346</u>	0.553	0.454	<u>0.242</u>	<u>0.341</u>	<u>0.463</u>	<u>0.413</u>	0.263
BSA(ours)	0.364	<u>0.520</u>	<u>0.442</u>	0.280	0.347	0.487	0.423	<u>0.274</u>

注:加粗表示最优结果,下划线表示次优结果。

由表 2 可知,BSA 模型在 5-shot 条件下的表现显著优于基准模型 SAIA,并在多项指标上超越了最新模型 FCC。具体而言,在 NELL-One 数据集上,BSA 模型在 MRR, Hit@10, Hit@5 和 Hit@1 指标上较 SAIA 模型分别提高了 3.2 个百分点、3.4 个百分点、3.4 个百分点和 2.5 个百分点;在 Wiki-One 数据集上,各项指标分别提高了 0.7 个百分点、0.6 个百分点、0.6 个百分点和 0.8 个百分点。与最新模型 FCC 相比,BSA 模型在 MRR, Hit@5 和 Hit@1 指标上分别提高了 0.4 个百分点、0.8 个百分点和 0.5 个百分点;在 Wiki-One 数据集上, MRR, Hit@10, Hit@5 和 Hit@1 指标分别提高了

1.9 个百分点、2.2 个百分点、2.2 个百分点和 2.5 个百分点。表 3 列出了 3-shot 条件下的链接预测结果。与最新模型 FCC 相比,BSA 模型在 NELL-One 数据集上 MRR 和 Hit@1 指标分别提高了 1.8 个百分点和 3.8 个百分点;在 Wiki-One 数据集上, MRR, Hit@10 和 Hit@5 分别提高 0.6 个百分点、2.4 个百分点和 1.0 个百分点。

实验结果表明,BSA 模型在不同稀疏度的数据集上呈现差异化性能:在邻域信息丰富的 NELL-One 数据集上,5-shot 条件下 BSA 模型的预测精度显著优于基线模型(如 SAIA);而在邻域稀疏的 Wiki-One 数据集上,BSA 在 Hit@10 和

Hit@5等中等排名指标上表现更优,但是整体的提升幅度较小。这一差异可归因于以下两点:1)Wiki-One的长尾分布特性(约10%的关系的三元组数量小于10条)导致实体间交互稀疏,实体对上下文相似度数值整体偏低,削弱了编码过程的噪声抑制效果;2)稀疏关系使关系编码器难以充分学习语义特征,导致低频关系的嵌入表征稳定性下降,进一步制约了模型精度提升。

4.5 消融实验

BSA模型框架主要由3个关键部分组成:1)基于多维度

表4 模型在NELL-One数据集上的消融实验结果

Table 4 Ablation results of the model on NELL-One dataset

Model	Components				5-shot on Nell-One			
	ESNE	GSRE	E-CIR-v1	E-CIR-v2	MRR	Hit@10	Hit@5	Hit@1
BSA	✓	✓	✓	✓	0.378	0.544	0.476	0.287
w/o ESNE	×	✓	×	✓	0.299	0.438	0.369	0.222
w/o GSRE	✓	×	✓	✓	0.348	0.512	0.436	0.261
w/o E-CIR	✓	✓	×	×	0.319	0.504	0.433	0.213
w/o E-CIR-v1	✓	✓	×	✓	0.355	0.507	0.436	0.271
w/o E-CIR-v2	✓	✓	✓	×	0.324	0.515	0.437	0.217

由表4可知,移除模型的任一模块都会造成链接预测的性能下降。其中,移除ESNE模块的性能下降幅度最大,在包含ESNE模块的基础上分别添加GSRE模块和E-CIR模块都能大幅提高模型的预测精度。这说明:1)高质量的实体嵌入表示对于FKGC任务十分重要,ESNE模块能基于实体上下文交互率优先聚合语义和结构上高度相识的邻域实体信息,获得中心实体的高质量嵌入表示;2)实体对之间的邻域结构相似性不仅有助于提高邻域编码器的编码质量,还有助于优化模型的打分函数,提高预测精度;3)建模知识图谱中丰富的背景关系信息,有助于获得高质量的目标关系嵌入表示。GSRE模块通过简单的相似性计算方法,能有效地获取结构相似的背景关系,补全目标向量的嵌入表示,提高模型的预测精度。

为了验证不同位置中E-CIR值对模型预测精度的影响,对ESNE模块和打分函数模块分别进行了消融实验。由表4可知,移除ESNE模块中的E-CIR-v1导致MRR下降2.3个百分点,而移除打分函数中的E-CIR-v2使MRR下降5.4个百分点。这表明ESNE模块通过E-CIR筛选结构相似邻居的能力可部分被其他模块(如打分函数)补偿,但其独立贡献仍显著。同时,E-CIR在评分阶段直接优化结构一致性的作用更为关键,它直接影响最终排序结果。此外,完全移除E-CIR导致Hit@1指标下降7.4个百分点,进一步验证了结构感知机制对小样本补全任务的核心贡献。

4.6 实体嵌入可视化分析

为了更好地展示BSA模型对实体编码的补全效果,本文参考文献[22]的方法对关系ID#1(Produced By)的候选实体集进行了二维的可视化分析。针对直接使用实体对预嵌入和ESNE模块得到的候选实体对嵌入两种情况分别进行了t-sne可视化实验,结果如图2和图3所示。通过对比图2和图3中的实体分布情况,发现结构相似或语义相近的节点经过ESNE模块处理后,在二维t-SNE图上的分布显示出更高的聚集度和更紧密的空间邻近性。这表明ESNE模块能够有效

特征融合的实体邻域编码器(ESNE);2)基于全局结构相似性的关系编码器(GSRE);3)实体对上下文信息交互率(E-CIR)。为了验证各个模块对BSA模型整体性能的影响,表4列出了模型在Nell-One数据集上的5-shot链接预测消融实验结果。其中,w/o ESNE使用TransE模型获得的预训练嵌入向量作为实体的嵌入表示;w/o GSRE表示不使用关系编码器;w/o E-CIR表示不使用实体上下文交互率;w/o E-CIR-v1和w/o E-CIR-v2分别表示移除ESNE模块和打分函数模块并使用实体上下文交互率。

捕捉实体间的隐含结构和语义关系,证明了该模块在增强实体表示学习方面的优越性能。

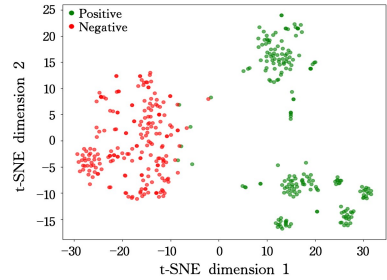


图2 预训练嵌入表示可视化

Fig. 2 Visualization of pre-trained embedding

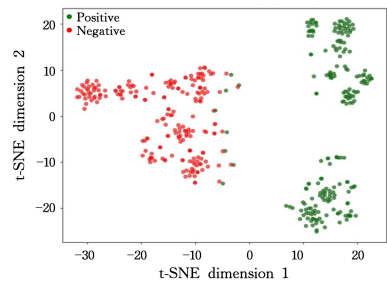


图3 经过ESNE模块处理的嵌入表示可视化

Fig. 3 Visualization of embedding processed by the ESNE module

4.7 超参数 λ 的影响分析

BSA模型使用了超参数 λ 控制实体对上下文信息交互率对打分函数的贡献度。为了确定 λ 的最佳取值,本文在NELL-One数据集的5-shot条件下,对 $[0, 0.5]$ 内 λ 取值进行了多次链接预测实验,结果如图4所示,其中横坐标表示 λ 的取值,纵坐标表示精确度。可以看到,各项指标的精确度在 λ 取值为0.15时普遍达到最大值,继续增大 λ 值后各项指标整体呈下降趋势。与 λ 值为0时的基准模型相比,BSA模型在MRR和Hits@1这两个指标上都至少提高了2个百分点和4个百分点。这说明在打分函数中引入目标实体对的上下文信

息交互率有助于提高模型的整体预测精度。

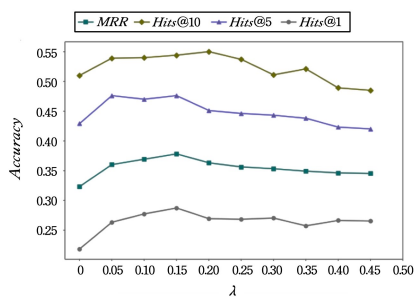


图4 不同超参数 λ 取值对于链接预测精度的影响

Fig. 4 Impact of different λ values on link prediction accuracy

4.8 超参数 T 的影响分析

BSA模型的关系编码器从背景知识图谱中抽取相似的背景关系补全任务关系 r ,并通过超参数 T 控制背景关系的数量。为了确定 T 的最佳取值,本文对 T 在 $[0,5]$ 内取值进行了多次链接预测实验, $T=0$ 时表示移除关系编码器模块。结果如图5所示,各项指标的精确度在 $T=1$ 时达到最大值,并随着 T 值的增大而降低。

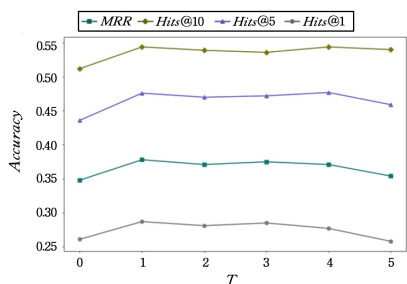


图5 不同超参数 T 取值对于链接预测精度的影响

Fig. 5 Impact of different T values on link prediction accuracy

这种现象出现的原因在于背景知识图谱中包含了大量的噪声关系,BSA模型设计的关系编码器还不能很好地降低噪声关系对编码质量的影响。目前设计的关系编码器取综合相似度最高的 T 条背景关系用于补全任务关系 r ,这意味着综合相似度的计算方法对编码质量起决定性影响。当综合相似度都普遍偏小时, T 值越大则引入的噪声信息越多,降低了编码质量。实验结果表明,将背景关系引入关系编码器可有效提升任务关系 r 的编码质量(取 $T=1$ 时,各项指标普遍提高2个百分点以上),而噪声关系带来的干扰仍需作为本研究的后续方向加以探索。

4.9 参考集大小 K 的影响分析

为了评估BSA模型在不同参考集大小下的性能,本文在NELL-One数据集上开展了一系列对比实验。实验结果如图6所示,其中横轴表示参考集大小 K ,纵轴表示评价指标MRR和Hits@1。

从图6中可以得出以下结论。1)在不同参考集实例数量下,BSA模型的预测精度显著优于大多数基准模型。与GANA模型相比,BSA在MRR和Hits@1指标上分别提高了7.6个百分点和8.1个百分点。2)BSA模型的预测精度随着参考集实例数量的增加而逐步改善,且在增加到一定程度后趋于稳定。这是由于初期更多的参考实例能够提供丰富的语

义信息和上下文关系,而当参考集过大时,新增实例可能引入冗余或噪声信息,削弱模型的有效性。3)BSA模型在参考实例数量较少的情况下仍能保持较高的预测精度,相比其他基准模型,其性能受参考实例数量的影响更小。这种现象可能与BSA模型在打分函数中引入的头尾实体对结构相似性评估有关,该机制在参考实例较少时为模型提供了额外的关联信息,从而在一定程度上保证了预测精度。同时,模型的其他设计,如E-CIR机制与注意力范围缩小,也可能共同起到了关键作用。

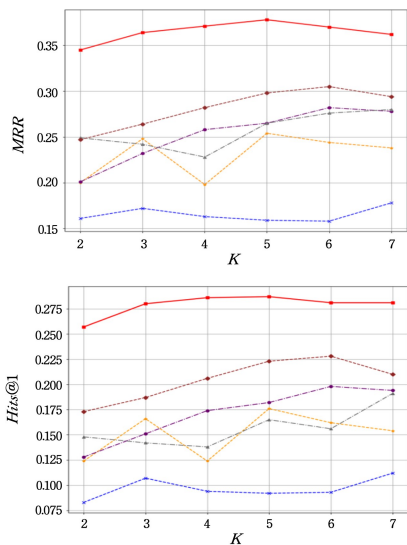


图6 NELL-One数据集上参考集大小对链接预测精度的影响
Fig. 6 Influence of reference set size on link prediction accuracy on the NELL-One dataset

4.10 讨论与未来工作

BSA模型在获得较好性能的同时,也存在以下局限:1)E-CIR模块依赖实体对上下文相似度的实时计算,若以预存储字典形式实现,则在数据集发生更新时需要重新更新字典,导致较高的计算开销与延迟;2)现有关系编码器对背景知识图谱中的噪声关系缺乏鲁棒性过滤机制,在训练过程中易受低质量背景关系的干扰;3)E-CIR的核心计算依赖于实体间的上下文关联,在Wiki-One等长尾分布显著的数据集中数值整体偏低,削弱了对编码过程的噪声抑制效果。

在未来工作中,将考虑通过以下方法提升模型性能:1)尝试在关系编码器中选用不同的实体对组合,观察不同组合下结构相似度对关系补全的效果;2)探索不同的任务关系 r 的编码方法。关系编码器和打分函数这两个模块需要对任务关系 r 进行建模,当前方法只使用了两个全连接层,在未来工作中,将探索利用Transformer架构来获取更高质量的任务关系 r 。

结束语 本文提出了一种基于背景结构感知的小样本知识图谱补全模型BSA。首先,BSA提出了一个实体对上下文信息交互指标E-CIR,在聚合邻域信息过程中计算各邻居节点的上下文信息,指导模型将注意力集中在与中心实体结构更相似的邻居节点上,提高了实体编码质量;其次,在关系表示学习过程中,BSA提出将关系间的结构相似度定义为关系对应头尾实体集合的结构相似度之和,并从背景知识图谱中

抽取语义和结构上相似的背景关系补全任务关系,增强了任务关系的表达能力;最后,BSA 在三元组的评分函数中引入头尾实体对的上下文信息交互指标,提升模型对复杂关系的推理能力。实验结果表明,BSA 模型在小样本知识图谱补全任务中表现出了显著的性能提升。特别是在长尾关系和稀疏邻域情况下,通过引入背景结构感知和实体对上下文信息交互机制,BSA 模型展示了其在复杂关系实体预测中的有效性。

参 考 文 献

- [1] CHEN X, JIA S, XIANG Y. A review: Knowledge reasoning over knowledge graph[J]. *Expert Systems with Applications*, 2020, 141(1): 112948.
- [2] FABIAN M S, GJERGJI K, GERHARD W. YAGO: a core of semantic knowledge[C]// *Proceedings of the 16th international conference on World Wide Web*. 2007: 697-706.
- [3] TAN H, ZHAO H, LI R, et al. A Pipeline Approach to Free-Description Question Answering in Chinese Gaokao Reading Comprehension[J]. *Chinese Journal of Electronics*, 2019, 28(1): 113-119.
- [4] 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.
- [5] SAXENA A, TRIPATHI A, TALUKDAR P. Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings[C]// *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020: 4498-4507.
- [6] WANG H, ZHANG F, ZHANG M, et al. Knowledge-aware Graph Neural Networks with Label Smoothness Regularization for Recommender Systems[C]// *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019: 968-977.
- [7] NGUYEN D Q, VU T, NGUYEN T D, et al. Preparing Network Intrusion Detection Deep Learning Models with Minimal Data Using Adversarial Domain Adaptation[C]// *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. 2019: 2180-2189.
- [8] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion [C]// *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2014: 601-610.
- [9] ZHANG N, DENG S, SUN Z, et al. Long-tail Relation Extraction via Knowledge Graph Embeddings and Graph Convolution Networks[C]// *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*. 2019: 3016-3025.
- [10] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating Embeddings for Modeling Multi-relational Data[C]// *Proceedings of the 26th International Conference on Neural Information Processing Systems*. 2013: 2787-2795.
- [11] WANG Z, ZHANG J, FENG J, et al. Knowledge Graph Embedding by Translating on Hyperplanes[C]// *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014: 1112-1119.
- [12] YANG B, YIH W, HE X, et al. Embedding Entities and Relations for Learning and Inference in Knowledge Bases[J]. *arXiv*: 1412.6575, 2014.
- [13] TROUILLON T, WELBL J, RIEDEL S, et al. Complex Embeddings for Simple Link Prediction[C]// *Proceedings of the 33rd International Conference on Machine Learning*. 2016: 2071-2080.
- [14] CHEN Z, WANG Y, ZHAO B, et al. Knowledge Graph Completion: A Review[J]. *IEEE Access*, 2020, 8: 192435-192456.
- [15] CHICAIZA J, VALDIVIEZO-DIAZ P. A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions[J]. *Information*, 2021, 12(6): 232.
- [16] XIONG W, YU M, CHANG S, et al. One-Shot Relational Learning for Knowledge Graphs[C]// *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. *ACL*, 2018: 1980-1990.
- [17] SHENG J, GUO S, CHEN Z, et al. Adaptive Attentional Network for Few-Shot Knowledge Graph Completion[C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. *ACL*, 2020: 1681-1691.
- [18] CHEN M, 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*. *ACL*, 2019: 4216-4225.
- [19] HE T, ZHOU H, ONG Y, et al. Not All Neighbors Are Worth Attending to: Graph Selective Attention Networks for Semi-supervised Learning[J]. *arXiv*: 2210.07715, 2022.
- [20] ZHANG M, WANG X, ZHU M, et al. Robust Heterogeneous Graph Neural Networks against Adversarial Attacks[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2022, 36(4): 4363-4370.
- [21] ZHANG C, YAO H, HUANG C, et al. Few-Shot Knowledge Graph Completion[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020: 3041-3048.
- [22] LI Y, YU K, HUANG X, et al. Learning Inter-Entity-Interaction for Few-Shot Knowledge Graph Completion[C]// *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. 2022: 7691-7700.
- [23] WANG J, WU T, ZHANG J. Incorporating Uncertainty of Entities and Relations into Few-Shot Uncertain Knowledge Graph Embedding[C]// *China Conference on Knowledge Graph and Semantic Computing*. Singapore: Springer, 2022: 16-28.
- [24] LIU J, FAN C F, ZHOU F, et al. Complete feature learning and consistent relation modeling for few-shot knowledge graph completion [J]. *Expert Systems with Applications*, 2024, 238: 121725.
- [25] NIU G, LI Y, TANG C, 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, ElectrNetwork. ACM, 2021: 213-222.
- [26] XIE P, ZHOU G, LIU J, et al. Incorporating global-local neighbors with Gaussian mixture embedding for few-shot knowledge graph completion[J]. Expert Systems with Applications, 2023, 234: 121086.
- [27] WU H, YIN J, RAJARATNAM B, et al. Hierarchical relational learning for few-shot knowledge graph completion[J]. arXiv: 2209.01205, 2022.
- [28] LUO L, LI Y F, HAFFARI G, et al. Normalizing flow-based neural process for few-shot knowledge graph completion[J]. arXiv: 2304.08183, 2023.
- [29] XU J, ZHANG J, KE X, et al. P-INT: A path-based interaction model for few-shot knowledge graph completion[C]// Findings of the Association for Computational Linguistics; EMNLP 2021. 2021: 385-394.
- [30] MA R, WU H, WANG X, et al. Multi-view semantic enhancement model for few-shot knowledge graph completion[J]. Expert Systems with Applications, 2024, 238: 122086.
- [31] LIANG Y, ZHAO S, CHENG B, et al. TransAM: Transformer appending matcher for few-shot knowledge graph completion [J]. Neurocomputing, 2023, 537: 61-72.
- [32] LI Q, YAO J, TANG X, et al. Capsule neural tensor networks with multi-aspect information for Few-shot Knowledge Graph Completion[J]. Neural Networks, 2023, 164: 323-334.
- [33] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling Relational Data with Graph Convolutional Networks[C]// 5th Conference on Semantic Web Challenges (SemWebEval Challenge) at 15th Extended Semantic Web Conference (ESWC). Berlin: Springer, 2018: 593-607.
- [34] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics. 2019: 4710-4723.
- [35] JAMBOR D, TERU K, PINEAU J, et al. Exploring the Limits of Few-Shot Link Prediction in Knowledge Graphs[J]. arXiv: 2102.03419, 2021.
- [36] LIN S, LU C H, JIANG W C, et al. Few-Shot Knowledge Graph Completion Based on Selective Attention[J]. Journal of Frontiers of Computer Science & Technology, 2024, 18(3): 646.
- [37] LI M, SUN Z, ZHANG W. Slan: similarity-aware aggregation network for embedding out-of-knowledge-graph entities [J]. Neurocomputing, 2022, 491: 186-196.



ZHANG Jing, born in 1972, postgraduate, associate professor, master's supervisor, is a member of CCF (No. K8905M). Her main research interests include big data, artificial intelligence, theories and applications of smart education, etc.



JIANG Wenchao, born in 1977, Ph.D., associate professor, master's supervisor, is a member of CCF (No. 78409M). His main research interests include cloud computing and complex network.

(责任编辑:何杨)