

基于大语言模型的知识图谱补全轻量级混合重排序框架

陈建方¹⁾ 邵宇轩²⁾ 张 凯¹⁾ 刘 淇¹⁾ 甘傲然¹⁾

¹⁾(认知智能全国重点实验室(中国科学技术大学) 合肥 230027)

²⁾(安徽大学计算机科学与技术学院 合肥 230039)

摘要 知识图谱补全(knowledge graph completion, KGC)旨在推断知识图谱(knowledge graph, KG)中缺失的事实三元组。现有基于大语言模型(large language model, LLM)的知识图谱补全重排序方法主要依赖微调或多轮交互策略。微调方法因模型参数量巨大而导致训练和部署成本高昂;多轮交互虽避免了额外训练开销,却面临推理效率低下和Token消耗过大的问题。为此,本文提出一种轻量级的混合重排序框架LHR-KGC,旨在协同本地知识图谱补全模型与LLM的互补优势,在无需额外训练的前提下高效优化知识图谱补全结果。为提升结果的鲁棒性并规避单一模型失效的风险,提出了一种混合重排序策略,动态融合本地知识图谱补全模型与LLM的输出结果。针对推理效率和开销问题,设计了一种轻量级交互接口,通过结构化提示词模板单次调用LLM即可实现高效重排序。实验结果表明,LHR-KGC能够在多个公开数据集上稳定提升本地知识图谱补全模型的排序质量和预测性能。具体而言,在Wiki27K数据集上,MRR和Hits@1指标相较于本地知识图谱补全模型分别提升了10.5%和13.1%;在FB15K-237-N数据集上,则分别提升了7.0%和8.6%。此外,通过在多个LLM上进行的实验,进一步验证了LHR-KGC在优化本地知识图谱补全模型预测结果方面的有效性、鲁棒性和通用性。

关键词 知识图谱;知识图谱补全;混合重排序;大语言模型;提示学习

中图法分类号 TP181 **DOI号** 10.11897/SP.J.1016.2026.01442

A Lightweight Hybrid Re-Ranking Framework Based on Large Language Models for Knowledge Graph Completion

CHEN Jian-Fang¹⁾ SHAO Yu-Xuan²⁾ ZHANG Kai¹⁾ LIU Qi¹⁾ GAN Ao-Ran¹⁾

¹⁾(State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei 230027)

²⁾(School of Computer Science and Technology, Anhui University, Hefei 230039)

Abstract Knowledge graph completion (KGC) aims to infer missing factual triples in knowledge graphs in order to improve their completeness and reliability for downstream knowledge-driven applications. Existing KGC approaches can be broadly categorized according to their primary sources of information into structure-based methods and text-based methods. Structure-based models capture relational patterns through low-dimensional embeddings and demonstrate strong capability in modeling graph structure, while text-based methods incorporate contextual semantic information from textual descriptions to support reasoning. Despite their respective advantages, both categories suffer from inherent limitations: structure-based approaches often face scalability and generalization challenges due to computational complexity and uneven structural distributions,

收稿日期:2025-06-04;在线发布日期:2026-02-14。本课题得到国家自然科学基金(62525606(青年A类),U25B2072(联合重点),62406303(青年C类))资助。陈建方,硕士,主要研究领域为知识图谱、数据挖掘和大语言模型。E-mail: jfc@mail.ustc.edu.cn。邵宇轩,本科生,主要研究领域为计算机视觉、大语言模型和数据挖掘。张凯(通信作者),博士,副研究员,主要研究领域为人工智能、自然语言处理和文本表示学习。E-mail: kkzhang08@ustc.edu.cn。刘淇,博士,教授,主要研究领域为数据挖掘与知识发现、机器学习方法、教育大数据分析。甘傲然,博士研究生,主要研究领域为数据挖掘、知识图谱和自然语言处理。

whereas text-based methods are highly dependent on the availability and quality of textual descriptions and tend to exhibit unstable performance in sparse-relation or long-tail entity scenarios. These limitations indicate that a single type of model is insufficient to fully address the diverse reasoning requirements encountered in practical KGC tasks. Moreover, existing methods commonly exhibit deficiencies in candidate entity ranking quality: although correct entities frequently appear in candidate lists with sufficient coverage (e. g. , Top-10), they are often ranked too low to be effectively utilized, resulting in poor accuracy at high-priority positions such as Top-1. This mismatch between candidate coverage and ranking effectiveness significantly constrains the practical utility of current KGC systems and motivates the exploration of post-processing techniques for improving ranking quality. In recent years, large language models (LLMs) have demonstrated strong capabilities in semantic understanding, logical reasoning, and long-context modeling, making them promising tools for enhancing candidate re-ranking. However, existing LLM-based re-ranking approaches typically rely on fine-tuning or multi-turn interaction strategies. Fine-tuning methods incur high training and deployment costs due to the large number of model parameters, while multi-turn interaction methods avoid additional training overhead but suffers from low inference efficiency and excessive token consumption, limiting its practicality. To address these challenges, we propose a Lightweight Hybrid Re-ranking framework for Knowledge Graph Completion (LHR-KGC), which is designed as a general, plug-and-play post-processing module rather than a new KGC model. LHR-KGC leverages the complementary strengths of local KGC models and LLMs to optimize candidate entity ranking without additional training. To enhance robustness and mitigate the risk of single-model failure, we introduce a hybrid re-ranking strategy that dynamically fuses the ranking scores produced by local KGC models and LLMs. In addition, to address efficiency and cost concerns, a lightweight interaction interface is designed based on structured prompt templates, which enables effective re-ranking through a single invocation of the LLM and avoids the overhead associated with multi-round interactions. Extensive experiments are conducted on multiple public benchmark datasets, including Wiki27K and FB15K-237-N, to evaluate the effectiveness of the proposed framework. The experimental results demonstrate that LHR-KGC consistently improves the ranking quality and prediction performance of local KGC models across different evaluation metrics. Specifically, on the Wiki27K dataset, the proposed framework improves the Mean Reciprocal Rank (MRR) and Hits@1 by 10.5% and 13.1%, respectively, compared with the corresponding local KGC model, while on the FB15K-237-N dataset, improvements of 7.0% in MRR and 8.6% in Hits@1 are achieved. Furthermore, experiments on multiple LLMs further verify the effectiveness, robustness, and general applicability of LHR-KGC in optimizing the prediction results of local KGC models.

Keywords knowledge graph; knowledge graph completion; hybrid re-ranking; large language model; prompt learning

1 引言

知识图谱(knowledge graph, KG)是一种结构化的语义知识库,以三元组形式存储事实,每个三元组由头实体 h 、关系 r 和尾实体 t 组成,即 (h, r, t) 。该

表示方式能够清晰刻画实体之间的语义联系,为多类知识驱动型应用奠定基础。然而,由于数据采集、融合与更新等环节存在局限,实际构建的知识图谱普遍存在不完整性。此类缺失会对依赖知识图谱的应用造成显著的负面影响,例如问答系统^[1-2]和推荐系统^[3-4]。在问答系统中,不完整的知识图谱可能导

致无法准确响应用户查询;在推荐系统中,缺失事实会降低推荐结果的准确性与多样性。因此,知识图谱补全(knowledge graph completion, KGC)成为提升知识图谱完整性的重要研究任务。其核心目标是通过预测缺失的事实三元组,填补知识图谱中的空白,提升其完整性,为下游应用提供更加可靠的知识支持。

现有的KGC方法根据其利用的主要信息源,可以划分为基于结构的方法和基于文本的方法。基于结构的方法,如TransE^[5]、ComplEx^[6]和RotatE^[7]等,通过低维向量嵌入捕捉实体关系模式,能够有效挖掘知识图谱中的潜在结构信息。这类方法在关系模式捕捉方面表现卓越,擅长通过向量运算刻画实体间的复杂关系;基于文本的方法,如KG-BERT^[8]、SimKGC^[9]和KnowC^[10]等,依托上下文语义,从非结构化文本中提取语义关联,为补全任务提供额外信息支持。尽管两类方法均取得一定进展,但仍面临局限:基于结构的方法计算复杂度高、扩展性不足,难以处理大规模知识图谱;基于文本的方法则依赖文本描述的完整性和质量,在稀疏关系或长尾实体场景中表现不稳定。上述限制表明,单一类型模型难以充分覆盖复杂补全场景中的多样化推理需求。此外,现有方法在候选实体^①排序质量方面也存在不足。多数模型在高优先级候选实体(如Top-1)上的准确率较低,同时大量正确候选实体虽然出现在覆盖性候选实体列表(如Top-10)中,却排名靠后未被有效利用。高优先级候选实体命中率偏低意味着模型在关键决策阶段容易产生偏差,而被埋没的优质候选实体未被识别则进一步限制了KGC方法的实际效果。上述现象严重制约了KGC方法的实际应用价值。因此,如何有效识别并提升排名靠后的正确候选实体已成为当前亟待解决的关键问题。

近年来,大语言模型(large language model, LLM)在语义理解、逻辑推理和长上下文建模方面展现出卓越能力,例如DeepSeek-v3^[11]和Qwen2.5^[12]等。受此启发,一些研究尝试利用LLM直接执行KGC任务^[13-14]。然而,未经微调的LLM在生成缺失实体时存在输出不可控、语义漂移及多样性过高等问题,往往需要依赖人工将生成结果与知识图谱中的实体进行对齐,因此难以在大规模数据集上实现自动化评估。尽管已有工作尝试通过训练LLM用于KGC任务^[14],但此类方法本质上仍沿用传统生成式语言模型(如T5^[15])的架构思路,在模型架构与推理机制方面缺乏实质性突破。此外,面向常识知识领域的

生成式方法^[16-17]允许生成新的实体,但这与传统KGC任务以补全现有知识图谱为目标的设定并不一致。

近期兴起的重排序技术为突破上述瓶颈提供了新的研究方向,这是一种优化现有模型预测结果的后处理技术。基于LLM的重排序方法^[18-19]能够利用LLM的丰富隐式知识与强语义推理能力,在保持现有模型优势的同时提升候选实体排序质量。然而,现有方法多依赖模型微调或多轮提示交互:前者由于模型规模庞大导致训练与部署成本高昂,后者虽然避免了额外训练,但其推理效率低下且token消耗过大,不利于实际应用。为解决上述问题,本文提出了一种轻量级混合重排序框架LHR-KGC。该框架通过结构化提示词模板与混合重排序策略,在无需额外训练的条件下实现高效的语义增强重排序。LHR-KGC不仅适用于基于结构的基础模型,还能与基于文本的方法协同工作,通过LLM的深度语义理解能力弥补本地KGC模型^②在复杂关系推理中的不足,从而为KGC任务提供一种更加普适、可部署的解决方案。实验结果表明,LHR-KGC能够在多个公开数据集上稳定提升本地KGC模型的排序质量和预测性能,其中在Wiki27K^[20]数据集上MRR和Hits@1指标较本地KGC模型分别提升10.5%和13.1%,在FB15K-237-N^[20]数据集上则分别提升7.0%和8.6%。此外,通过在不同LLM上的综合实验,进一步验证了LHR-KGC框架在优化本地KGC模型预测结果方面的有效性、鲁棒性以及通用性。总而言之,本文的贡献如下:

(1)提出了一种通用、轻量且无需训练的混合重排序框架LHR-KGC。该框架并非用于构建新的KGC模型,而是作为一种可插拔的后处理模块,对任意本地KGC模型的候选实体进行排序优化。与依赖微调或多轮交互的已有重排序方法不同,LHR-KGC通过一次结构化提示词调用LLM即可完成高效重排序,实现了真正的零训练范式,极大降低了部署成本并具备良好的跨模型与跨数据集迁移能力;

(2)设计了一种混合重排序策略,利用本地KGC模型与LLM的互补性实现鲁棒性提升。该策略基于两类模型在结构关系建模与语义推理能力上

^① 候选实体是指KGC模型针对不完整的三元组预测出来的一组可能实体序列,其中排序越靠前表示可能性越高。

^② 本地KGC模型指基于结构或基于文本的传统KGC模型,比如TransE、KG-BERT等。

的互补性,通过权重融合的方式提升预测鲁棒性,显著降低单一模型失效带来的风险。消融实验表明,混合策略在不同LLM与不同数据集上均表现出稳定增益;

(3)在多个公开数据集和多种LLM上进行了系统实验,验证了框架的有效性、鲁棒性与通用性。实验结果表明,LHR-KGC框架不仅能提升本地KGC模型的预测排序质量,还在不同规模与不同来源的数据集上保持一致的改进效果,体现了其作为轻量级后处理框架在KGC任务中的实用价值。

2 相关工作

2.1 基于结构的知识图谱补全方法

现有KGC方法大部分是基于结构的,这类方法通常将实体和关系映射到低维向量空间,并通过不同的评分函数评估三元组的合理性。TransE^[5]是该领域的代表性基础模型,其将关系建模为线性平移,并通过向量距离衡量三元组得分。DistMult^[21]使用双线性评分函数来衡量实体嵌入和关系嵌入之间的相容性。随后研究引入了更具表达能力的表示方式。例如,ComplEx^[6]利用复数值嵌入来建模复杂的关系模式,而RotatE^[7]则通过在复平面上的旋转变换来捕捉对称、反对称等多类关系属性。然而,由于这些方法采用的是浅层网络结构,所以在表示能力上仍存在局限。近年来,更强大的神经网络结构被用于提升KGC的推理能力,包括图神经网络^[22]、卷积神经网络^[23]和Transformer^[24]等。这些方法通常通过聚合局部结构上下文增强节点嵌入表示,并在多个数据集上取得显著性能提升。但是该类结构化方法仍面临两大挑战:知识图谱结构分布不均导致模型在不同关系类型间表现不均衡,以及长尾实体因知识稀疏而难以学习到充分表征。

2.2 基于文本的知识图谱补全方法

随着自然语言处理(natural language processing, NLP)技术的发展,基于文本的KGC方法逐渐成为研究热点。这类方法通过融合知识图谱的结构化信息与文本的语义信息,有效缓解了纯结构方法的计算压力与表达不足问题。早期研究如DKRL^[25]首次将实体描述引入卷积神经网络以增强实体嵌入表示。随后,预训练语言模型(pre-trained language model, PLM)的兴起推动了KGC方法的范式转变。KG-BERT^[8]开创性地将三元组转换为自然语言序列,通过掩码语言模型预测缺失实体。MTL-

KGC^[26]引入多任务学习方法以捕捉知识图谱中更多的关系特征和词汇相似性。GenKGC^[27]提出将KGC任务转化为序列生成问题,通过关系引导和分层解码策略显著提升了推理效率。PKGC^[20]通过自然语言提示词设计有效弥合了结构化知识与自然语言的鸿沟,并引入软提示技术增强语义传达能力。SimKGC^[9]创新性地将损失函数从基于边界的排序损失改为信息噪声对比估计损失(InfoNCE loss),通过难负样本挖掘提升模型区分能力。KnowC^[10]提出知识上下文建模方法,通过附加提示引导预训练语言模型进行KGC任务。SKG-KGC^[28]则提出了一种多级共享知识引导的学习方法用于KGC任务。生成式语言模型的兴起进一步拓展了基于文本的KGC范式。KGT5^[29]将补全任务转化为序列生成问题,利用T5模型的强大生成能力直接输出缺失实体。ReSKGC^[30]结合检索与生成机制,通过动态选择相关文本增强推理准确性。KGT5-context^[31]引入实体邻域上下文信息,增强了模型对查询实体的理解。KGC-ERC^[32]动态融合实体邻域和关系上下文,进一步提升了模型的推理能力。这些方法利用Transformer架构^[33]的长距离依赖建模能力,在复杂关系推理中展现出独特优势。近年来,KGC研究开始探索与LLM的结合。例如,Li等人^[34]提出一种语境化蒸馏框架,通过将结构化三元组转化为富含上下文信息的文本片段,并设计重建与语境化辅助任务,使较小的KGC模型能够从增强后的三元组中获益。

2.3 知识图谱补全重排序方法

随着KGC技术持续发展,重排序方法已逐步成为提升补全性能的重要手段。该类方法的核心目标在于对现有KGC模型所生成的候选实体列表进行优化,通过引入额外语义线索与上下文信息进一步提升预测精度与稳定性。随着知识图谱规模不断增大和应用场景的日益繁复,传统的单一模型补全方法往往难以全面捕捉知识图谱中丰富的语义关联和结构特性。因此需要借助重排序策略进一步提升KGC性能。CEAR^[35]利用BERT^[36]对现有KGC模型的输出进行重排序,与以往对每个实体进行独立评分的方法不同,CEAR借助BERT中的跨实体注意力机制,对基于结构的模型所生成的候选实体进行联合评分,从而实现了对候选集的全局优化。近年来,随着LLM在自然语言处理、文本生成、问答系统、语义理解等众多任务中展现出广泛且卓越的性能^[37-38],部分研究开始尝试将其用于KGC重排序。

LLM具备强大的语义理解能力,能够精准剖析文本中的语义信息,捕捉词汇与句子之间的深层次关联;拥有出色的逻辑推理能力,可以依据给定的信息进行合理推导,得出准确的结论;还具备长上下文处理能力,能够有效处理较长的文本段落,从中提取关键信息并进行综合分析。例如,KICGPT^[18]采用传统的嵌入方法 RotatE^[7]作为本地 KGC 模型,并通过多轮交互式提示利用 ChatGPT (gpt-3.5-turbo) 对候选实体进行二次排序。尽管这种方式在一定程度上提升了补全的准确性,但多轮交互带来了较高的时间成本和资源消耗。KC-GenRe^[19]则依据不同数据集的特点,选用不同的模型作为第一阶段模型,然后运用 QLORA^[39]方法对 LLaMA-7b^[37]进行微调,进而对第一阶段的预测结果进行重排序。虽然这种灵活选择模型和微调的策略能够更好地适应不同数据集的特点,提高重排序的效果,但也面临着模型选择和微调参数设置的挑战。综上,当前重排序方法普遍面临训练成本高、推理效率低、参数敏感性强等问题。因此,在保证重排序效果的同时进一步提升效率、降低计算成本并增强可解释性,是未来 KGC 重排序研究的重要方向。

3 预备知识

3.1 知识图谱

知识图谱是一种关于实体及其相互关系的结构化表示。形式上,知识图谱可以表示为 $\mathcal{G}=(\mathcal{E}, \mathcal{R}, \mathcal{T})$,其中 \mathcal{E} 是实体的集合, \mathcal{R} 是关系的集合, $\mathcal{T}=\{(h, r, t) \mid h \in \mathcal{E}, r \in \mathcal{R}, t \in \mathcal{E}\}$ 是事实三元组的集合, h 表示头实体, r 表示关系, t 表示尾实体。

3.2 知识图谱补全

KGC 旨在通过利用现有信息预测缺失的三元组来填补知识图谱中的空白。对于一个不完整的三元组,例如 $(h, r, ?)$ 或 $(?, r, t)$,其中头实体或尾实体缺失,目标是确定最有可能填补该三元组的实体。具体而言,对于不完整的三元组 $(h, r, ?)$,知识图谱补全的任务目标就是找到一个尾实体 $t \in \mathcal{E}$,使得三元组 (h, r, t) 为真的可能性最大化。

3.3 基于大语言模型的知识图谱补全重排序

基于 LLM 的知识图谱补全重排序是指利用 LLM 对本地 KGC 模型预测的候选实体序列进行二次排序,使正确的实体排名提升,优化现有的预测结果。

假设给定一个查询 $q=(h, r, ?)$,需要预测其中缺失的尾实体 t 。重排序过程分三个阶段:第一阶段,通常使用高效的本地 KGC 模型生成一组候选尾实体及相应的分数,用 $E_{local}=[t_1^{local}, t_2^{local}, \dots, t_K^{local}]$ 表示预测的前 K 个候选实体,用 $S_{local}=[s_1^{local}, s_2^{local}, \dots, s_K^{local}]$ 表示其相应的得分;第二阶段,将上一阶段预测的候选实体 E_{local} 和得分 S_{local} 、不完整的三元组 q 、 q 相关的上下文 $C(q)$ 以及实体 h 的描述信息 $D(h)$ 一起通过结构化提示词模板转化成一段文本,然后输入 LLM 以输出一个重排序的候选实体列表 $E_{LLM}=[t_1^{LLM}, t_2^{LLM}, \dots, t_K^{LLM}]$,其中列表 E_{LLM} 只对 E_{local} 的候选实体位置进行置换,不改变所包含的候选实体;第三阶段,根据前面两个阶段的输出结果 E_{local} 、 S_{local} 和 E_{LLM} 计算出最终的候选实体序列 E_{final} 。

4 框架设计

本文提出的轻量级混合重排序框架 LHR-KGC 旨在通过协同本地 KGC 模型与 LLM 的互补优势,准确高效地完成 KGC 任务。该框架的核心在于利用 LLM 对本地 KGC 模型的预测结果进行优化,且无需额外的训练过程。框架的整体结构如图 1 所示,以下详述各模块设计与数学形式化表达。

4.1 本地候选实体生成

给定查询 $q=(h, r, ?)$ (以预测尾实体为例讲述原理,下同),本地 KGC 模型 M_{local} 生成 Top-K 候选实体列表 $E_{local}=[t_1^{local}, t_2^{local}, \dots, t_K^{local}]$ 及对应得分 $S_{local}=[s_1^{local}, s_1^{local}, \dots, s_K^{local}]$,形式化为

$$E_{local}, S_{local} = M_{local}(q, K, C(q), D(h)) \quad (1)$$

其中, $C(q)$ 表示知识图谱中与查询 q 相关的上下文信息,如不需要可设置为 \emptyset ; $D(h)$ 表示实体 h 的描述信息,如不需要可设置为 \emptyset ; S_{local} 是原始得分归一化后的得分。该设计使得 LHR-KGC 框架可兼容多种不同的本地 KGC 模型。如图 1 所示,不同输入信息通过统一接口适配,最终输出标准化的候选实体列表 E_{local} 及其得分 S_{local} 。

4.2 结构化提示词模板构建

基于提示学习(prompt learning)范式,本节设计了一个结构化提示词模板来实现与 LLM 的高效交互。给定查询 $q=(h, r, ?)$,首先获取本地候选实体列表 E_{local} 及其得分 S_{local} ,同时结合头实体的描述信息 $D(h)$ 。这些信息通过结构化提示词模板,转化为适合 LLM 处理的自然语言形式的提示词 \mathcal{P} ,然后输

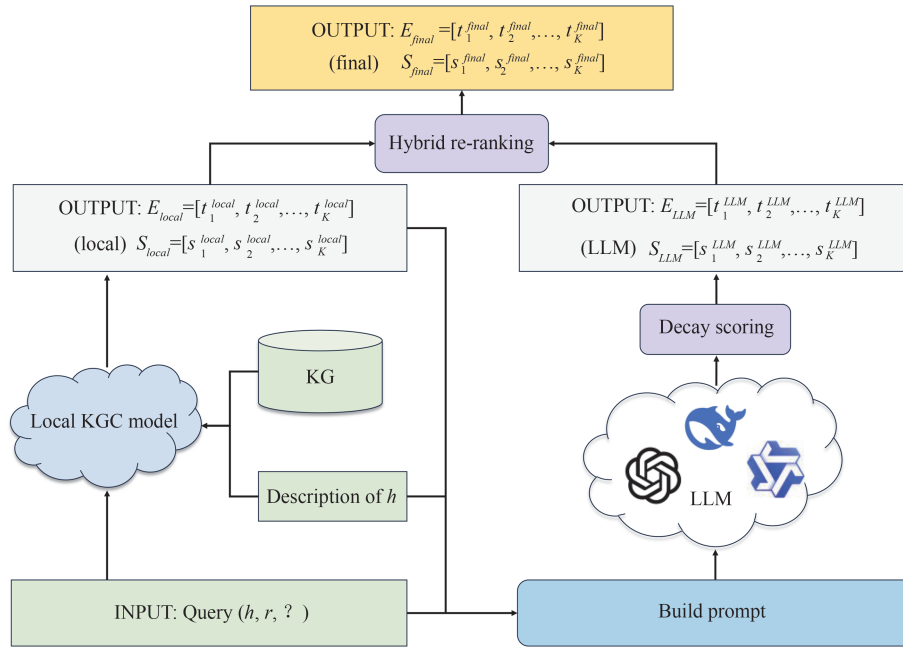


图1 LHR-KGC框架图

入给LLM。该设计旨在引导LLM根据其语言理解和推理能力对候选实体进行重排序。如表1所示,提示词模板采用类型敏感的差异化设计策略,针对不同实体类型优化模板结构。模板中的 $\{\cdot\}$ 部分使用时需替换成实际的文本。此外,为避免实体名称歧义并降低token消耗,在处理本地候选实体列表 E_{local} 及其得分 S_{local} 时会同时给每个候选实体分配一个ID,对应模板中的变量“ $\{candidates_with_scores_and_ids\}$ ”,然后要求LLM只需返回一个重排序后的ID列表。这种设计既保留了LLM的语义理解优势,又通过提示学习技术有效约束了模型的推理方向。

同时,本文在所有实验中均使用同一份结构化提示词模板,不针对数据集或模型进行专门调参。模板采用人工设计与模型反馈结合的方式构建:首先根据任务结构人工编写初始模板,再通过少量实例与LLM进行交互,根据中间推理结果和模型响应逐步调整模板表达,使模型能够准确理解任务目标。最终,通过对输出格式进行严格约束,使模型生成符合要求的最简结构化结果。

此外,由于该提示词模板主要依赖任务中“候选实体排序”这一通用目标,与具体数据集和本地KGC模型无耦合,因此具有良好的迁移性。在不同规模的数据集以及不同LLM上均无需修改即可使用。该现象表明该提示词模板在任务级别上具有较强的通用性,并能适配多种LLM的推理范式。

4.3 大语言模型重排序

以提示词模板构建出的提示词 \mathcal{P} 作为输入,通过API调用大语言模型 M_{LLM} 获取重排序结果 $\pi = [\pi_1, \pi_2, \dots, \pi_K]$,其中 $\pi_j \in \{1, 2, \dots, K\}$ 为候选实体ID。

为了统一评估方式,设计指数衰减赋分策略将LLM排序结果量化为用得得分表示的权重:

$$\omega(\pi_j) = e^{-\lambda \cdot (j-1)}, \quad 0 < \lambda < 1 \quad (2)$$

其中 λ 用于控制得分的衰减率, λ 越大表示越重视头部的候选实体, j 表示排名,越小得分越高。给定位置 i 的候选实体,其LLM的分数计算为

$$S_{LLM}(i) = \begin{cases} \sum_{j=1}^K \omega(\pi_j) \cdot \delta(\pi_j = i), & \text{if } i \in \pi \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

其中 $\delta(\cdot)$ 为指示函数。由于只对本地预测结果的前 K 个候选实体进行重排序,所以未出现在 π 中的候选将保留原始本地排序。通过 π 映射还原LLM输出的候选实体序列为

$$E_{LLM} = \text{Map}(E_{local}, \pi) \quad (4)$$

最终本模块的输出为

$$\begin{aligned} E_{LLM} &= [t_1^{LLM}, t_2^{LLM}, \dots, t_K^{LLM}] \\ S_{LLM} &= [s_1^{LLM}, s_2^{LLM}, \dots, s_K^{LLM}] \end{aligned} \quad (5)$$

4.4 混合重排序

混合重排序的目的是降低单一模型失效的风险,协同本地KGC模型与LLM的互补优势,提升预测结果的鲁棒性。

表 1 两种结构化的提示词模板

预测实体	提示词模板
尾实体	<p>You are provided with the following:</p> <ul style="list-style-type: none"> - Head entity: "{head_entity}" - Relation: "{relation}" - Description of the head entity: "{description_of_head_entity}" - A list of candidate tail entities with ids and scores predicted by a local knowledge graph completion model: "{candidates_with_scores_and_ids}" <p>The score represents the probability that the candidate is correct. The higher the score, the higher the likelihood. The local model's performance is limited, and the predicted results may not be entirely accurate. Use the provided information and your own knowledge to re-rank the candidate tail entities in descending order of likelihood.</p> <p>Return a JSON-formatted list of the candidate entity ids in the new order. The list must be the same length as the candidate list and should only contain the ids.</p> <p>For example: [id_1, id_2, ...]</p> <p>Do not include any additional text.</p>
头实体	<p>You are provided with the following:</p> <ul style="list-style-type: none"> - Tail entity: "{tail_entity}" - Relation: "{relation}" - Description of the tail entity: "{description_of_tail_entity}" - A list of candidate head entities with ids and scores predicted by a local knowledge graph completion model: "{candidates_with_scores_and_ids}" <p>The score represents the probability that the candidate is correct. The higher the score, the higher the likelihood. The local model's performance is limited, and the predicted results may not be entirely accurate. Use the provided information and your own knowledge to re-rank the candidate head entities in descending order of likelihood.</p> <p>Return a JSON-formatted list of the candidate entity ids in the new order. The list must be the same length as the candidate list and should only contain the ids. For example: [id_1, id_2, ...]</p> <p>Do not include any additional text.</p>

首先,为消除评分的尺度差异,分别对本地KGC模型得分 S_{local} 和LLM得分 S_{LLM} 实施Min-Max归一化:

$$S_{local}^{norm} = \frac{S_{local} - \min(S_{local})}{\max(S_{local}) - \min(S_{local})} \quad (6)$$

$$S_{LLM}^{norm} = \frac{S_{LLM} - \min(S_{LLM})}{\max(S_{LLM}) - \min(S_{LLM})} \quad (7)$$

其中, S^{norm} 表示基于当前列表的归一化值,这里用 S_{local}^{norm} 表示本地KGC模型的归一化分数,用 S_{LLM}^{norm} 表示LLM的归一化分数。最终分数是归一化本地KGC模型分数和LLM分数的凸组合:

$$S_{final} = \alpha \cdot S_{local}^{norm} + (1 - \alpha) \cdot S_{LLM}^{norm} \quad (8)$$

$$= [s_1^{final}, s_2^{final}, \dots, s_K^{final}]$$

其中, $\alpha \in [0, 1]$ 为本地KGC模型权重调节器,体现对本地KGC模型预测确定性的信任程度,越大表示越信任本地KGC模型。然后,根据 S_{final} 输出最终排名:

$$E_{final} = [t_1^{final}, t_2^{final}, \dots, t_K^{final}] \quad (9)$$

此外,LHR-KGC框架在实现混合重排序时无需对LLM进行任何形式的微调,也不需要额外构建训练数据,仅通过一次结构化提示词调用LLM即可完成候选实体的排序优化。这种零训练范式带来了以下重要特性:(1)极低的资源成本。微调LLM通常需要大规模GPU资源,而LHR-KGC框架完全避免训练过程,使其可在资源受限环境中部署;(2)模型级通用性。由于不涉及参数更新,LHR-KGC框架可直接适配不同架构的LLM。在GPT-4o-mini、Qwen-plus、Qwen-max、DeepSeek-V3等模型上的实验均验证了其适配能力;(3)跨图谱迁移能力。结构化提示词模板能够激活LLM的通用知识,使LHR-KGC框架在迁移到不同规模的知识图谱时无需额外训练即可保持稳定表现;(4)鲁棒性更高。与需要微调的重排序方法相比,零训练范式不会引入新的数据偏差,能够在复杂场景中保持较高稳健性。综上所述,零训练特性构成了LHR-KGC框架轻量、高效、易部署的重要基础,为后续性能优化奠定了条件。

4.5 框架优势

LHR-KGC框架在KGC任务中展现出良好的实践价值,其优势可概括如下:

(1)无需额外训练的快速集成能力:依托提示学习范式,LHR-KGC框架可以直接与任意本地KGC模型组合使用,实现即时部署,不依赖模型变化调整流程;

(2)单轮重排序带来的高效率:结构化提示词模板使重排序过程能够在一次API调用中完成,减少了多轮复杂交互带来的延迟,有效提升了系统的推理速度;

(3)基于混合策略的稳健预测性能:通过结合本地KGC模型与LLM的预测得分,LHR-KGC框架能够在不同数据集和不同推理条件下保持稳定且一致的性能增益;

(4)广泛的适配能力与可扩展性:LHR-KGC框架设计为完全可插拔,不依赖特定模型架构,可自由搭配不同的本地KGC模型与LLM,以适配多样的任务与应用场景。

5 实验设置

5.1 数据集

本文选用了三个公开数据集来进行性能评估,分别是:Wiki27K^[20]、FB15K-237-N^[20]和Wikidata5M^[40]。数据集的详细统计信息如表2所示。与Wiki27K和FB15K-237-N相比,Wikidata5M的规模要大得多。具体来说,Wiki27K数据集来源于Wikidata,包含大约2.7万个实体和62种关系。FB15K-237-N数据集是从FB15K-237^[41]中移除涉及中介节点的关系后得到的,包含大约1.3万个实体和93种关系。相比之下,Wikidata5M数据集将Wikidata知识图谱与维基百科相结合,涵盖了近500万个实体和大约2000万个三元组。关于数据集中使用的实体、关系和实体描述的文本信息的来源,其中FB15K-237-N和Wiki27K数据集来自PKGK^[20],Wikidata5M数据集自于KGT5^[29]。

表2 数据集统计信息

数据集	实体	关系	训练集	验证集	测试集
FB15K-237-N	13,104	93	87,282	7,041	8,226
Wiki27K	27,122	62	74,793	10,121	10,122
Wikidata5M	4,594,485	822	20,614,279	5,163	5,133

5.2 评估指标

为了全面评估框架性能,本文采用了四种常用的评估指标:平均倒数排名(MRR)和Hits@ k ,其中 $k \in \{1, 3, 10\}$ 。MRR用于衡量测试三元组的正确实

体在候选预测实体中排名倒数的平均值,而Hits@ k 表示在前 k 个结果中包含正确预测实体的比例。在生成提示词之前,本地KGC模型的预测结果会提前过滤掉知识图谱中不存在的非法实体,以确保预测结果的有效性。在出现排名并列的情况下,采用平均排名的方法以避免误导性结果^[42]。所有指标均在“过滤”设置^[5]下计算,该设置从排序列表中过滤掉了真实目标实体之外的有效实体,从而确保评估结果的准确性。另外,为确保全面分析,本文报告了头实体和尾实体预测指标的平均值。

5.3 实现细节

尽管本文提出的LHR-KGC框架可以使用各种本地KGC模型和LLM,但是出于实际意义考虑,应优先选用具备较强推理能力的本地KGC模型,因为只有对高性能的本地KGC模型有提升作用才能体现所提方法的价值。具体而言,LHR-KGC框架使用KGC-ERC^[32]作为本地KGC模型。该模型的优势在于能够有效地融合多种上下文信息,并具备较强的推理能力,在当前的基线模型中也取得了最佳的性能表现。在LLM的选择上,综合考量模型架构、参数量以及应用场景等多方面因素,最终选定GPT-4o-mini^[38]、DeepSeek-v3^[11]、Qwen-plus^[12]和Qwen-max^[12]四款常用的模型。KGC-ERC的参数设置与原文保持一致,候选实体数量 $K \in \{10, 15, 20\}$ 。衰减系数 λ 从 $\{0.1, 0.2, \dots, 0.7\}$ 中选择,权重系数 α 从 $\{0.0, 0.1, \dots, 1.0\}$ 中选择。通过在验证集上进行全面且细致的网格搜索,确定了一系列超参数的最佳配置,具体信息如表3所示。LLM的参数均使用官方默认设置,未进行单独调整,以确保实验的公平性和可复现性。提示词模板在所有实验中保持一致,以验证LHR-KGC在不同模型组合下的有效性和稳定性。

表3 最佳超参数

模型	Wiki27K			FB15K-237-N			Wikidata5M		
	α	λ	K	α	λ	K	α	λ	K
LHR-KGC (GPT-4o-mini)	0.5	0.3	20	0.6	0.3	20	0.8	0.4	20
LHR-KGC (Qwen-plus)	0.5	0.6	20	0.6	0.2	20	0.7	0.4	20
LHR-KGC (Qwen-max)	0.6	0.4	20	0.5	0.2	20	0.6	0.4	20
LHR-KGC (DeepSeek-v3)	0.4	0.4	20	0.5	0.3	20	0.7	0.4	20

6 实验结果与分析

6.1 基线模型

对于Wiki27K和FB15K-237-N数据集,选择与

以下几种基线模型进行比较:(1)基于结构的方法,包括TransE、ConvE、TuckER和RotatE。(2)基于文本的模型,如KG-BERT、LP-RP-RR、PKGK、VKGC和KGC-ERC。对于Wikidata5M数据集,选择与以下几种基线模型进行比较:(1)基于结构的方

法,包括 TransE、DistMult、SimplE、RotatE、QuatE 和 ComplEx。(2)基于文本的方法,包括 KEPLER、KGT5、ReSKGC、KGT5-context、SimKGC、SKG-KGC、KnowC 和 KGC-ERC,其中 KGT5、ReSKGC、KGT5-context 以及 KGC-ERC 均以生成式语言模型作为其核心模型。这些基线模型涵盖了从传统基于结构的方法到最新的生成式语言模型,为 LHR-KGC 提供了一个具有挑战性的比较基准。

6.2 主要结果

在 Wiki27K、FB15K-237-N 和 Wikidata5M 数据集上的实验结果分别如表 4 和表 5 所示。实验结果清晰地表明,LHR-KGC 在所有选定的数据集上均展现出显著优于基线模型的性能,取得了最高的 MRR、Hits@1 和 Hits@3,对于本地 KGC 模型 KGC-ERC 有较大的性能提升。在多个不同 LLM 上的实验结果也体现了 LHR-KGC 的有效性、稳定性以及通用性。

表 4 在 Wiki27K 和 FB15K-237-N 数据集上的性能比较结果

模型	Wiki27K				FB15K-237-N			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
基于结构的方法								
TransE ^{[5]*}	0.155	0.032	0.228	0.378	0.255	0.152	0.301	0.459
ConvE ^{[23]*}	0.226	0.164	0.244	0.354	0.273	0.192	0.305	0.429
TuckER ^{[43]*}	0.249	0.185	0.269	0.385	0.309	0.227	0.340	0.474
RotatE ^{[7]*}	0.216	0.123	0.256	0.394	0.279	0.177	0.320	0.481
基于文本的方法								
KG-BERT ^{[8]*}	0.192	0.119	0.219	0.352	0.203	0.139	0.201	0.403
LP-RP-RR ^{[26]*}	0.217	0.138	0.235	0.379	0.248	0.155	0.256	0.436
PKGK ^[20]	0.285	0.230	0.305	<u>0.409</u>	0.332	0.261	0.346	<u>0.487</u>
VKGC ^[44]	0.296	0.241	0.320	0.422	0.338	0.266	0.369	0.489
KGC-ERC ^[32]	0.305	0.267	0.322	0.381	0.343	0.280	0.375	0.465
LHR-KGC (GPT-4o-mini)	0.322	0.284	0.344	0.393	0.362	0.301	0.395	0.475
LHR-KGC (Qwen-plus)	<u>0.334</u>	<u>0.297</u>	0.357	0.400	0.367	<u>0.303</u>	0.407	0.484
LHR-KGC (Qwen-max)	0.333	0.296	<u>0.354</u>	0.400	0.367	0.304	<u>0.403</u>	0.484
LHR-KGC (DeepSeek-v3)	0.337	0.302	0.357	0.399	<u>0.365</u>	<u>0.303</u>	0.400	0.480

注:带*符号后缀的结果来自 PKGC^[20],其他结果均来自对应文献,最优结果加粗表示,次优结果加下划线表示。

在规模较小的数据集 Wiki27K 和 FB15K-237-N 中,LHR-KGC 展现了显著的性能提升。具体而言,在 Wiki27K 数据集上,LHR-KGC (DeepSeek-v3) 取得了最高的 MRR 值 0.337 和 Hits@1 值 0.302,相较于本地 KGC 模型 KGC-ERC, MRR 提升了 10.5%, Hits@1 提升了 13.1%。在 FB15K-237-N 数据集上,LHR-KGC (Qwen-plus) 和 LHR-KGC (Qwen-max) 均取得了最高的 MRR 值 0.367,相较于本地 KGC 模型的 0.343,提升幅度达到 7.0%。Hits@1 指标从本地 KGC 模型的 0.280 提升至 0.303 和 0.304,提升比例为 8.2% 和 8.6%。这些结果表明,LHR-KGC 框架能够有效地利用 LLM 对本地 KGC 模型的预测结果进行优化,精准地识别和调整候选实体的排序,从而提升 KGC 的性能。

在大规模数据集 Wikidata5M 中,LHR-KGC 依旧保持稳定且出色的性能表现,LHR-KGC (DeepSeek-v3) 取得了最高的 MRR 值 0.436 和

Hits@1 值 0.416。这些结果进一步验证了 LHR-KGC 框架在大规模数据集上的有效性。值得注意的是,LHR-KGC 框架在四个不同的 LLM (GPT-4o-mini、Qwen-plus、Qwen-max 和 DeepSeek-v3) 上均表现出色,这表明该框架具有良好的通用性和稳定性,能够与不同的 LLM 有效结合,实现对本地 KGC 模型预测结果的优化。

整体上看,LHR-KGC 框架在优化本地 KGC 模型预测结果方面呈现出较高的效能。该框架借助 LLM 强大的语义理解与逻辑推理能力,对本地 KGC 模型生成的候选实体列表进行重排序,在一定程度上提高预测结果的准确性与可靠性。以 Wiki27K 数据集为例,本地 KGC 模型 KGC-ERC 的 MRR 和 Hits@1 分别 0.305 和 0.267,经 LHR-KGC 优化后提升至 0.337 和 0.302,提升幅度分别达 10.5% 和 13.1%。这些结果充分展示了 LHR-KGC 框架在优化本地 KGC 模型预测结果方面的强大能力。

表 5 在 Wikidata5M 上的性能比较结果

模型	MRR	Hits@1	Hits@3	Hits@10
基于结构的方法				
TransE*	0.253	0.170	0.311	0.392
DistMult*	0.253	0.209	0.278	0.344
Simple*	0.296	0.252	0.317	0.377
RotatE*	0.290	0.234	0.322	0.390
QuatE*	0.276	0.227	0.301	0.359
ComplEx**	0.308	0.255	-	0.398
基于文本的方法				
KEPLER	0.210	0.173	0.224	0.277
KGT5	0.300	0.267	0.318	0.365
ReSKGC	0.363	0.334	0.386	0.416
SimKGC	0.358	0.313	0.376	0.441
SKG-KGC	0.366	0.323	0.382	0.446
KGT5-context	0.426	0.406	0.440	0.460
KnowC	0.426	0.373	0.447	0.525
KGC-ERC	0.433	0.412	0.448	0.467
LHR-KGC (GPT-4o-mini)	0.434	0.414	0.450	0.466
LHR-KGC (Qwen-plus)	<u>0.435</u>	<u>0.415</u>	0.450	0.467
LHR-KGC (Qwen-max)	0.434	0.413	0.452	<u>0.468</u>
LHR-KGC (DeepSeek-v3)	0.436	0.416	<u>0.451</u>	0.467

注：带*符号后缀的结果来自 Graphvite^[45]提供的最佳预训练模型，带**符号后缀的结果来自文献[46]，其他结果均来自对应文献。最优结果加粗表示，次优结果加下划线表示。

此外，不同 LLM 在 LHR-KGC 框架下的实验结果展现出高度一致性。实验结果表明，LHR-KGC 在

GPT-4o-mini、Qwen-plus、Qwen-max 和 DeepSeek-v3 四个 LLM 上均取得了优异的性能，且绝对数值上差异很小，这表明该框架并非只能选用某一个 LLM，而是能够与多种 LLM 有效结合。例如，在 FB15K-237-N 数据集上，LHR-KGC (Qwen-plus) 和 LHR-KGC (Qwen-max) 均取得了最高的 MRR 值 0.367，而表现最差的 LHR-KGC (GPT-4o-mini) 仅低 0.005，三者在 Hits@1 上的差异更缩小至 0.003。这种稳定性在大规模数据集 Wikidata5M 中同样显著。图 2 的可视化结果进一步验证了框架的通用性与稳定性。在四个关键指标 (MRR、Hits@1、Hits@3、Hits@10) 上，不同 LLM 的性能表现非常接近，这在直观反映了该框架对不同 LLM 的兼容性。

综上所述，全面评估所有实验结果，明确展示了 LHR-KGC 框架在大多数评估场景中的显著有效性。无论是在小型数据集还是大规模数据集上，LHR-KGC 框架均能有效地利用 LLM 对本地 KGC 模型的预测结果进行优化，从而实现有效的性能提升。此外，该框架在不同 LLM 上的稳定表现也为其在实际应用中的广泛适用性提供了有力保障。

6.3 混合重排序有效性分析

为深入解析混合重排序策略的性能优势，本节基于 Wiki27K 数据集对本地 KGC 模型 (KGC-ERC) 与 LLM (GPT-4o-mini) 的 Top-1 预测结果进

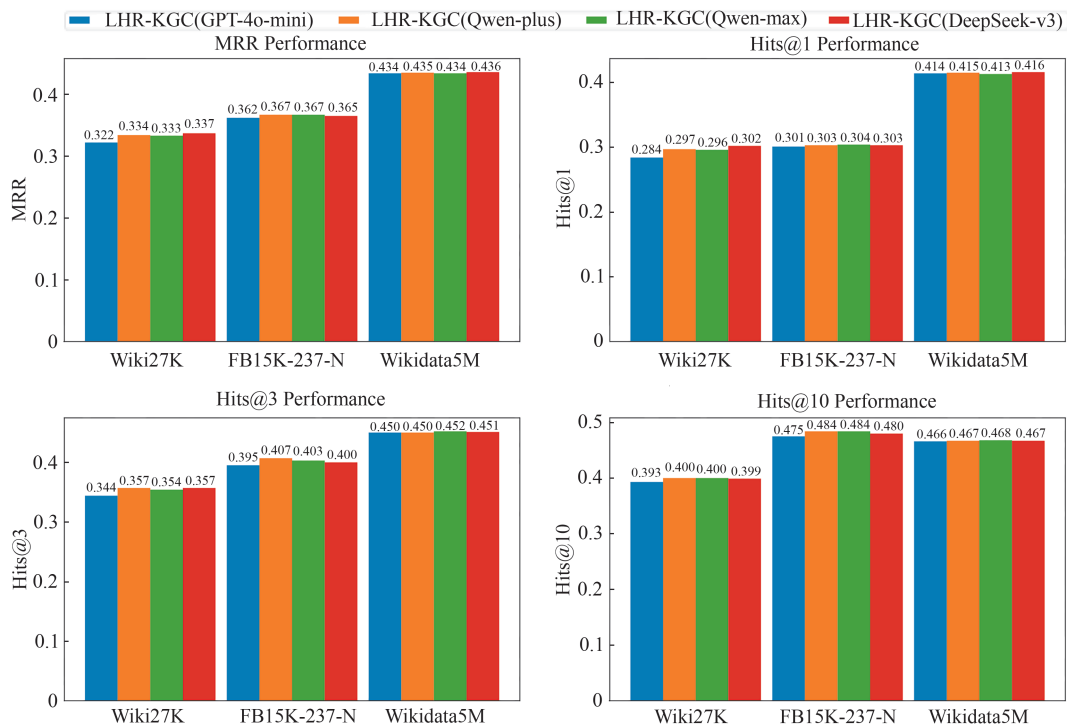


图 2 LHR-KGC 使用不同 LLM 的性能对比情况

行对比分析。如图3所示,通过韦恩图可视化展示了两种模型的预测分布特征。

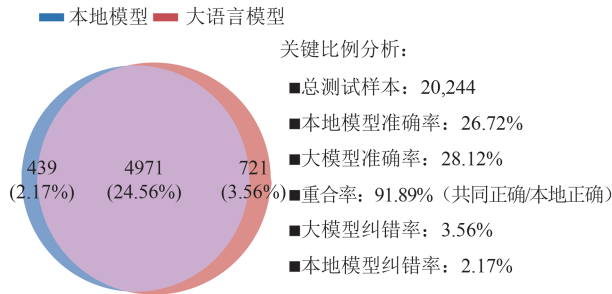


图3 Wiki27K数据集 Top-1 预测正确性分布

注:比例计算基于总测试三元组数(20,244),包含头/尾实体预测任务。本地KGC模型预测正确5410例,独有正确439例;LLM预测正确5692例,独有正确721例;两者共同正确4971例。

6.3.1 预测重合度分析

(1)本地KGC模型与LLM共同预测正确的三元组共4,971例,占本地KGC模型正确预测的91.89%;

(2)LLM独有的正确预测721例,成功纠正了本地KGC模型3.56%的预测偏差;

(3)本地KGC模型独有的正确预测439例,保留了LLM遗漏的2.17%正确答案。

6.3.2 互补性机制解析

上述分布提示了两类模型可能存在互补性,这一现象在以往的研究中也有所讨论。尽管实验结果无法直接分离并验证各自的内部机制,但结合已有研究,可从以下角度进行理解:

(1)本地KGC模型通常更擅长利用知识图谱结构性信息,在特定的知识图谱内部具有较强的模式捕捉能力;

(2)LLM具有更丰富的外部知识储备与语义推理能力,在处理复杂语义关联时可弥补本地KGC模型的不足。

因此,混合重排序策略在实验中表现更稳定,可能源于两类模型在不同信息来源上的互补性。需要指出的是,该解释属于基于文献与现象的推断,而非由本研究的实验直接证明。

6.3.3 参数调节机制

混合策略通过参数 α 实现动态权重分配:

(1)当LLM提供强语义证据时,降低 α 放大其纠错能力;

(2)当LLM受文本相似性干扰时,提高 α 维持本地KGC模型的稳定性。

此动态调节机制是LHR-KGC在复杂场景中保持鲁棒性的关键,当LLM缺乏特定于知识图谱的领域知识时,提高 α 可有效抑制噪声干扰。

6.4 消融实验与分析

前述混合重排序有效性分析揭示了混合策略的理论优势,本节通过消融实验量化验证各组件的实际贡献。为验证LHR-KGC框架中各关键组件的必要性,设计消融实验对比多种配置:

(1)Local Model Only:是指仅使用本地KGC模型KGC-ERC进行实体预测;

(2)LLM Re-ranking:是指使用LLM直接对本地KGC模型的预测结果进行重排序,并以LLM的输出作为最终结果;

(3)Hybrid Re-ranking:是指分别采用本地KGC模型预测结果与LLM重排序结果进行混合重排序。

实验结果如表6所示,不同配置在各数据集上的差异化表现揭示了框架各组件的作用。

表6 消融实验结果

配置	Wiki27K		FB15K-237-N		Wikidata5M	
	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1
Local Model Only	0.305	0.267	0.343	0.280	0.433	0.412
LLM Re-ranking						
GPT-4o-mini	0.321	0.281	0.360	0.298	0.427	0.401
Qwen-plus	0.333	0.296	0.357	0.289	0.431	0.408
Qwen-max	0.330	0.293	0.361	0.299	0.428	0.403
DeepSeek-v3	0.336	0.301	0.361	0.299	0.432	0.410
Hybrid Re-ranking						
GPT-4o-mini	0.322	0.284	0.362	0.301	0.434	0.414
Qwen-plus	0.334	0.297	0.367	0.303	0.435	0.415
Qwen-max	0.333	0.296	0.367	0.304	0.434	0.413
DeepSeek-v3	0.337	0.302	0.365	0.303	0.436	0.416

注:第一组Local Model Only表示仅使用本地KGC模型的结果;第二组LLM Re-ranking表示直接使用LLM重排序后的结果;第三组Hybrid Re-ranking表示使用LLM与本地KGC模型结合后的混合重排序的结果。每组最优结果加粗表示。

在Wiki27K和FB15K-237-N数据集上,从仅使用本地KGC模型到直接使用LLM重排序,再到混合重排序的配置,性能呈现出递增趋势。在FB15K-237-N中,本地KGC模型的MRR为0.343,LLM重排序(Qwen-max)提升至0.361,混合重排序(Qwen-max)进一步提高到0.367;Hits@1指标从本地KGC模型的0.280,经LLM重排序(如Qwen-max)提升到0.299,混合重排序(Qwen-max)达到

0.304。在Wiki27K数据集上也有类似趋势。这一现象表明,引进LLM进行重排序是有效的,能够在一定程度上优化本地KGC模型的结果。而混合重排序则在此基础上进一步提升性能,显示出其独特的优势。

在Wikidata5M数据集上,呈现出与前两个数据集不同的现象,直接使用LLM进行重排序的结果甚至不如本地KGC模型。其中表现最佳的LLM重排序(DeepSeek-v3)的MRR为0.432,略低于本地KGC模型的0.433;Hits@1为0.410,同样略低于本地KGC模型的0.412。然而,采用混合重排序后,各指标均达到最优结果,例如混合重排序(DeepSeek-v3)的MRR为0.436,Hits@1为0.416,均高于本地KGC模型。这表明单纯依赖LLM并不一定能直接优化本地KGC模型的输出。然而,当采用混合重排序策略时,性能得到有效提升,达到了最优水平。这一结果凸显了混合重排序策略的重要性,其不仅能够结合本地KGC模型和LLM的优势,还能有效降低单一模型失效的风险,确保在不同数据集上都能获得稳定且优异的性能。

综上所述,消融实验结果有力地证明了LHR-KGC框架中各组件的重要性。LLM重排序为提升性能提供了新的途径,而混合重排序则是一种更为可靠和有效的方法,其不仅在性能上优于单独使用本地KGC模型或LLM,还能够根据不同数据集的特点,灵活整合本地KGC模型和LLM的优势,实现稳定的性能增强。这一框架设计不仅适用于多种数据集,还为解决单一模型在复杂场景下的局限性提供了可行的解决方案。

6.5 混合重排序策略的敏感性与适用边界分析

在前述实验结果与消融分析中,本文已验证了LHR-KGC框架在多个数据集和不同LLM配置下的整体有效性。然而,仅从宏观性能指标出发,仍不足以全面刻画混合重排序策略在不同条件下的行为特征及其适用范围。鉴于此,本节从机制层面对混合重排序策略的权重敏感性、适用边界以及典型失效场景进行进一步分析,以加深对方法内在工作机制的理解。

6.5.1 融合权重的敏感性分析

LHR-KGC采用混合重排序策略,对本地KGC模型得分与LLM重排序结果进行加权融合。该策略的核心在于通过权重参数在结构信息与语义推理能力之间取得平衡,从而实现更稳健的预测排序结果。从已有实验与消融结果可以观察到,融合权重

并非对性能提升呈线性影响,而是在一定区间内表现出相对稳定的增益趋势。

当融合权重过度偏向本地KGC模型时,最终排序结果主要受限于原始模型的结构表示能力,难以充分发挥LLM在复杂语义推理和隐含关系理解方面的优势;相反,当权重过度偏向LLM时,虽然部分语义相关性较强的候选实体排名有所提升,但在关系约束较强或结构模式明显的场景中,容易引入语义上合理但结构上不一致的候选,导致整体排序稳定性下降。实验结果表明,在不同数据集和不同LLM配置下,融合权重在一定范围内变化时,性能指标保持相对平稳,说明该混合策略对权重参数具有一定的鲁棒性,但仍存在合理的取值区间。

这一现象表明,混合重排序策略的优势并不来源于对单一模型的完全替代,而在于对结构约束与语义推理能力的协同利用。合理的权重设置有助于在两类信息源之间建立有效平衡,从而实现稳定且可解释的性能提升。

6.5.2 不同关系类型下的适用性分析

混合重排序策略在不同类型关系上的表现存在一定差异,这与关系本身的结构复杂性和语义依赖程度密切相关。对于模式较为规则、结构约束明确的关系(如一对一或对称关系),本地KGC模型通常能够给出较为可靠的候选排序,此时混合策略更多起到稳定排序和微调排名的作用,性能提升相对有限但较为稳定。

相比之下,在涉及多跳语义推理、抽象语义关联或上下文依赖较强的关系类型中,LLM的语义理解能力能够有效弥补本地KGC模型在结构建模方面的不足。混合重排序在此类关系上往往表现出更明显的性能增益,说明该策略在处理复杂关系推理任务时具有更高的适用价值。

与此同时,对于语义歧义较强、文本描述信息不足或关系定义本身较为宽泛的场景,混合重排序的效果可能出现波动。一方面,LLM可能受到语义相似性干扰;另一方面,本地KGC模型提供的结构约束又难以完全消除歧义。这类关系构成了混合策略的典型适用边界之一。

6.5.3 典型错误案例与失效场景分析

尽管LHR-KGC在整体上显著提升了候选实体的排序质量,但是在部分场景中仍然可能出现重排序效果受限甚至失效的情况。通过对预测结果的定性分析可以发现,典型失效场景主要集中在以下几类情况中。

首先,当本地KGC模型未能在候选列表中召回正确实体时,混合重排序无法从根本上修正这一错误。这一问题源于框架“先召回、后排序”的基本假设,其性能上限天然受限于候选实体集合的质量。其次,在候选实体语义高度相似但结构角色差异较大的情况下,LLM可能倾向于提升语义相关性较高但结构上不匹配的实体排名,从而对最终排序产生负面影响。再次,对于长尾实体或文本描述极度稀疏的节点,LLM可利用的语义线索有限,混合策略的增益相对有限。

上述失效场景表明,混合重排序并非对所有错误类型均具备同等修正能力,其优势主要体现在“候选正确但排序不足”的情形,而非“候选缺失”的根本性错误。

6.5.4 泛化能力与方法适用边界讨论

从泛化能力角度来看,LHR-KGC框架不依赖特定的本地KGC模型或LLM,其混合重排序策略在多个数据集和不同模型组合下均表现出一致的性能增益,说明该方法具有良好的跨模型与跨数据集适用性。然而,其适用边界同样较为明确:当本地KGC模型的候选召回能力较弱,或关系与实体语义信息高度稀疏时,混合重排序的提升空间将受到显著限制。

此外,融合权重的合理区间、候选实体数量的选择以及提示词模板的表达方式,均会对方法的最终表现产生影响。这些因素共同决定了混合重排序策略在不同应用场景下的适用性,需要在实际部署中结合具体任务需求进行权衡。

综上所述,混合重排序策略在合理假设条件下能够稳定发挥结构建模与语义推理的互补优势,但其效果仍受候选质量、关系类型及融合机制设置等因素制约。上述分析从机制层面明确了LHR-KGC框架的适用边界,也为后续从系统视角讨论其整体局限性提供了理论依据。

6.6 LHR-KGC框架的局限性分析与讨论

6.6.1 性能的上限约束

在前文已从混合重排序机制层面分析其适用边界的基础上,进一步从整体系统视角审视LHR-KGC框架,其在实际应用中仍存在若干不可忽视的局限性。从本质上讲,LHR-KGC框架依赖本地KGC模型生成的候选实体列表进行后续重排序操作,这意味着本地KGC模型输出的质量直接划定了框架性能的天花板。以Wikidata5M数据集为例,如表5所示,本地KGC模型KGC-ERC在该数据集上

的Hits@1为0.412,Hits@10为0.467。如果只对Top-10的结果进行重排序,那么,其理论上最多只能将Hits@1提升至0.467,实际应用中往往难以达到这一上限。这清晰地表明,本地KGC模型若无法提供高质量、高召回率的候选实体序列,即便LHR-KGC框架运用LLM的强大能力,也无法实现突破性的性能提升。当本地KGC模型在某些复杂关系或长尾实体的识别上存在缺陷时,框架难以通过重排序完全纠正错误,导致最终KGC结果的准确性受限。

此外,增加重排序的候选数量虽然在理论上可以提升性能,但也带来了新的挑战。例如,将重排序的范围从Top-10扩展到Top-20或Top-50,虽然可以增加正确实体出现在候选列表中的概率,但同时也显著增加了推理成本和token消耗。LLM在处理更多候选实体时,需要更多的计算资源和时间,这可能导致效率下降。更重要的是,性能提升并不是随着候选实体数量的增加而线性提升的,而是呈现出收益率递减的趋势。当增加候选实体数量时,干扰性候选实体的数量也会增多,LLM准确识别出正确实体的难度也会加大,导致性能提升的边际效益逐渐降低。因此,在实际应用中,需要在候选实体数量和推理成本之间进行权衡,以实现性能和效率的平衡。

6.6.2 对提示词的依赖

LHR-KGC框架的性能在很大程度上依赖于结构化提示词模板的设计。提示词模板作为本地KGC模型预测结果与LLM之间的桥梁,其表达方式、信息量以及对任务需求描述的清晰度,直接影响LLM对任务的理解和推理能力。尽管LHR-KGC框架在多个LLM上均展现出良好的通用性和稳定性,但不同LLM对提示词的理解和处理能力存在差异,这可能导致框架在不同模型上的表现存在波动。因此,提示词模板的设计不仅需要适应本地KGC模型的输出,还需考虑LLM的特点,以确保最佳的优化效果。未来的研究可以进一步探索如何精细化设计提示词模板,以提高LLM对任务的理解能力,包括优化提示词的表达方式、增加信息量以及适应不同LLM的特点。

6.6.3 扩展性与实际适用性讨论

本节在前文明确混合重排序机制适用边界的基础上,进一步从系统层面讨论其扩展性与实际适用性。尽管LHR-KGC框架不需要对LLM进行微调,但其适用性仍与候选实体数量和调用成本相关。由

于本框架仅对本地KGC模型生成的Top-K候选进行重排序,因此对知识图谱规模(百万至数千万实体)的敏感度主要由本地KGC模型的检索能力决定,而非LLM的推理复杂度。也就是说,即使在大规模知识图谱上,本框架的计算开销依然保持稳定,因为LLM的输入规模始终限制在K个候选实体范围内,而K通常在实际场景中总是远小于知识图谱中实体的总量(如 $K=10$ 或 30)。

从可部署性角度来看,LHR-KGC框架推理仅需要一次LLM调用,不依赖额外训练数据,也不增加检索阶段的复杂度,具有良好的可迁移性和部署效率,可直接集成到现有KGC系统的推理阶段。此外,在无需GPU训练、不新增模型参数、调用开销线性随候选数量变化的前提下,本框架适用于多种实际应用场景,例如在线知识库补全、智能问答系统中的实体生成验证、企业知识工程流程中的质量控制等。

总体而言,LHR-KGC框架提供了一种在大规模知识图谱环境中仍具可行性、可部署性和成本优势的后处理方案。

7 结束语

本文提出了一种基于LLM的知识图谱补全轻量级混合重排序框架LHR-KGC。该框架通过提示学习范式充分结合本地KGC模型与LLM的互补优势,在无需额外训练的条件下实现高效且稳定的重排序优化。在Wiki27K、FB15K-237-N和Wikidata5M三个数据集上的实验结果表明,LHR-KGC在MRR和Hits@1等关键指标上均较本地KGC模型取得显著提升。混合重排序有效性分析和消融实验进一步验证了LLM重排序与混合策略的重要性。LHR-KGC采用的动态权重融合机制能够在保留本地KGC模型结构优势的同时,充分发挥LLM在语义理解与逻辑推理方面的强项。特别是在Wikidata5M数据集上的消融实验中,混合策略的性能提升更加明显,凸显其在大规模复杂知识图谱场景中的必要性。局限性分析展示了LHR-KGC框架存在的瓶颈问题,探讨了提升框架性能的可行方案。未来研究可进一步探索引入高性能的本地KGC模型,以生成高质量、高召回率的候选实体序列,并精细化设计提示词模板,以充分挖掘LLM的潜力,从而提升LHR-KGC框架在更复杂应用场景中的推理表现。

致谢 衷心感谢在审稿过程中对本论文提出宝贵意见和建议的所有评审专家!

参 考 文 献

- [1] Saxena A, Tripathi A, Talukdar P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle, USA, 2020: 4498-4507
- [2] Sun Y, Shi Q, Qi L, et al. JointLK: Joint reasoning with language models and knowledge graphs for commonsense question answering//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Dublin, Ireland, 2022: 5049-5060
- [3] Zhang F, Yuan N J, Lian D, et al. Collaborative knowledge base embedding for recommender systems//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 353-362
- [4] Yang Y, Huang C, Xia L, et al. Knowledge graph contrastive learning for recommendation//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid, Spain, 2022: 1434-1443
- [5] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data//Proceedings of the Advances in Neural Information Processing Systems. New York, USA, 2013, 26
- [6] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction//Proceedings of the International Conference on Machine Learning. New York, USA, 2016: 2071-2080
- [7] Sun Z, Deng Z H, Nie J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space. arXiv preprint arXiv:1902.10197, 2019
- [8] Yao L, Mao C, Luo Y. KG-BERT: BERT for knowledge graph completion. arXiv preprint arXiv:1909.03193, 2019
- [9] Wang L, Zhao W, Wei Z, et al. Simkgc: Simple contrastive knowledge graph completion with pre-trained language models//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long papers). Dublin, Ireland, 2022: 4281-4294
- [10] Yang G, Liu Y, Zhang L, et al. Knowledge context modeling with pre-trained language models for contrastive knowledge graph completion//Proceedings of the Association for Computational Linguistics ACL 2024. Bangkok, Thailand, 2024: 8619-8630
- [11] Liu A, Feng B, Xue B, et al. DeepSeek-v3 technical report. arXiv preprint arXiv:2412.19437, 2024
- [12] Yang A, Yang B, Zhang B, et al. Qwen2.5 technical report. arXiv preprint arXiv:2412.15115, 2024
- [13] Zhu Y, Wang X, Chen J, et al. LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities. World Wide Web, 2024, 27(5): 58

- [14] Yao L, Peng J, Mao C, et al. Exploring large language models for knowledge graph completion//Proceedings of the ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Hyderabad, India, 2025: 1-5
- [15] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020, 21(140): 1-67
- [16] Bosselut A, Rashkin H, Sap M, et al. COMET: Commonsense transformers for automatic knowledge graph construction//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 4762-4779
- [17] Yang Z, Du X, Cambria E, et al. End-to-end case-based reasoning for commonsense knowledge base completion//Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. Dubrovnik, Croatia, 2023: 3509-3522
- [18] Wei Y, Huang Q, Zhang Y, et al. Kicgpt: Large language model with knowledge in context for knowledge graph completion//Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore, 2023: 8667-8683
- [19] Wang Y, Hu M, Huang Z, et al. KC-GenRe: A Knowledge-constrained Generative Re-ranking Method Based on Large Language Models for Knowledge Graph Completion//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). Torino, Italy, 2024: 9668-9680
- [20] Lv X, Lin Y, Cao Y, et al. Do pre-trained models benefit knowledge graph completion? a reliable evaluation and a reasonable approach.//Proceedings of the Association for Computational Linguistics: ACL 2022. Dublin, Ireland, 2022: 3570-3581
- [21] Yang B, Yih W T, He X, et al. Embedding entities and relations for learning and inference in knowledge bases//International Conference on Learning Representations. Banff, Canada, 2014
- [22] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks//Proceedings of the European Semantic Web Conference. Crete, Greece, 2018: 593-607
- [23] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018, 32: 1811-1818
- [24] Chen S, Liu X, Gao J, et al. Hitter: Hierarchical transformers for knowledge graph embeddings//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican Republic, 2021: 10395-10407
- [25] Xie R, Liu Z, Jia J, et al. Representation learning of knowledge graphs with entity descriptions//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix, USA, 2016, 30(1)
- [26] Kim B, Hong T, Ko Y, et al. Multi-task learning for knowledge graph completion with pre-trained language models//Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain, 2020: 1737-1743
- [27] Xie X, Zhang N, Li Z, et al. From discrimination to generation: Knowledge graph completion with generative transformer//Proceedings of the ICLR 2022 Workshops. New York, USA, 2022: 162-165
- [28] Shan Y, Zhou J, Peng J, et al. Multi-level shared knowledge guided learning for knowledge graph completion. *Transactions of the Association for Computational Linguistics*, 2024, 12: 1027-1042
- [29] Saxena A, Kochsiek A, Gemulla R. Sequence-to-Sequence Knowledge Graph Completion and Question Answering//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Dublin, Ireland, 2022: 2814-2828
- [30] Yu D, Yang Y. Retrieval-enhanced generative model for large-scale knowledge graph completion//Proceedings of the 46th International ACM Sigir Conference on Research and Development in Information Retrieval. Taipei, China, 2023: 2334-2338
- [31] Kochsiek A, Saxena A, Nair I, et al. Friendly neighbors: Contextualized sequence-to-sequence link prediction//Proceedings of the 8th Workshop on Representation Learning for NLP (RepL4NLP 2023). Toronto, Canada, 2023: 131-138
- [32] Chen J, Zhang K, Gan A, et al. Enhancing knowledge graph completion with entity neighborhood and relation context//International Conference on Database Systems for Advanced Applications. Singapore, 2025: 537-546
- [33] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Advances in Neural Information Processing Systems, 2017, 30
- [34] Li D, Tan Z, Chen T, et al. Contextualization Distillation from Large Language Model for Knowledge Graph Completion//Findings of the Association for Computational Linguistics: EACL 2024. Malta, 2024: 458-477
- [35] Kolluru K, Chauhan M S, Nandwani Y, et al. Cear: Cross-entity aware reranker for knowledge base completion. arXiv preprint arXiv:2104.08741, 2021
- [36] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding//Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (long and short papers). Minneapolis, USA, 2019: 4171-4186
- [37] Touvron H, Lavril T, Izacard G, et al. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023
- [38] Achiam J, Adler S, Agarwal S, et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023
- [39] Dettmers T, Pagnoni A, Holtzman A, et al. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems*, 2023, 36: 10088-10115
- [40] Wang X, Gao T, Zhu Z, et al. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 2021, 9: 176-194
- [41] Toutanova K, Chen D. Observed versus latent features for

- knowledge base and text inference//Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. Beijing, China, 2015: 57-66
- [42] Sun Z, Vashishth S, Sanyal S, et al. A re-evaluation of knowledge graph completion methods//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Seattle, USA, 2020: 5516-5522
- [43] Balažević I, Allen C, Hospedales T. Tucker: Tensor factorization for knowledge graph completion//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019: 5185-5194
- [44] Han S, Guan Z, Li S, et al. Knowledge graph completing with dual confrontation learning model based on variational information bottleneck method//Proceedings of the 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security (QRS). Chiang Mai, Thailand, 2023: 741-750
- [45] Zhu Z, Xu S, Tang J, et al. Graphvite: A high-performance cpu-gpu hybrid system for node embedding//Proceedings of the World Wide Web Conference. San Francisco, USA, 2019: 2494-2504
- [46] Kochsiek A, Gemulla R. Parallel training of knowledge graph embedding models: a comparison of techniques//Proceedings of the 47th International Conference on Very Large Data Bases. Copenhagen, Denmark, 2021, 15(3): 633-645



CHEN Jian-Fang, M. S. His main research interests include knowledge graph, data mining, and large language models.

SHAO Yu-Xuan, B. S. candidate. His main research interests include computer vision, large language models, and data mining.

ZHANG Kai, Ph. D., associate professor. His main research interests include artificial intelligence, natural language processing, and text representation learning.

LIU Qi, Ph. D., professor. His main research interests include data mining and knowledge discovery, machine learning method, and educational big data analysis.

GAN Ao-Ran, Ph. D. candidate. His main research interests include data mining, knowledge graph, and natural language processing.

Background

Knowledge graph completion (KGC) is a fundamental research problem in the field of knowledge engineering and artificial intelligence, aiming to infer missing factual triples in incomplete knowledge graphs. Knowledge graphs provide structured semantic representations of real-world entities and relations, and have been widely adopted in a variety of downstream applications, including intelligent question answering, recommendation systems, and semantic search. However, due to limitations in data acquisition, integration, and continuous updating, most real-world knowledge graphs inevitably suffer from incompleteness, which significantly restricts their practical utility. As a result, improving the quality and reliability of knowledge graph completion remains an important and challenging research topic.

Existing KGC approaches have achieved substantial progress at the international level. Structure-based methods model entities and relations through low-dimensional embeddings and effectively capture graph structural patterns, while text-based methods incorporate contextual semantic information from textual descriptions to enhance reasoning capability. In recent years, large language models (LLMs) have further demonstrated strong

semantic understanding and reasoning abilities, motivating researchers to explore their potential for knowledge graph completion and related tasks. Despite these advances, current methods still face notable limitations. In particular, many models exhibit insufficient ranking accuracy at top positions, where correct entities are often present in candidate lists but fail to be ranked highly enough to be effectively utilized. Moreover, Existing LLM-based solutions frequently rely on model fine-tuning or multi-step interactions, leading to high computational cost, limited deployability, and reduced efficiency in practical scenarios.

This work addresses the above challenges by focusing on the post-processing stage of knowledge graph completion. Instead of designing a new KGC model, the proposed research aims to improve the prediction quality of existing models through a lightweight hybrid re-ranking framework. By integrating the complementary strengths of local KGC models and large language models, the proposed approach enhances candidate ranking without introducing additional training procedures or modifying model parameters. As demonstrated in the experimental results, this framework consistently improves ranking performance and

robustness across multiple datasets and large language models, while maintaining low computational overhead.

The research presented in this paper is conducted under projects related to intelligent knowledge processing and knowledge-driven artificial intelligence systems. These projects aim to enhance the reliability, efficiency, and scalability of knowledge-based technologies in real-world applications.

The significance of this work lies in providing a practical and deployable solution for improving knowledge graph completion systems. By offering a general, lightweight, and training-free re-ranking framework, this study contributes to bridging the gap between advanced language models and existing KGC systems, facilitating their effective integration in large-scale and resource-constrained application scenarios.