

基于大语言模型与适配器驱动的知识图谱补全算法

姜昀奇¹, 韩晓同², 田原²

(1. 吉林大学 计算机科学与技术学院, 长春 130012; 2. 吉林大学 人工智能学院, 长春 130012)

摘要: 针对基于 Transformer 为骨干网络的知识图谱补全方法在前馈网络参数冗余、常识场景下尾实体识别困难以及对比学习嵌入存在偏差等问题, 提出一种融合大语言模型与多正样本对比学习的适配器增强知识图谱补全算法。该算法通过在前馈网络中引入多头适配器减少冗余特征, 并利用大型语言模型提升常识推理能力, 同时通过多正样本对比学习校正嵌入偏差。实验结果表明, 相比于当前最佳模型, 该算法在数据集 WN18RR 和 FB15k-237 上的 MRR 分别提高了 5.4% 和 9.2%, 在更复杂的数据集 Wikidata5M 下的转导与归纳设定中分别提高 3.6% 和 6.7%, 并在低资源与复杂场景中展现出更佳的泛化能力。

关键词: 知识图谱补全; 知识图谱; 大语言模型; 对比学习; 适配器学习

中图分类号: TP391.1 **文献标志码:** A **文章编号:** 1671-5489(2026)02-0291-10

Knowledge Graph Completion Algorithm Based on Large Language Models and Adapter Driver

JIANG Yunqi¹, HAN Xiaotong², TIAN Yuan²

(1. College of Computer Science and Technology, Jilin University, Changchun 130012, China;

2. College of Artificial Intelligence, Jilin University, Changchun 130012, China)

Abstract: Aiming at the problems that the knowledge graph completion method based on Transformer as the backbone network, included parameter redundancy in feed-forward networks, difficulties in identifying tail entities under commonsense scenarios, and embedding biases in contrastive learning, we proposed an adapter-enhanced knowledge graph completion algorithm that integrated large language models with multi-positive sample contrastive learning. The algorithm reduced redundant features by introducing multi-head adapters in feed-forward network, and utilized large language models to enhance commonsense reasoning ability. At the same time, it corrected embedding biases through multi-positive sample contrastive learning. Experimental results show that, compared to the current state-of-the-art models, the algorithm improves MRR by 5.4% and 9.2% on WN18RR and FB15k-237 datasets, respectively, and by 3.6% and 6.7% in transductive and inductive settings on the more complex Wikidata5M dataset, respectively, and demonstrates superior generalization ability under low-resource and complex scenarios.

收稿日期: 2024-12-19.

第一作者简介: 姜昀奇(1999—), 男, 汉族, 硕士研究生, 从事知识图谱补全与大语言模型的研究, E-mail: jiangyq22@mails.jlu.edu.cn. **通信作者简介:** 田原(1985—), 男, 汉族, 博士, 副教授, 从事知识图谱与大语言模型的研究, E-mail: yuantian@jlu.edu.cn.

基金项目: 国家重点研发计划项目(批准号: 2023YFF0905400)、国家自然科学基金“叶企孙”科学基金(批准号: U2341229)和吉林省发展和改革委员会基金(批准号: 2024C003).

Keywords: knowledge graph completion; knowledge graph; large language model; contrastive learning; adapter learning

目前,知识图谱已成为表示各领域结构化信息的重要工具,广泛应用于问答系统和推荐系统等多个场景^[1].知识图谱通常由以 (h,r,t) 形式表示的三元组构成,其中 h 表示头实体, r 表示关系, t 表示尾实体.尽管知识图谱得到了广泛应用,但由于数据收集的局限性、数据质量问题以及现实世界的动态变化,知识图谱通常是不完整的.因此,知识图谱补全任务应运而生,旨在通过算法模型根据已知的两个元素预测三元组中的第三个元素.在知识图谱补全任务中,研究者们已提出了许多模型,这些模型不仅可补充知识图谱中缺失的三元组,还能推断图谱中潜在的新知识.

现有的知识图谱补全方法主要可分为基于结构的方法和基于文本的方法两大类.基于结构的方法主要利用现有的图谱结构,通过分析实体之间的模式和关系预测缺失的链接. TransE^[2]和 TransR^[3]等平移模型将关系表示为嵌入空间中的平移操作. RotatE^[4]和 ComplEx^[5]捕捉了包括对称性、反对称性和反转等更广泛的关系模式. R-GCN^[6]将图神经网络扩展到处理关系数据. ConvE^[7]则结合卷积神经网络用于学习图嵌入. 尽管这些模型取得了显著成功,但它们过度依赖结构信息,通常忽视了知识图谱上下文中蕴含的丰富语义信息. 基于文本的方法则利用与实体相关的文本描述和外部信息弥补这些不足,进而提升知识图谱补全任务的模型性能. 例如, KG-BERT^[8]通过将三元组视为文本序列,针对知识图谱补全任务微调预训练语言模型. LASS^[9]则进一步整合了文本和结构信息以增强表示学习. 此外, GenKGC^[10]将知识图谱补全视为一个序列生成任务,通过现有的文本信息生成新的三元组. SimKGC^[11]则利用对比学习最大化正三元组之间的相似性,并最小化负三元组之间的相似性,从而学习嵌入表征.

尽管基于文本的方法已有研究进展,但仍然存在一些问题. 首先,这些方法通常使用基于 Transformer^[12]的模型作为骨干网络,而前馈层(feed-forward network, FFN)是关键组件. 尽管 FFN 具有大量可训练参数,但它常在相邻的 Transformer 层之间学习到相似的特征,从而限制了模型捕捉数据多样性的能力^[13]. 其次,由于基于 Transformer 的模型预训练阶段主要依赖语义一致的文本数据,导致使用语义不一致的构造数据进行微调可能会破坏 FFN 中存储的丰富通用知识^[14],从而影响知识图谱补全性能. 最后,尽管对比学习的引入提高了知识图谱补全模型的性能,但其损失函数通常仅将包含头实体、关系和注释尾实体的三元组视为正样本,这种方式忽视了知识图谱中其他正确的尾实体,从而导致头-关系对嵌入的学习存在偏差. 因此,生成的嵌入缺乏表示性,且可能仅识别出一小部分正确的尾实体.

为解决上述问题,本文提出一种高效的适配器模型——AdpKGC,该模型通过大型语言模型(large language models, LLM)^[15]和多正样本对比学习进行增强. 在嵌入阶段,采用由查询编码器和键编码器构成的 Siamese 网络^[16],并在每个 Transformer 层中加入与 FFN 兼容的多头适配器(multi-head, MH). 该适配器能通过更少的参数有效学习知识图谱补全任务的新知识,避免了微调 FFN,防止其丰富的通用知识遭到破坏. 并且通过一个门控机制自适应地融合适配器和 FFN 的输出,缓解了 FFN 在相邻 Transformer 层之间捕捉到相似特征的问题. 在推理阶段,当模型无法准确识别给定头实体和关系对应的尾实体时,则利用 LLM 丰富的常识知识和推理能力,对备选尾实体进行重新排序. AdpKGC 将原始模型和 LLM 共同识别的尾实体作为最终答案,从而提高了常识知识图谱补全任务的性能. 在反向传播阶段,本文提出一种多正样本对比学习(MPS)损失函数,该损失函数由两部分组成:自正样本损失(SPS)和他正样本损失(OPS). SPS 损失仅将批次中标记为正样本的三元组视为正样本,而 OPS 损失则通过用知识图谱中的其他正确实体替换批次中正样本的尾实体补充这点. 该方法纠正了头-关系对嵌入学习的偏差问题,并通过学习更具区分性的头-关系对嵌入提升模型的泛化能力.

1 方法设计

本文提出的 AdpKGC 模型采用双编码器架构,如图 1 所示. 一个编码器负责将头-关系对编码为

查询向量, 另一个编码器负责将尾实体编码为键向量, 每个编码器均基于 bert-base-uncased 模型^[17] 构建, 该模型包含 12 层 Transformer 基本层. 为进一步提升模型的任务适应性, 在每一层 Transformer 的前馈网络 FFN 中设计了一个多头 MH 适配器, 以更少的参数高效学习特定任务的知识, 同时保留 FFN 中丰富的通用知识. 不仅缓解了信息模式冲突的问题, 还有效解决了相邻 Transformer 层间的特征冗余和相似性挑战. AdpKGC 模型的整体架构由三部分组成: 基于 MH 适配器的编码阶段、基于大语言模型增强的推理阶段和基于多正样本对比学习的损失计算阶段. 各部分协同作用以提升知识图谱推理的建模能力和推理精度.

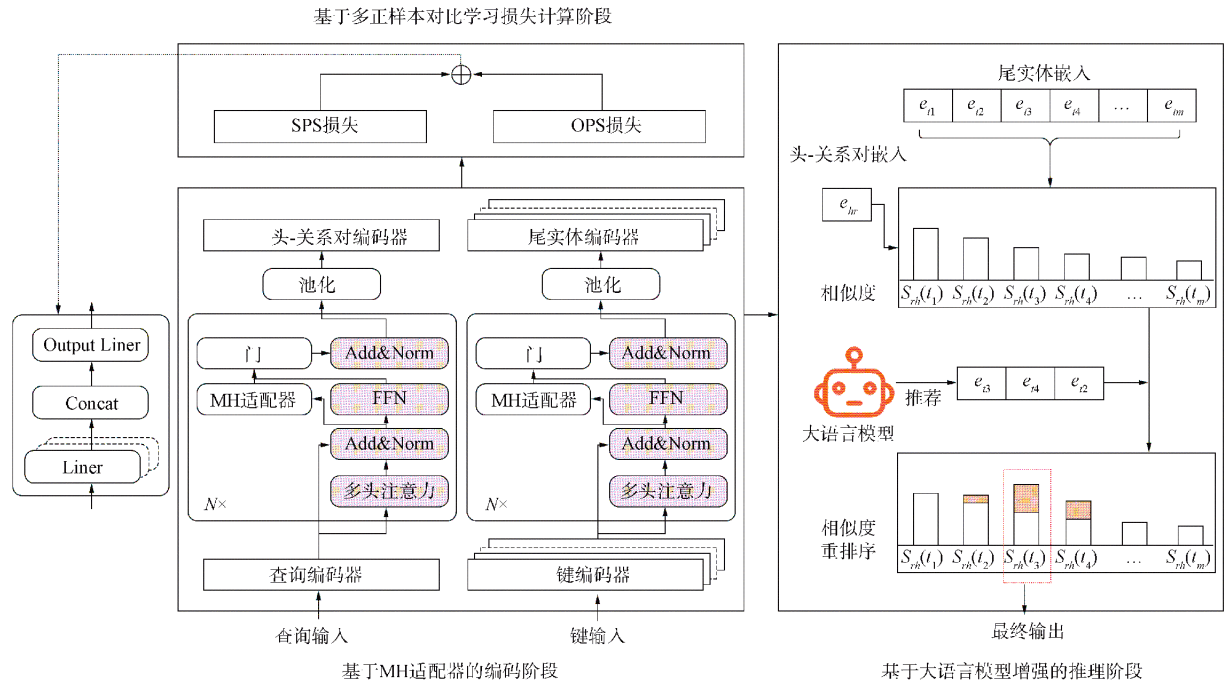


图 1 AdpKGC 模型架构

Fig. 1 Architecture of AdpKGC model

1.1 基于 MH 适配器的编码阶段

给定一个三元组 \$(h, r, t')\$, 其中包含头实体 \$h\$、关系 \$r\$ 和候选尾实体 \$t'\$, 分别具有描述 \$d_h\$ 和 \$d_{t'}\$, 查询编码器和键编码器的输入序列 \$\text{Input}_{hr}\$ 和 \$\text{Input}_{t'}\$ 分别定义为

$$\text{Input}_{hr} = ([\text{CLS}], h, d_h, h_{\text{neighbor}}, [\text{SEP}], r, [\text{SEP}]), \quad (1)$$

$$\text{Input}_{t'} = ([\text{CLS}], t', d_{t'}, t'_{\text{neighbor}}, [\text{SEP}]), \quad (2)$$

其中 \$h_{\text{neighbor}}\$ 表示与头实体 \$h\$ 相关的尾实体, \$t'_{\text{neighbor}}\$ 表示与尾实体 \$t'\$ 相连的头实体, \$[\text{CLS}]\$ 和 \$[\text{SEP}]\$ 表示输入序列中的特殊标记.

本文在每个 Transformer 编码器层的 FFN 中设计了一个 MH 适配器. 通过冻结 FFN 中的参数, 仅训练 MH 适配器, 模型能以较少的参数学习与知识图谱补全任务相关的新知识, 同时最大限度地保留 FFN 中的通用知识. 较少的参数也提高了学习效率, 使模型的每一层都能捕捉不同的特征, 更好地反映数据的多样性. 本文在适配器中采用多头机制, 使每个头在不同的空间中学习, 每个头特定关注数据中的不同信息子集或模式, 从而获得多样的表征. 以第 \$l\$ 层 Transformer 的 MH 适配器为例, 假设 \$\mathbf{X} \in \mathbb{R}^{n \times d_{\text{model}}}\$ 表示注意力机制输出经过归一化后的结果, 其中 \$n\$ 是输入序列中的标记数量. \$\mathbf{X}\$ 被切分成 \$m\$ 小段, 表示为 \$(\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m)\$, 其中 \$\mathbf{X}_i \in \mathbb{R}^{n \times d_{\text{adp}}}\$, 并且 \$d_{\text{adp}} = \frac{d_{\text{model}}}{m}\$. 每段 \$\mathbf{X}_i\$ 都经过 \$\mathbf{W}_i^{(l)}\$ 和 \$\mathbf{b}_i^{(l)}\$ 进行如下线性变换:

$$\mathbf{z}_i^{(l)} = \mathbf{X}_i \mathbf{W}_i^{(l)} + \mathbf{b}_i^{(l)}. \quad (3)$$

然后将经过每个头在不同子空间得到的 \$\mathbf{z}_i^{(l)}\$ 拼接成为一个完备的表征, 并通过变换矩阵 \$\mathbf{W}_o^{(l)}\$ 及偏置项

$\mathbf{b}_o^{(l)}$ 映射至同一空间中:

$$\text{Adapter}^{(l)}(\mathbf{X}) = \text{Concat}(\mathbf{z}_1^{(l)}, \mathbf{z}_2^{(l)}, \dots, \mathbf{z}_m^{(l)})\mathbf{W}_o^{(l)} + \mathbf{b}_o^{(l)}. \quad (4)$$

平衡 MH 适配器学习的新知识与 FFN 中保留的通用知识是一个巨大挑战. 本文通过设计一个门控机制解决该问题, 动态地平衡冻结的 FFN 和激活的 MH 适配器的贡献. 门控机制定义为

$$\mathbf{e}^{(l)} = \text{AddNorm}(\text{Gate}(\mathbf{x})) = \text{AddNorm}(\alpha^{(l)} \cdot \text{FFN}^{(l)}(\mathbf{X}) + (1 - \alpha^{(l)}) \cdot \text{Adapter}^{(l)}(\mathbf{X})), \quad (5)$$

其中 $\alpha^{(n)}$ 为一个可训练参数, 用于控制 FFN 输出 $\text{FFN}^{(n)}(\mathbf{X})$ 和 MH 适配器输出 $\text{Adapter}^{(n)}(\mathbf{X})$ 的融合程度. 该机制确保了模型在保留丰富通用知识的同时, 具有学习和泛化能力, 能适应知识图谱补全任务. 然后应用平均池化对最后一层 Transformer 编码器的输出进行池化操作, 以生成最终的嵌入表示:

$$\mathbf{e} = \text{Pooling}(\mathbf{e}^{(L)}), \quad (6)$$

其中 L 表示最后一层 Transformer 编码器. 通过上述过程, 得到了查询编码器中头-关系对的嵌入 \mathbf{e}_{hr} 以及键编码器中候选尾实体的嵌入 \mathbf{e}_t .

1.2 基于大语言模型增强的推理

在推理阶段, AdpKGC 模型通过计算头-关系对的嵌入与潜在尾实体嵌入之间的相似度分数识别给定头实体和关系的正确尾实体. 相似度分数的计算公式为

$$S_{hr}(t') = \mathbf{e}_{hr} \cdot \mathbf{e}_{t'}. \quad (7)$$

然后将潜在尾实体的相似度分数按降序排列为 $(S_1^*, S_2^*, S_3^*, \dots)$, 有最高相似度分数的尾实体被认为是正确的结果. 然而, 这种策略在某些情况下可能无效, 特别是当排名最高尾实体与排名较低尾实体的相似度分数非常接近时, 表明模型在区分正确尾实体和错误尾实体时存在困难, 从而导致预测结果的可靠性较低. 为解决该问题, 本文引入大语言模型 LLM, 利用其广泛的常识知识和强大的推理能力, 对前 10 个候选尾实体进行重新排序, 将更高的排名赋予那些同时得到原始模型和 LLM 推荐的尾实体. 本文设计一个指标 ΔS 衡量模型预测结果的困惑度, 通过比较候选尾实体中最高和第二高相似度分数之间的差异定义该指标:

$$\Delta S = \frac{S_1^* - S_2^*}{S_1^*}. \quad (8)$$

当 ΔS 超过设定阈值时, 表示预测可靠性较低, 则利用提示工程指导大语言模型推荐 3 个最有可能的尾实体 $\{a_1, a_2, a_3\}$. 每个推荐实体 a_j 根据因子 $\omega_j * S_1^*$ 赋予新的相似度分数, 其中 ω_j 表示根据大语言模型的排名给 a_j 分配的经验权重. 更新每个候选尾实体相似度分数的完整过程如下:

$$S'_{hr}(t') = \begin{cases} S_{hr}(t') + \omega_j * S_1^*, & \Delta S < \Delta_{th} \text{ 且 } t' = a_j, \\ S_{hr}(t'), & \Delta S \geq \Delta_{th} \text{ 或 } t' \neq a_j, \end{cases} \quad (9)$$

其中 Δ_{th} 表示预设的阈值. 最终, 预测的尾实体 \hat{t} 为具有最高更新后相似度分数的实体:

$$\hat{t} = \underset{t \in E}{\text{argmax}} S_{hr}(t'). \quad (10)$$

上述复合推理框架结合了小模型的高效性和大语言模型的深度推理能力. 通过使用小模型进行初步筛选, 使用大模型进行精细化处理, 保证了模型运行的速度和精度. 该框架通过重新排序候选尾实体并更新相似度分数, 进而提高了预测结果的可信度.

1.3 多正样本对比学习损失计算

尽管使用对比学习显著提高了知识图谱补全模型的性能, 但传统的对比学习损失函数通常只将由头实体、关系和批次中标注的尾实体组成的三元组视为正样本, 而忽略了知识图谱中其他许多正确的尾实体. 这一局限性导致学习到的头-关系对嵌入只能识别部分正确的尾实体, 从而导致有偏学习. 为解决该问题, 本文提出一种多正样本对比学习 MPS 损失函数, 该函数包括两部分损失.

第一个损失是自正样本 SPS 损失, 它将批次中的头实体、关系和标注尾实体的三元组视为正样本, 同时将批次中头实体、关系和所有错误尾实体的三元组视为负样本. L_{SPS} 确保模型能准确预测给定头实体和关系的最正确尾实体. 其公式遵循 InfoNCE 损失^[18]的形式, 定义为

$$L_{\text{SPS}} = -\log \frac{e^{(S_{hr}(t) - \gamma)/\tau}}{e^{(S_{hr}(t) - \gamma)/\tau} + \sum_{j=1}^{|N|} e^{S_{hr}(t'_j)/\tau}}, \quad (11)$$

其中 t 表示给定头实体 h 和关系 r 的真实尾实体, N 表示当前批次中所有错误尾实体 t_j^- 的集合, γ 和 τ 分别表示控制边界分离和概率分布锐度的超参数. 尽管该损失可有效引导模型预测正确的尾实体, 但在某些情况下, 即在头实体和尾实体之间存在一对多关系时, 实际上是正确但没有在当前批次中标注的尾实体, 它们也可以作为宝贵的信息补充. 通过这种方式, 模型能学习到与批次中标注尾实体不完全相关的头-关系对的嵌入, 进而获得一个更通用的嵌入, 这些嵌入能接近所有正确的尾实体, 从而反映出模型对知识图谱结构信息的更深理解. 为实现这一目标, 本文引入了第二个损失优化 L_{SPS} .

第二个损失是其他正样本 OPS 损失, 它在有效解决头-关系对嵌入有偏学习问题并增强模型的泛化能力方面具有关键作用. L_{OPS} 将由头实体、关系和整个知识图谱中所有其他正确尾实体组成的三元组视为正样本, 同时将批次中头实体、关系和所有错误尾实体的三元组视为负样本. 其定义为

$$L_{\text{OPS}} = -\log \frac{\frac{1}{|P|} \sum_{i=1}^{|P|} e^{(S_{hr}(t_i^+) - \gamma)/\tau}}{\sum_{i=1}^{|P|} e^{(S_{hr}(t_i^+) - \gamma)/\tau} + \sum_{j=1}^{|N|} e^{S_{hr}(t_j^-)/\tau}}, \quad (12)$$

其中 P 表示与头实体 h 在关系 r 下相关的所有其他正确尾实体 t_i^+ 的集合. 最终的损失函数 L_{MSP} 则通过加权的方式将 L_{SPS} 和 L_{OPS} 相结合:

$$L_{\text{MSP}} = \lambda L_{\text{SPS}} + (1 - \lambda) L_{\text{OPS}}, \quad (13)$$

其中 λ 为一个超参数, 通过动态调整使模型在准确性和泛化性之间取得平衡.

2 实验设计

2.1 数据集和评价指标

本文使用分别来自 FB15k-237^[19], WN18RR^[7] 和 Wikidata5M^[20] 的 4 个数据集评估模型的性能. 表 1 列出了各数据集的相关信息. 数据集 FB15k-237 来自 Freebase^[21], 包含约 15 000 个实体和 237 个关系, 总计 310 116 个三元组. 数据集 WN18RR 基于 WordNet^[22], 包含约 41 000 个实体和 11 个关系, 共 93 003 个三元组. 这两个数据集都去除了原始数据集 FB15K 和 WN18 中的逆向关系. 数据集 Wikidata5M 则以其大规模著称, 提供了转导和归纳两种设置, 支持多种实验场景. 该数据集包含约 480 万个实体和 822 个关系, 总计约 2 120 万个三元组. 转导设置是指针对已知测试数据的推理, 而不泛化到新数据的某种学习范式, 即测试集出现的实体均会出现在训练集中. 而归纳设置是指学习一个能泛化到未见过的新数据学习范式, 即已训练完成的模型可在测试时处理之前完全未见过的数据.

表 1 各数据集的信息

Table 1 Information of each dataset

数据集	实体数	关系数	训练集	验证集	测试集
WN18RR	40 943	11	86 835	3 034	3 134
FB15k-237	14 541	237	272 115	17 535	20 466
Wikidata5M-Trans	4 594 485	822	20 614 279	5 163	5 133
Wikidata5M-Ind	4 579 609	822	20 496 514	6 699	6 894

为评估模型在知识图谱补全任务中的性能, 本文采用 4 个广泛使用的评估指标, 并在过滤设置下进行测试. 这些指标包括 MRR, Hit@1, Hit@3 和 Hit@10. MRR 是正确实体倒数排名的平均值, 综合反映了预测实体的排名质量; 而 Hit@1, Hit@3 和 Hit@10 分别衡量正确实体出现在前 1 名、前 3 名和前 10 名预测结果中的比例. 本文采用双 BERT-base-uncased 作为编码器, 向量维度为 768, 最大句长为 512, 并使用 AdamW 优化器进行训练, 学习率设置为 $5 \times 10^{-4} \sim 10^{-5}$, 批次大小为 1 024, 温度参数为 0.05, Margin 为 0.02, 选用 GPT-4o 作为大语言模型.

2.2 整体实验

下面对 AdpKGC 模型在 4 个知识图谱补全任务常用的数据集 WN18RR, FB15k-237, Wikidata5M-Trans 和 Wikidata5M-Ind 上的性能与基线方法进行对比分析. 表 2 列出了不同模型在数据集 WN18RR 和 FB15k-237 上的实验结果. 由表 2 可见, 数据集 WN18RR 主要描述词汇之间的层次关

系,在该数据集中,基于文本的方法性能优于基于结构的方法,而 AdpKGC 模型在所有模型中性能最佳. AdpKGC 模型在 MRR 上超过了排名第二的 SimKGC 模型 5.4%,在 Hit@1, Hit@3 和 Hit@10 上分别提高了 8.2%, 3.2% 和 2.2%. 数据集 FB15k-237 的关系连接较稠密,在该数据集中,基于结构的方法性能较好,但 AdpKGC 模型超越了所有基线模型,在 MRR 上提高了 9.2%,在 Hit@1, Hit@3 和 Hit@10 上分别提高了 12.0%, 7.2% 和 1.5%. 实验结果表明, AdpKGC 模型在每个数据集的所有评价指标上均表现优异,从而充分证明了 AdpKGC 模型在知识图谱补全任务中的高准确性得益于三项关键创新.

表 2 不同模型在数据集 WN18RR 和 FB15k-237 上的实验结果

Table 2 Experimental results of different models on WN18RR and FB15k-237 datasets

%

方法	网络模型	WN18RR				FB15k-237			
		MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
基于结构	TransE	24.3	4.3	44.1	53.2	27.9	19.8	37.6	44.1
	ComPLEX	44.9	40.9	46.9	53.0	27.8	19.4	29.7	45.0
	RotatE	47.6	42.8	49.2	57.1	33.8	24.1	37.5	53.3
	DistMult	44.4	41.2	47.0	50.4	28.1	19.9	30.1	44.6
	SEPA	48.1	44.1	49.6	56.2	33.2	24.3	36.3	50.9
基于文本	KG-BERT	21.6	4.1	30.2	52.4	—	—	—	42.0
	MTL-KGC	33.1	20.3	38.3	59.7	26.7	17.2	29.8	45.8
	StAR	40.1	24.3	49.1	70.9	29.6	20.5	32.2	48.2
	GenKGC	—	28.7	40.3	53.5	—	19.2	35.5	43.9
	SimKGC	66.6	58.7	71.7	80.0	33.6	24.9	36.2	51.1
	AdpKGC	70.2	63.5	74.0	81.8	36.9	27.9	40.3	54.1

在数据集 Wikidata5M 的两种设置下, AdpKGC 模型同样取得了最佳性能. 表 3 列出了不同模型在数据集 Wikidata5M-Trans 和 FB15k-Wikidata5M-Ind 上的实验结果. 由表 3 可见,在转导设置中, AdpKGC 模型在 MRR 上超过最好的基线模型 3.6%,在 Hit@1, Hit@3 和 Hit@10 上分别提高了 4.8%, 3.7% 和 0.9%. 特别是在低资源条件下的归纳设置中, AdpKGC 模型在 MRR 上提高了 6.7%,在 Hit@1, Hit@3 和 Hit@10 上分别提高了 9.9%, 6.2% 和 5.6%. 实验结果展示了 AdpKGC 模型出色的泛化能力.

表 3 不同模型在数据集 Wikidata5M-Trans 和 FB15k-Wikidata5M-Ind 上的实验结果

Table 3 Experimental results of different models on Wikidata5M-Trans and Wikidata5M-Ind datasets

%

方法	网络模型	Wikidata5M-Trans				Wikidata5M-Ind			
		MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
基于结构	TransE	25.3	17.0	31.1	39.2	—	—	—	—
	RotatE	29.0	23.4	32.2	39.0	—	—	—	—
基于文本	DKPL	16.0	12.0	18.1	22.9	23.1	5.9	32.0	54.6
	KEPLER	21.0	17.3	22.4	27.7	40.2	22.2	51.4	73.0
	BLP-Complex	—	—	—	—	48.9	26.2	66.4	87.7
	BLP-SimplW	—	—	—	—	49.3	28.9	63.9	86.6
	SimKGC	35.8	31.3	37.6	44.1	71.4	60.9	78.5	78.5
	AdpKGC	37.1	32.8	39.0	44.5	76.2	66.9	83.4	92.6

2.3 消融实验

为验证 AdpKGC 模型中各模块的贡献,在两个数据集上进行消融实验,结果列于表 4. 由表 4 可见,移除 MH 适配器会导致在 MRR 上下降 1.2 个百分点,在两个数据集上的 Hit@1 分别下降 0.8 个百分点和 1.1 百分点. 表明 MH 适配器的引入对提升模型性能至关重要,不仅有效地为知识图谱补全任务学习新知识,同时保留 FFN 中丰富的语义知识. 此外,当将 MH 适配器的位置从前馈网络层调整至注意力计算层时,模型的性能显著下降,从而凸显了 MH 适配器与 FFN 协作的重要性. 本文在推理过程中移除了大模型增强推理模块,在数据集 WN18RR 中, MRR 和 Hit@1 均下降了 0.5 百分点. 在

数据集 FB15k-237 上, MRR 和 Hit@1 分别下降了 1.3 个百分点和 1.5 个百分点, 表明大模型所包含的丰富常识知识, 在以常识知识构建的知识图谱中起到了更重要的作用. 本文移除了 MPS 损失, 在数据集 WN18RR 中, 模型的 MRR 和 Hit@1 分别下降了 0.4 个百分点和 0.7 百分点. 而在数据集 FB15k-237 中, MRR 和 Hit@1 分别下降了 1.7 个百分点和 0.9 百分点. 表明 MPS 损失能解决与头-关系对嵌入的偏差学习问题, 并在复杂的知识图谱 FB15k-237 中产生更明显的效果. 此外, MPS 损失通过利用整个知识图谱中丰富的尾实体信息, 引导模型学习更多的结构化信息, 尤其在需要学习丰富结构化信息的数据集上效果显著.

表 4 消融实验结果

Table 4 Results of ablation experiment

网络模型	WN18RR				FB15k237			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
AdpKGC	70.2	63.5	74.0	81.8	36.9	27.9	40.3	54.1
移除 MH 适配器模块	69.0	62.7	72.7	80.0	35.7	26.8	39.2	53.6
移除大模型增强推理模块	69.7	63.0	73.2	81.8	35.6	26.4	38.9	54.1
移除 MPS 损失	69.8	62.8	74.0	81.6	35.2	27.0	38.3	50.0
更改 MH 适配器位置	68.3	61.1	72.2	81.3	34.8	6.2	38.3	50.3

3 实验分析

3.1 MH 适配器分析

FFN 前馈层在 Transformer 中扮演至关重要的角色, 先前的研究已证明大量的通用知识存储其中. 但近期研究表明, FFN 在相邻的 Transformer 层之间输出的相似性较高, 从而阻碍了模型有效捕捉数据多样性的能力. 在本文研究中观察到了相同现象. 为评估 FFN 在每个 Transformer 层输出之间的相似性, 本文采用线性中心化核对齐方法^[23]. 查询编码器和键编码器的结果如图 2 所示, 其中颜色越深表示层间特征相似度高, 即模型输出的特征越相似, 模型的参数冗余度越高. 在不同前馈网络层之间, 无论是查询编码器还是键编码器, 模型在较低层均存在严重的特征一致性, 即相邻前馈网络层输出的特征高度相似. 多头适配器的引入使各层之间的特征表达更多样化, 有效减少了参数冗余,

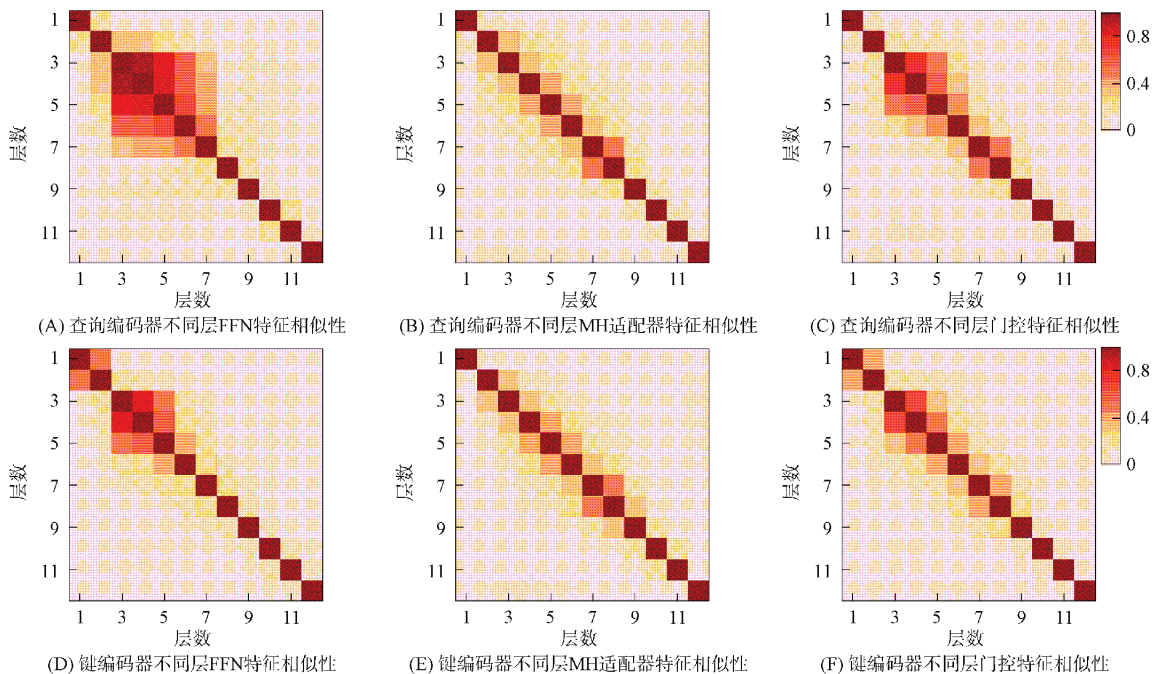


图 2 查询编码器和键编码器不同层之间的特征相似性 CKA 热力图

Fig. 2 CKA heatmaps of feature similarity between different layers in query and key encoders

提高了不同层次的表征能力. 此外, 本文还计算了 MH 适配器的输出与每个 Transformer 层中门控机制输出之间的相似性, 发现 MH 适配器通过减少各层之间的相似性, 增强了模型捕捉更多样化和丰富表示的能力, 从而使模型能更好地理解 and 区分输入数据中的各种特征.

MH 适配器在查询编码器的每个 Transformer 层中对 FFN 的贡献比例如图 3 所示, 其中 Alpha 为一个可训练的门控参数, 用于调节 FFN 与 MH 适配器之间的融合权重, 其值越大, 模型对 FFN 中知识的利用程度越高; 反之, 则更依赖于 MH 适配器所提供的知识. 由图 3 可见, MH 适配器对模型初始层的影响更显著. 门控机制能自适应地调整 MH 适配器的贡献, 当 Transformer 初始层中 FFN 输出的相似性较高时, MH 适配器能发挥更大作用.

MH 适配器另一个优势是能最大程度地保留

FFN 中已经存在的通用知识, 同时学习针对知识图谱补全任务的特定任务知识. 保留这些通用知识对成功完成知识图谱补全任务至关重要. 图 4 为不同训练数据比例 FFN 层的贡献. 由图 4 可见, 在低资源场景中, MH 适配器能获取的新知识越来越有限, 模型的能力更多依赖 FFN 中存储的通用知识. 此外, MH 适配器有效减少了需要训练的参数数量. AdpKGC 模型的可训练参数数量为 6 100 万个, 而 SimKGC 基线模型则需要 2.18 亿个参数, 与 SimKGC 模型相比, AdpKGC 模型所需的可训练参数少于其 1/3.

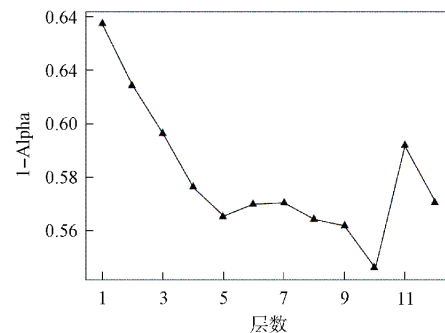


图 3 MH 适配器在不同层上的贡献

Fig. 3 Contribution of MH adapter on different layers

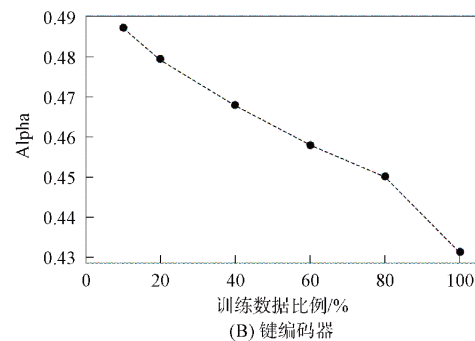
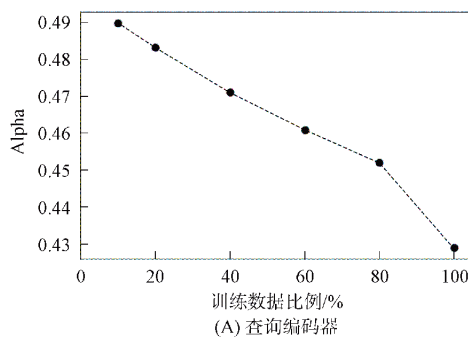


图 4 不同训练数据比例 FFN 层的贡献

Fig. 4 Contribution of FFN layer with different training data ratios

3.2 大模型模块分析

图 5 为各数据集使用大模型增强推理模块前后 Hit@1 性能对比. 由图 5 可见, 在集成大语言模型增强推理模块后, AdpKGC 模型在 Hit@1 和 Hit@3 指标上均表现出有效提升. 尤其在数据集 FB15k-237 和 Wikidata5M-Ind 上, 性能提升尤为显著. 这一改进主要归因于大语言模型所蕴含的丰富常识知识和强大的推理能力, 显著增强了模型的整体性能.

3.3 多正样本对比学习模块分析

AdpKGC 模型通过增强头-关系对嵌入的区分能力, 在复杂场景中提升了泛化能力. 本文通过计算每个头实体的平均尾数量 tphr 和每个尾实体的平均头数量 hptr 对关系进行分类. 如果 tphr 和 hptr 都低于 1.5, 则该关系被归类为简单场景; 否则, 归类为复杂场景. SimKGC, AdpKGC 和去除 MPS 损失后的 AdpKGC 模型在简单和复杂场景下的性能比较结果如图 6 所示. 由图 6 可见, 尽管随着复杂度的增加, 3 种模型的性能均有下降, 但 AdpKGC 模型在复杂场景下始终优于其他模型, 充分证明了 AdpKGC 模型的泛化能力.

综上所述, 针对基于 Transformer 为骨干网络的知识图谱补全方法在前馈网络参数冗余、常识场

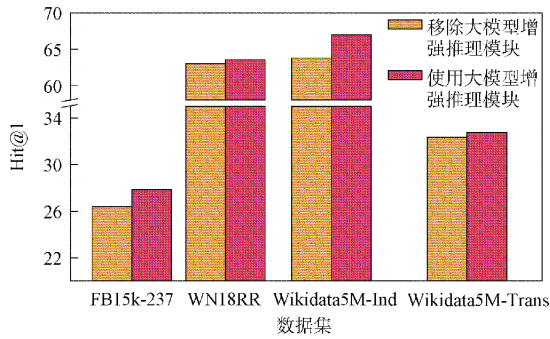


图 5 各数据集使用大模型增强推理模块前后 Hit@1 性能对比

Fig. 5 Comparison of Hit@1 performance with and without large model-enhanced inference module on each dataset

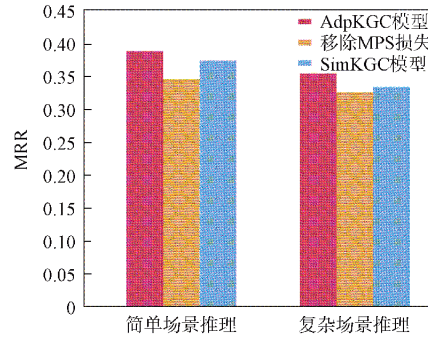


图 6 各模型在简单场景及复杂场景下的 MRR 性能对比

Fig. 6 Comparison of MRR performance of various models in simple and complex scenarios

景下尾实体识别困难以及对比学习嵌入存在偏差等问题, 本文提出了一种融合大语言模型与多正样本对比学习的适配器增强知识图谱补全算法. AdpKGC 模型通过引入多头自注意力适配器并结合门控机制, 在保留 FFN 层通用知识的同时, 更充分地学习任务特定知识, 从而有效缓解了相邻层学习相似特征的问题; 同时, 采用 MPS 损失函数进一步解决了头-关系对嵌入中的偏置学习问题. 在此基础上, 引入大规模语言模型以增强模型的常识推理能力. 实验结果表明, AdpKGC 模型在多个基准数据集上均取得了最优性能, 验证了其有效性与良好的泛化能力.

参 考 文 献

- [1] 陈成, 陈跃国, 刘宸, 等. 意图知识图谱的构建与应用 [J]. 大数据, 2020, 6(2): 57-68. (CHEN C, CHEN Y G, LIU C, et al. Constructing and Analyzing Intention Knowledge Graphs [J]. Big Data Research, 2020, 6(2): 57-68.)
- [2] BORDES A, USUNIER N, GARCÍA-DURÁN A, et al. Translating Embeddings for Modeling Multi-relational Data [C]//Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. [S.l.]: MIT, 2013: 2787-2795.
- [3] LIN Y K, LIU Z Y, SUN M S, et al. Learning Entity and Relation Embeddings for Knowledge Graph Completion [C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2015: 2181-2187.
- [4] 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)[2024-10-01]. <https://arxiv.org/abs/1902.10197>.
- [5] TROUILLON T, WELBL J, RIEDEL S, et al. Complex Embeddings for Simple Link Prediction [C]//Proceedings of the 33rd International Conference on Machine Learning. New York: ACM, 2016: 2071-2080.
- [6] SCHLICHTKRULL M S, KIPF T N, BLOEM P, et al. Modeling Relational Data with Graph Convolutional Networks [C]//The Semantic Web-15th International Conference. Berlin: Springer, 2018: 593-607.
- [7] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D Knowledge Graph Embeddings [C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2018: 1811-1818.
- [8] YAO L, MAO C S, LUO Y. KG-BERT: BERT for Knowledge Graph Completion [EB/OL]. (2019-09-07)[2024-12-16]. <https://arxiv.org/abs/1909.03193>.
- [9] SHEN J, WANG C, GONG L, et al. Joint Language Semantic and Structure Embedding for Knowledge Graph Completion [C]//Proceedings of the 29th International Conference on Computational Linguistics. [S.l.]: ACL, 2022: 1965-1978.
- [10] XIE X, ZHANG N Y, LI Z B, et al. From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer [C]//Companion of the Web Conference 2022. New York: ACM, 2022: 162-165.

- [11] WANG L, ZHAO W, WEI Z Y, et al. SimKGC: Simple Contrastive Knowledge Graph Completion with Pre-trained Language Models [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. [S.l.]: ACL, 2022: 4281-4294.
- [12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C]//Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017. [S.l.]: MIT, 2017: 5998-6008.
- [13] PIRES T, LOPES A V, ASSOGBA Y, et al. One Wide Feedforward Is All You Need [C]//Proceedings of the Eighth Conference on Machine Translation. [S.l.]: ACL, 2023: 1031-1044.
- [14] GEVA M, SCHUSTER R, BERANT J, et al. Transformer Feed-Forward Layers Are Key-Value Memories [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. [S.l.]: ACL, 2021: 5484-5495.
- [15] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and Efficient Foundation Language Models [EB/OL]. (2023-02-27)[2024-12-16]. <https://arxiv.org/abs/2302.13971>.
- [16] CHOPRA S, HADSELL R, LECUN Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005: 539-546.
- [17] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding [C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.]: ACL, 2019: 4171-4186.
- [18] CHEN T, KORNBLITH S, NOROUZI M, et al. A Simple Framework for Contrastive Learning of Visual Representations [C]//Proceedings of the 37th International Conference on Machine Learning. New York: ACM, 2020: 1597-1607.
- [19] TOUTANOVA K, CHEN D, PANTEL P, et al. Representing Text for Joint Embedding of Text and Knowledge Bases [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. [S.l.]: ACL, 2015: 1499-1509.
- [20] WANG X Z, GAO T Y, ZHU Z C, et al. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation [J]. Trans Assoc Comput Linguistics, 2021, 9: 176-194.
- [21] BOLLACKER K D, EVANS C, PARITOSH P K, et al. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge [C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM, 2008: 1247-1250.
- [22] MILLER G A. WORDNET: A Lexical Database for English [C]//Speech and Natural Language: Proceedings of a Workshop Held at Harriman. [S.l.]: ACL, 1992: 468.
- [23] GRETTON A, BOUSQUET O, SMOLA A J, et al. Measuring Statistical Dependence with Hilbert-Schmidt Norms [C]//Algorithmic Learning Theory. Berlin: Springer, 2005: 63-77.

(责任编辑:韩 啸)