



专题：AI赋能通信网络

基于大语言模型与知识图谱融合的多跳问答技术研究

蒋献, 王涵亦, 杨诗婷, 陈星妤, 周梦瑶, 董黎刚
(浙江工商大学统计与数据科学学院, 浙江 杭州 310018)

摘要: 随着大语言模型在自然语言处理任务中的广泛应用, 提升其在垂直领域问答任务中的表现成为当前研究的重点方向。针对传统方法在复杂多跳推理任务中存在的局限, 提出了一种融合知识图谱的多跳问答方法 (LLMKG)。该方法通过在 Prompt 中注入知识图谱中的事实知识, 有效提升了大语言模型在特定领域问答中的推理能力。在 COKG-DATA 数据集上的对比实验显示, LLMKG 在 Hits@1 指标上较最优基线模型提升了 3.5%。该方法具备零样本能力, 适用于各类大语言模型, 且无须额外参数更新。还探讨了时间性知识增强与多模态知识融合的潜力, 并提出构建多模态知识图谱 (MMKG) 作为未来的发展方向。该方法为面向垂直领域的智能问答系统提供了新的研究思路和可行的技术路径。

关键词: 大语言模型; 知识图谱; 多跳问答; 多跳推理

中图分类号: TP301

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025238

Multi-hop question answering by integrating large language models and knowledge graphs

JIANG Xian, WANG Hanyi, YANG Shiting, CHEN Xingyu, ZHOU Mengyao, DONG Ligang
School of Statistics and Data Science, Zhejiang Gongshang University, Hangzhou 310018, China

Abstract: With the widespread application of large language models (LLM) in natural language processing tasks, improving their performance in domain-specific question answering has become a key research focus. To address the limitations of traditional methods in complex multi-hop reasoning tasks, a knowledge-graph-enhanced multi-hop question answering approach, LLMKG, was proposed. Factual knowledge from knowledge graphs was injected into Prompt to enhance the reasoning capability of LLM in vertical domains. Comparative experiments conducted on the COKG-DATA dataset show that LLMKG outperforms the best baseline by 3.5% in terms of Hits@1. The method operates in a zero-shot setting, requires no additional parameter updates, and is applicable to various types of LLM. Temporal knowledge enhancement and multimodal knowledge fusion strategies were further explored, and a multi-modal knowledge graph (MMKG) was proposed as a future direction. This approach offers a novel and effective pathway for

收稿日期: 2025-05-29; 修回日期: 2025-07-10

通信作者: 董黎刚, donglg@zjgsu.edu.cn

基金项目: 浙江省领雁研发项目 (No.2025C02024)

Foundation Item: Zhejiang Lingyan Research and Development Project (No.2025C02024)



advancing intelligent question answering systems in specialized domains.

Key words: large language model, knowledge graph, multi-hop question answering, multi-hop reasoning

0 引言

近年来，随着大语言模型（large language model, LLM）在语言理解与生成任务中的广泛应用，基于其强大语言建模能力的问答系统成为智能搜索与人机交互领域的研究热点。尽管LLM在通用问答任务中已取得显著进展，但在垂直领域、复杂语义理解及多跳推理等任务中仍面临知识不足与推理链条不完整的问题。这种限制主要源于其知识获取方式依赖预训练语料，缺乏与结构化知识源的高效交互能力。

知识图谱（knowledge graph, KG）作为结构化知识的有效表达方式，具有强语义关联性和显式可追溯性，为增强语言模型的知识表示与推理能力提供了可行路径。近期研究逐步探索将知识图谱中的实体和关系注入语言模型，从而提升其在复杂问答场景中的表现。其中，基于自然语言提示（Prompt）的知识注入策略因其轻量、高效、易迁移的优势而引起广泛关注。然而，如何构建适用于多跳问答任务的Prompt模板，并实现知识图谱与LLM之间的无缝衔接，仍是当前研究中的重要问题。

现有方法在多跳推理场景中仍存在诸多瓶颈，例如，缺乏对问题驱动的事实链条建构机制，导致模型生成路径模糊、推理过程不可控；在涉及事件演化与时间限制的问题中，由于缺乏显式时序表示，模型难以准确把握事实的时间逻辑；同时，传统知识图谱大多局限于结构化文本模态，在多模态语境中面对图像、视频等异构信息时无法提供统一的表示与推理支撑，严重制约了其在真实应用场景中的泛化能力。

为了解决上述问题，本文提出了一种面向多跳问答任务的大语言模型与知识图谱融合推理框

架，旨在实现更高效、更精准、更可解释的复杂问答能力。不同于以往仅通过实体检索或静态三元组注入实现知识增强的做法，本文方法从“结构逻辑引导+语义适配”双重视角出发，设计了一套高度适配多跳推理过程的Prompt构建机制，可将知识图谱中与问题语义强相关的多跳事实路径以结构化方式动态嵌入语言模型输入中，形成显式推理链条。这一机制不仅在保持语言生成能力的同时增强了知识可控性，还显著提升了推理过程的连贯性和可解释性。通过该方式，模型能够在理解复杂问题语义的基础上，借助图谱的实体关系结构构建起推理支撑路径，从而在多跳问答中实现对高阶逻辑关系的有效建模与精准回答，展现出较现有方法更强的结构感知能力与事实一致性。

相比现有研究，GMeLLo^[1]通过引入图谱记忆机制提升LLM对外部知识的保留能力，Paths-over-Graph方法^[2]则通过多跳路径构建显式引导模型推理逻辑，均在知识增强方向上取得一定突破。但前者在路径组织上的结构控制能力有限，后者则未充分考虑时序性与知识裁剪等实际应用中的关键问题。本文提出的方法首次结合了“问题驱动的多跳路径显式注入”机制与动态子图构建策略，有效提升了推理的准确性与连贯性，并显著降低了知识注入过程中的冗余与计算开销，在多跳问答任务中展现出更强的结构控制力与实用性。

1 相关工作

随着LLM在自然语言处理领域的持续突破，如何将其与结构化知识图谱深度融合，以提升问答系统的准确性、可解释性与知识覆盖率，已成为当前研究的热点方向。特别是在多跳问答任务

中,模型需基于多个知识片段完成链式推理,对LLM的知识感知、推理路径选择及语义组织能力提出了更高要求。

1.1 知识图谱与LLM融合策略

知识图谱与大语言模型融合的核心在于增强LLM对结构化知识的感知与利用能力。早期研究主要聚焦于将KG三元组转化为Prompt,通过提示工程引导LLM生成更具知识依据的结果。例如,CogMG框架^[3]通过设计“知识感知、信息融合与生成优化”三阶段模块,实现了KG与LLM的深度协同,缓解了知识更新滞后与覆盖不足的问题。当前主流的知识注入范式被系统总结^[4],涵盖实体对齐、嵌入融合和提示设计等方法,为多模态信息融合提供了理论指导。在无监督知识注入场景下,PURE(prompt-based unsupervised relation extraction)模型^[5]被提出,通过构建上下文感知的提示模板进行关系抽取,有效降低了知识抽取过程中的幻觉问题。

近年来,随着任务复杂性的提升,融合方式也进一步向“问题驱动+检索优化+推理链生成”的方向演进。Linders & Tomczak^[6]提出的KG-RAG方法,集成问题分解模块与KG检索机制,构建多跳推理链条,不仅提升了生成准确性,也增强了推理路径的可解释性。Lu等^[7]提出了KEDKG(knowledge editing with dynamic KGs for multi-hop QA),引入动态知识图进行知识修订,并结合细粒度实体关系检索显著提升了LLM的生成质量,在动态信息环境中展现出更强的鲁棒性。

1.2 推理路径构建与多跳机制优化

针对多跳推理任务中路径建模的挑战,研究者尝试引入显式路径规划来提升模型的逻辑连贯性。Tan等^[2]提出了PoG(paths-over-graph)方法,通过在KG中预先构建推理路径,引导LLM生成过程遵循事实链,有效增强了可解释性和精度。相较之下,Xu等^[8]提出的GoG(generate-on-graph)方法更具前瞻性:他们将LLM视为图谱

生成器 and 使用者双重角色,借助动态填补知识图谱缺口以应对知识不完备的挑战,该方法在多个不完整知识图谱问答基准上展现了优异的泛化能力。此外,最新研究CuriousLLM框架提出了一种“好奇心驱动”的多跳推理范式,训练基于Mistral-7B的LLM代理主动生成后续问题,引导检索路径构建,而非直接预测缺失证据^[9]。该方法引入动态终止机制,当模型判断已获取充分信息时输出“NA”终止检索,显著降低了延迟与无关信息干扰。此外,研究者还构建了Follow-upQA数据集,以训练LLM代理在桥接、比较等问题中中学会构造逻辑合理的后续查询。实验表明,CuriousLLM在HotpotQA等4个多跳问答基准上准确率平均提升3%,同时运行效率优于传统KGP-T5方法,展现了路径建模与检索效率的双重提升潜力。

1.3 复杂问题建模与原子级知识检索

为了提升LLM处理复杂、嵌套问句的能力,Li等提出了“问题分解树”策略^[10],将复杂问句分解为多个原子问题,在KG中分别检索相关信息,再交由LLM综合生成答案。该方法不依赖大规模微调,在Few-shot场景下表现出强大的跨任务迁移能力。在教育场景中,另一项由Li团队推动的工作^[10]将教科书内容结构化为图谱后,构建知识驱动的Prompt,用以提升LLM在教材问答任务中的准确率。实验结果表明,该方法在CK12-QA数据集上带来了5.7%的性能提升,展示了结构化知识在增强模型领域适应性的潜力。

同时,为了提升LLM在结构化知识图谱上的路径建模能力,Shu等^[11]提出KG-LLM框架,通过将多跳实体关系链转化为自然语言提示(如“A的X是B,B的Y是C”),并结合指令微调(IFT)与上下文学习(ICL),实现多跳路径的生成式预测。该方法支持Flan-T5、LLaMA2和Gemma等模型,借助特殊标记增强对KG中实体与关系的辨识能力。在多跳链接预测与关系补全



任务中，KG-LLM在WN18RR、FB15k-237等多个标准数据集上取得显著领先效果。实验表明，在5跳推理中其性能下降最小，且具备良好的零样本泛化能力，表明该框架在复杂路径建模场景下具有优越的解释性与扩展性。

2 方法概述

在COKG-DATA数据集上的实验证明，所提方法在零样本设定下，相较于当前最优基线模型，其Hits@1性能提升达3.5%。该方法无须对LLM进行参数更新，具备良好的通用性与可迁移性。本文为提升LLM在知识驱动的问答任务中的表现提供了新思路，也为多源异构知识与语言模型的高效融合提供了理论基础与实践路径。

为了减少LLM产生幻觉的现象，本文利用知识图谱中的事实信息来增强LLM回答问题的准确性。本节设计的LLMKG方法主要包括3个步骤：信息检索、知识生成和问题回答。通过信息检索，可以筛选和获取与用户查询相关的真实信息，从而提高LLM回答问题的可靠性。此外，为了使LLM能够有效理解知识图谱中的信息，提出了基于简单连接的知识库生成方法，将结构化知识转换为LLM可理解的自然语言文本，形成只与该查询相关的知识库。在最后的问题回答

阶段，根据提出的问题和相关知识库利用LLM进行回答，从而提高回答的准确性。融合LLM和知识图谱的多跳问答方法总览如图1所示。

LLMKG方法具体的流程如下：首先，根据给定的问题识别问题文本中的主题实体，并在知识图谱中检索与主题实体相关的关系。接着通过基于Prompt的关系筛选方法得到与问题相关的top-K个关系。然后，根据上一步得到的top-K个关系，提取主题实体和top-K个关系相关的所有三元组，并采用本文提出的简单连接策略，将提取的三元组转化为语义一致的文本描述，构建与问题相关的知识库。最后，以Prompt的形式将构建的知识库添加到问题中，将其转发给LLM以生成最终的答案。

在执行多跳问答时，构建出与问题相关的知识库后，系统将通过知识判断策略评估其是否包含足够信息以回答问题。若信息充足，则生成答案；否则，对三元组中的尾实体进行下一跳迭代。在这个过程中，LLMKG可以在多层级知识结构中灵活导航，以确保能够全面而精准地提供复杂问题的答案。这种动态的评估和决策机制使得多跳问答系统更智能化，能够适应问题不同复杂程度的需求。

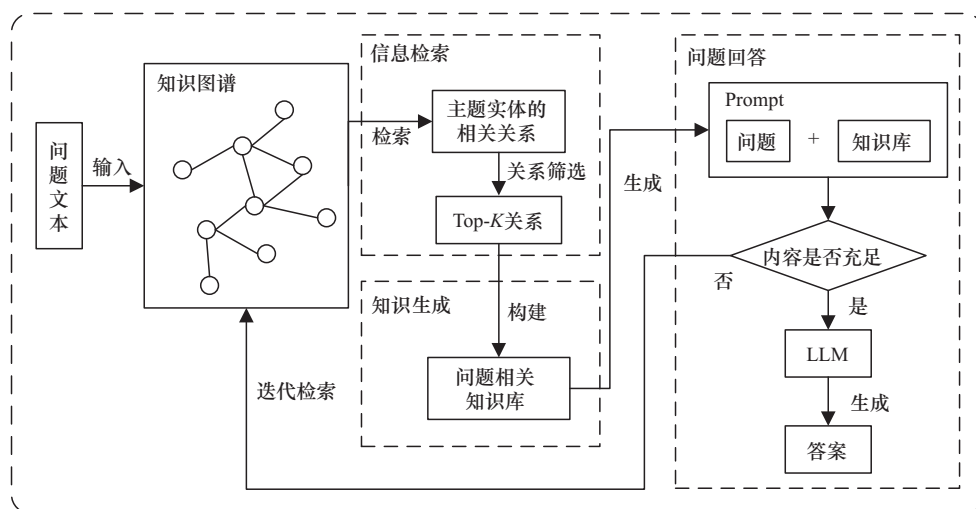


图1 融合LLM和知识图谱的多跳问答方法总览

3 基于 Prompt 的关系筛选

为了降低 LLM 产生幻觉的可能性，本文通过信息检索有效地筛选和获取与给定问题相关的信息，并将检索到的信息处理后传入 LLM，从而增强了 LLM 在回答问题时的可靠性。首先，模型在信息检索时会根据给定问题 q 识别出问题文本中的主题实体 e_h ，并检索整个知识图谱 G 中与主题实体相关的关系路径 z 。信息检索模块的最终目标是在知识图谱 G 中检索出能回答 q 的推理路径 p_z 的集合 P_z ，且检索过程是由问题的主题实体 e_h 开始的， P_z 的求解过程为：

$$P_z = \left\{ p_z(e_h, r^*, e^*) \mid p_z(e_h, r^*, e^*) = e_h \rightarrow_{r_1} e_1 \rightarrow_{r_2} \dots \rightarrow_{r_l} e_a, p_z(e_h, r^*, e^*) \in G \right\} \quad (1)$$

其中， (e_h, r^*, e^*) 表示一个三元组，关系 r^* 连接了 e_h 和 e^* ， e_a 表示最后得到的答案实体。

由于知识图谱中往往涵盖了丰富的事实知

识，在检索知识图谱时，通常会获得与主题实体相关的大量关系。如果不对这些关系进行筛选，可能导致下游知识生成任务中包含大量与问题无关的信息，影响下游问答任务的性能。因此，为了应对这一问题，本节提出一种基于 Prompt 的关系筛选策略，该策略通过过滤与问题无关的关系，限制 LLM 生成答案的范围，降低 LLM 生成虚假信息的可能性，从而改善问答系统的性能。

研究者发现，与 LLM 进行交互时，通过设置合适的 Prompt，可以有效缓解幻觉的现象，使得模型能够更加专注地理解并满足用户的实际任务需求，从而显著提高模型的性能表现^[12]。先前的研究和实践充分证明，高质量的 Prompt 应当具备多个关键特征，包括清楚的任务描述、精准的语言表达，以及提供一些案例提示等。这些特征不仅能够有效引导 LLM 的学习和生成过程，还能够确保模型对于不同任务和用户输入的适应性。基于 Prompt 的关系筛选示例如图 2 所示。本

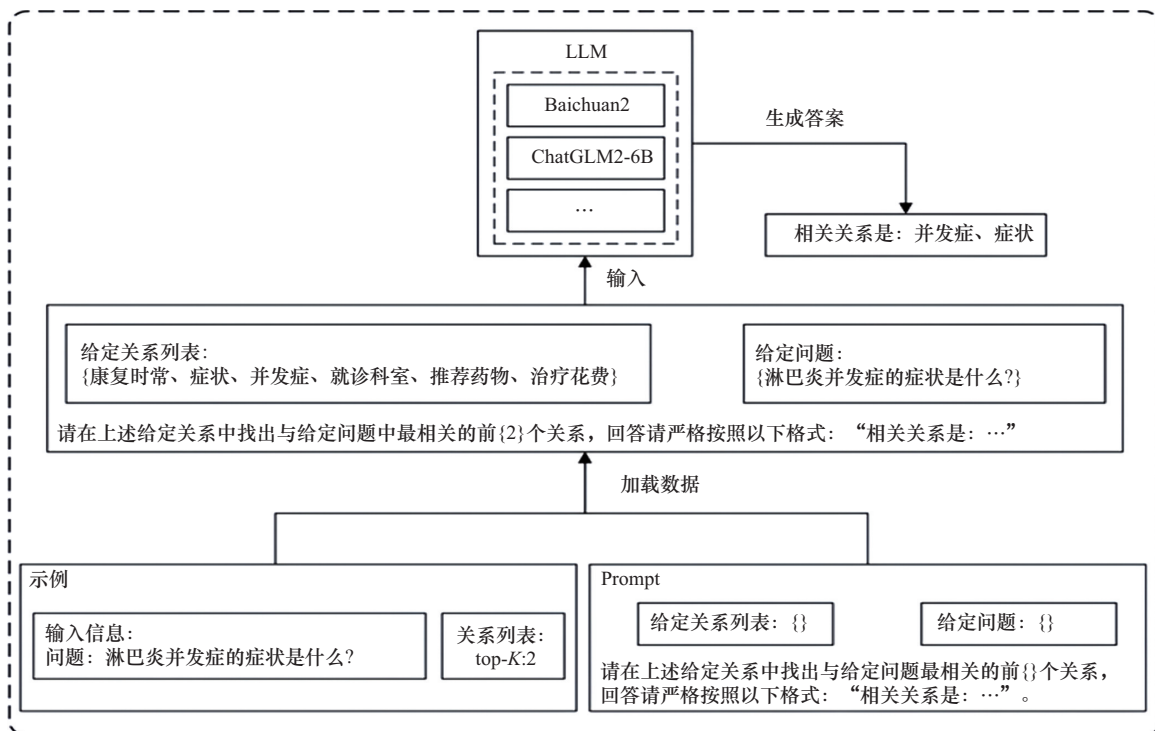


图2 基于 Prompt 的关系筛选示例



节提出的基于 Prompt 的关系筛选策略经历了以下几个关键步骤。

(1) 根据关系筛选任务设计特定的 Prompt 模板。在这个模板中，需要明确表达任务要求，以及明确定义输出格式，以便从关系列表中筛选出与问题最相关的前 K 个关系。

(2) 加载数据。将 Prompt 模板中需要的信息填充进去，随后输入预训练的 LLM 进行处理，这一过程有助于优化模型学习和适应特定任务的能力。

(3) 得到 LLM 输出的预测结果。通过传递具体信息和指令给 LLM，获得最终的答案文本。基于 Prompt 的关系筛选策略无须额外的标记数据或模型参数更新，本质上是一种零样本学习方法，通过有效地利用已有信息，使得模型在未见过的问题或领域中也取得良好的性能。

近年来，基于 Prompt 的关系筛选策略得到了进一步发展，新的研究成果为该领域带来了新的技术和应用。例如，Huang 等^[5]提出了一种基于 Prompt 学习的无监督关系抽取模型 (PURE)，该模型通过设计上下文感知的 Prompt 模板，将无监督关系抽取任务转换为掩码预测任务，并结合链接预测模块，从文本中有效提取关系。这一方法通过预测关系三元组中的缺失实体，显著提升了模型性能。此外，AdaPrompt 方法^[13]将关系抽取任务重新定义为掩码语言建模，并引入自适应标签选择机制，通过可变数量的标签令牌来处理复杂的多标签空间，取得了在少样本和监督设置下的优异表现。在零样本关系抽取方面，2023 年的一项专利提出了一种基于 Prompt 多模板融合的零样本关系抽取方法，旨在解决在缺乏标注数据的情况下如何从文本中有效提取关系^[14]。通过这些创新，基于 Prompt 的关系筛选策略变得更加高效和适应不同的应用场景。

基于 Prompt 的关系筛选策略能够获取与问题相关的 top- K 个关系，辅助生成主题实体和 top- K 个

关系相关的所有三元组。信息检索模块通过使用 LLM 选取与问题最相关的关系，为后续任务提供更有针对性的信息，使得 LLMKG 在处理复杂问题时更加灵活和精准。

基于 Prompt 的关系筛选策略的筛选过程可以被公式化为：给定问题 q 和最终生成的答案 a 之间，存在推理路径 $p_z = e_h \xrightarrow{r_1} e_1 \xrightarrow{r_2} \dots \xrightarrow{r_l} a$ ，那么 $z = \{r_1, r_2, \dots, r_l\}$ 即与 q 相关的所有关系集合。为了得到与 q 相关的 top- K 个关系，可以通过最小化 p_z 的后验分布的 KL 散度 (Kullback-Leibler divergence) 来实现^[15]，后验分布 Q_z 的表达式如式 (2) 所示。这表示路径 z 是否在图 G 中是合法的，且与问题 q 所蕴含的关系匹配。

$$Q_z \cong Q(z|a, q, G) = \begin{cases} 1, & q_z(e_h, a) \in G \\ 0, & \text{其他} \end{cases} \quad (2)$$

为了得到有效的关系集合，本文只考虑 e_h 和 a 之间的最短关系路径，使用 KL 散度 $D_{\text{KL}}(P(z|q)||Q(z))$ 衡量两种分布的差异。KL 散度 D_{KL} 的计算式如式 (3) 所示：

$$D_{\text{KL}}(P(z|q)||Q(z)) = \sum_{z \in Q(z|a, q, G)} P(z|a) \log \left(\frac{P(z|a)}{Q(z|a, q, G)} \right) \quad (3)$$

因为 $Q(z|a, q, G)$ 是 one-hot 形式，所以最终退化为：

$$= - \sum_{z \in Q(z|a, q, G)} \log P(z|a)$$

其中， $P(z|a)$ 等价于 LLM 对路径 z 生成答案 a 的概率。通过最小化 D_{KL} ，可最大化 LLM 生成“忠实于知识图谱关系路径”的概率，确保推理结果与知识图谱结构一致。

4 基于简单连接的知识库生成

LLMKG 是一种零样本问答方法。零样本问答的特殊之处在于测试集包含了训练集中不存在的类别，因此在零样本学习中需要引入额外的信

息。然而，这种方法目前仍然存在一些挑战。首先，LLM 依赖于参数中的知识，容易导致生成与事实相悖的答案，因为 LLM 中的知识可能不准确，并且随着时间的变化，存在过时的可能性。其次，LLMKG 需要不断纠正错误并且增加新知识。然而，通过额外的参数更新（即微调）来调整 LLMKG 内部知识是一项非常耗时且繁琐的任务。此外，现有方法在知识更新时缺乏高效的增量学习机制，使得整个系统在面对新知识时的适应性较低^[16]。最后，LLM 在生成问题答案时所用到的知识来源不够清晰，这限制了它们对输出结果的可解释性，并影响用户对答案的信任度。

为了解决上述问题，本文提出了基于简单连接的知识库生成方法。该方法的核心思想是将结构化知识图谱中的信息转化为语义一致的自然语言文本，并将其嵌入问题中。这样一来，LLM 在回答问题时不再仅依赖于内部参数中的知识，而是通过从外部知识图谱中额外获取和注入知识来提升回答的专业性。这种注入的知识包含了有助于回答问题的准确和最新的事实，从而有效地改善了 LLM 在垂直领域问题上的表现。

知识图谱是一种结构化的语义知识库，用于存储实体之间的关系和属性，并以图形的方式表示各种实体（如人、地点、事物）及其相互之间的多种联系。知识图谱通常被定义为 $G = \{(eh, r, et) \mid eh, et \in \mathbf{E}, r \in \mathbf{R}\}$ ，其中 \mathbf{E} 表示实体集合， \mathbf{R} 表示关系集合。三元组 (eh, r, et) 表示关系 r 链接了头实体 e_h 和尾实体 e_t 。例如，现有的关系事实为“淋巴炎的并发症是颈部淋巴转移癌”可以表示为两个实体 $e_h = \text{“淋巴炎”}$ ， $e_t = \text{“颈部淋巴转移癌”}$ 以及关系 $r = \text{“并发症”}$ 组成的一个三元组（淋巴炎，并发症，颈部淋巴转移癌），将该三元组作为外部知识嵌入问题回答模块的 Prompt 模板中，以提示 LLM 生成基于知识图谱中事实知识的正确答案。然而，LLM 仅处理自然语言文本作为输入内容，而不直接处理结构化数据。因此，

在将结构化知识图谱嵌入 LLM 之前，需要先将其转换为 LLM 可以理解的自然语言文本。

基于简单连接的知识库生成方法首先定义了一个 Prompt 模板 p ，它的作用是引导 LLM 将结构化的三元组转化为自由格式的文本。 p 中的内容为：“给定一个三元组： $\{ \}$ ，你需要使用简单的连接词将它转换为一个句子”。本文将正确的输出序列定义为 $y = [y_1, y_2, \dots, y_T]$ ，模型输出定义为 $o = [o_1, o_2, \dots, o_u]$ 。LLM 预测的概率分布为 $D'_{o,t}$ ，在第 t 步时模型输出是基于 p 的并与前 $t-1$ 步的正确输出序列 y_1, y_2, \dots, y_{t-1} 有关， $D'_{o,t}$ 和 d'_{o_c} 的表达式如式 (4) 和式 (5) 所示^[17]：

$$D'_{o,t} = [d'_{o_1}, d'_{o_2}, \dots, d'_{o_u}] \quad (4)$$

$$d'_{o_c} = P(o_c | p, y_1, y_2, \dots, y_{t-1}), c = 1, 2, \dots, u \quad (5)$$

d'_{o_c} 表示在给定 p 和 y_1, y_2, \dots, y_{t-1} 的情况下生成 o_c 的概率。如果 LLM 预测的下一个 token o_c 与 y_t 相等，则预测概率为 1，不相等则为 0，事实概率分布 $D_{o,t}$ 和 $d_{o,t}$ 的表达式如式 (6) 和式 (7) 所示：

$$D_{o,t} = [d_{o_1}, d_{o_2}, \dots, d_{o_u}] \quad (6)$$

$$d_{o_c} = \begin{cases} 1, & o_c = y_t \\ 0, & o_c \neq y_t \end{cases}, c = 1, 2, \dots, u \quad (7)$$

知识库生成方法的主要目标在于将三元组转化为其相应的文本字符串，即进行语言化。本文采用了一种简单的连接方法，即通过 LLM 将头实体、关系和尾实体组合成自然语言文本。根据第 6.7 节中的实验结果可知，相较于采用复杂的方法进行从图形到文本的转换，本文提出的方法在 LLM 提示问答中表现更为出色。例如，上述示例三元组（淋巴炎，并发症，颈部淋巴转移癌）在经过简单连接后被表示为“淋巴炎的并发症是颈部淋巴转移癌”。经由该方法生成的相关知识将被整合到知识库中。对于多跳问题，如果需要进行迭代式的知识检索，之前加载到知识库



中的文本信息将持续存在，直到LLM判断知识库已经包含足够信息来回答当前给定的问题。这种机制有助于提高LLMKG在处理复杂问题时的效率和准确性，同时确保了知识库中的信息始终保持与问题的要求相匹配。

5 基于知识判断的迭代检索

为了提高LLM在回答问题时的专业性，本文在模型回答问题时用Prompt的形式将垂直领域知识注入LLM中，扩大了LLM所能解决的问题范围。这种方法的独特之处在于，它能够让LLM在特定专业领域内表现出色，而无须再次训练。相较于引入专业数据并进行微调，这种提示注入的方式更加高效，同时确保了模型对专业性问题的应对能力。因此，上述方法不仅能够提升模型在垂直领域问答中的表现，还具备了更快速、灵活的部署优势。

本文提出的知识判断策略的重要性在于它解决了多跳问题所面临的挑战，即需要检索多个知识点问题。在进行最终的问题回答之前，知识判断策略首先通过预设的Prompt模板，将当前问题(q)和已构建的知识库(B)提交给大语言模型(LLM)，要求LLM评估现有信息是否足以解答给定的问题。该评估的核心标准是知识覆盖的完整性(即知识库需包含问题涉及的所有核心实体并能形成从主题实体到答案实体的完整、逻辑连贯的关系路径)和语义相关性(确保知识片段与问题高度相关)。通过这种方式，可以在实际回答问题之前对知识库的信息覆盖程度进行全面评估，确保模型能够有效处理多跳问题，从而避免信息缺失或不完整而导致错误答案的情况发生。这一策略的实施不仅能够显著提高回答准确性，同时也加强了系统鲁棒性，使其更好地适应并解决各种复杂问题的挑战。

知识判断的过程如下：假设给定问题 q 已经检索出3个与当前问题相关的三元组，并经过语

言化处理，将它们纳入与当前问题相关的知识库 B 中。该模块会先基于知识库判断当前已知信息是否足以回答问题。在获得LLM回答“是”的情况下，立即执行知识注入步骤，将知识库嵌入问题回答Prompt。借助LLM在语义理解方面的优势，得出最终答案。此外，上述方法是一种零样本学习方法，与引入专业相关数据对LLM进行微调的方式相比，具有明显的优势，因为它无须重新训练模型就能够快速部署一个在专业领域具备强大语言处理能力的模型。问题回答模块根据 B 中的知识推理出问题答案，其推理出正确答案的概率如式(8)所示：

$$P_{\theta}(a|q, G) = \sum_{p_z \in B} P_{\theta}(a|q, B, G) P_{\theta}(p_z|q) \quad (8)$$

其中， $P_{\theta}(a|q, B, G)$ 表示基于知识库 B 回答出正确答案 a 的概率， $P_{\theta}(p_z|q)$ 表示给定问题 q 生成推理路径 p_z 的概率， θ 表示LLM的参数。

Prompt设计及问题回答模块示例如图3所示，该示例首先将问题和知识库中的内容嵌入知识库判断Prompt中，让LLM判断给定的知识是否足以回答该问题，在模型回复“是”以后，再把问题和知识库中的内容嵌入问题回答Prompt中，利用LLM推理出问题答案。

本文方法所依赖的Prompt结构具备良好的可扩展性与通用性，能够轻松迁移至其他垂直领域问答任务中，实现跨任务、跨知识库的快速部署。这也体现了Prompt注入策略在当前大模型应用范式中的独特优势，即无须重新训练模型便可实现领域能力的迁移与扩展。

LLM的知识注入策略在学术研究和实际应用中展现出显著潜力。Cadeddu等^[18]系统研究了LLM在科学文章分类任务中的知识注入方法，比较了4种主流策略(知识增强编码、知识嵌入对齐、实体链接注入与知识提示生成)的效果及适用场景。研究结果表明，这些注入方法能显著提升模型的领域适应性与专业理解能力。该研究还

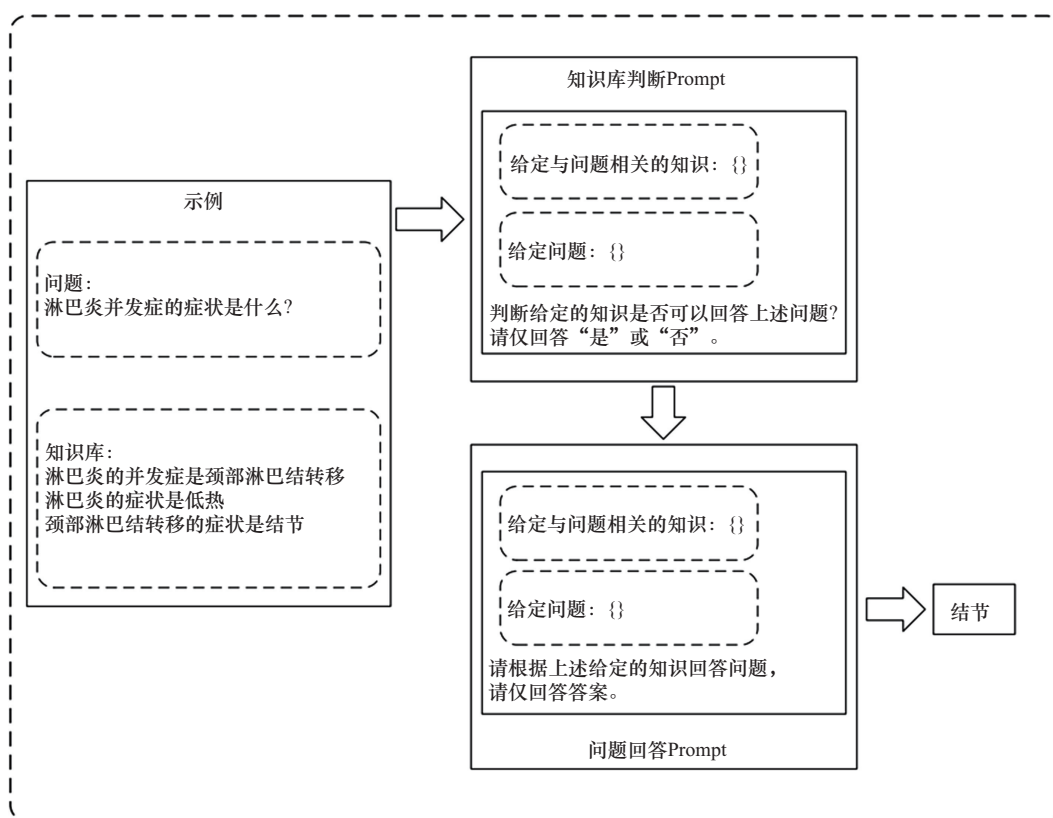


图3 Prompt 设计及问题回答模块示例

构建了涵盖丰富三元组的专业知识图谱及评估体系，为后续知识增强型 LLM 提供了方法论支持和实验参考。

6 实验结果及分析

本节针对融合大语言模型与知识图谱的多跳问答方法进行了实验设计和结果分析，以验证该方法的有效性。

6.1 数据集

为了验证 LLMKG 的有效性，本文使用了由达观数据与同济大学合作发布的新冠病毒知识图谱多跳问答数据集 COKG-DATA^[19]进行实验。COKG-DATA 是基于 OpenKG 发布的知识图谱 OpenKG-COVID19 的一个大型中文多跳问答数据集。该数据集包含一个知识图谱和与 COVID-19 七方面（即百科、防控、物资、诊疗、健康、流行病、英雄）相关的单跳问题和多跳问题。其中，

COKG-DATA 的知识图谱中包含了 112 246 个实体、209 个关系和 787 056 个三元组。

COKG-DATA 数据集的统计数据量信息见表 1。

表 1 COKG-DATA 数据集的统计数据量信息

数据集	训练集	验证集	测试集
COKG-DATA 1hop	165 795	55 239	55 239
COKG-DATA 2hop	48 147	16 049	16 049
COKG-DATA 3hop	2 811	927	927

本节实验所采用的 LLM 都在 Python3.8 和 PyTorch2.0.1 环境下基于 NVIDIA GeForce RTX 4 090 硬件运行，显存为 24 GB。

6.2 评估指标

为了评估模型回答复杂问题的能力，本文使用 Hits@1 作为检验模型效果的评估指标。Hits@1 表示在所有候选答案中，问题正确答案在候选答案中排名第一的比例。即如果模型预测问句的



top-1 答案存在于标注答案中，则该问句的 Hits@1 为 1，高 Hits@1 值说明模型具有较高的预测性能。这种评估指标很受欢迎，也得到了研究者的认可，并在最近的许多 KGQA 工作中使用^[20-21]。

在验证基于知识图谱嵌入的排名框架的有效性时，还用到了 Hits@10 指标，它表示在所有候选答案中，问题正确答案在候选答案中排名前 10 的比例。

6.3 大语言模型

LLMKG 方法中的各个模块都需要与 LLM 进行问答交互：基于 Prompt 的关系筛选策略、基于简单连接的知识库生成方法、知识判断策略和最终进行的问题回答。为了验证 LLMKG 框架在不同 LLM 上的性能，在 COKG-DATA 上设计了实验。本节使用了 11 种中文开源 LLM：Moss-moon-003-sft-int4^[22]、Chinese-Llama2-7B^[23]、ChatGLM-6B^[24]、ChatGLM2-6B、Baichuan-7B、Baichuan2-7B^[25]、Qwen1.5-110B^[26]、Qwen1.5-14B-MoE^[26]、Qwen3-8B^[27]、DeepSeek-VL-7B-base^[28]、LLama-3-8B^[29]。LLMKG 框架实验中使用的中文开源 LLM 概览见表 2。

LLMKG 方法的性能受所使用基座 LLM 的影响，使用不同的 LLM 可能影响 LLMKG 方法的效果。在 LLMKG 方法的性能评估中，Qwen1.5-110B 与 Qwen1.5-14B-MoE 未被选为测试基准模型，主要基于以下两方面考量。

(1) 资源效率瓶颈：Qwen1.5-110B 需 4×A100-80G GPU 进行 FP16 推理^[30]，其千亿级参数的微调与推理成本远超常规实验条件^[31]，而 Qwen1.5-14B-MoE 虽激活参数仅 14 B，但 140 B 总参量导致显存占用显著高于同规模稠密模型^[32]，二者均不符合大规模重复实验的资源约束。

(2) 任务适配缺陷：LLMKG 要求精准的知识编辑与稳定的实时响应，但 Qwen1.5-110B 的超大参数量导致知识注入可控性低且推理延迟过高 (>5 s)^[33]；Qwen1.5-14B-MoE 的专家路由机制则引入输出波动性，可能干扰知识对齐结果的复现性^[34]。

因此，在 LLMKG 这类强依赖可重复实验、低成本迭代、高知识可控性的任务中，选用其他 6 种模型进行实验以平衡计算可行性、知识编辑

表 2 LLMKG 框架实验中使用的中文开源 LLM 概览

模型名称	参数规模	特点与描述
Moss-moon-003-sft-int4	—	预训练数据包括约 7 000 亿中英文及代码单词，具备对话指令微调、插件增强学习和人类偏好训练能力，支持多轮对话和多插件使用
Chinese-LLaMA2-7B	70 亿	中文版 LLaMA2 模型，全部开源，完全可商用
ChatGLM-6B	62 亿	支持中英双语的对话语言模型，基于 General Language Model 架构
ChatGLM2-6B	62 亿	ChatGLM 的第二代版本，基座模型升级，整体性能显著提升
Baichuan-7B	70 亿	基于 Transformer 结构，在约 1.2 万亿 tokens 上训练，支持中英双语，上下文窗口长度为 4 096
Baichuan2-7B	70 亿	在数学、代码、安全、逻辑推理和义理解等方面均较前代有显著提升
Qwen1.5-110B	1100 亿	对话能力增强，支持多语言文本与代码处理，稳定支持最长 32 K 上下文，适合长文档分析与长对话任务
Qwen1.5-14B-MoE	140 亿	动态路由，总参数量为 14 B，但通过专家子网络动态路由，实际激活参数量仅 2.7 B，平衡性能与计算成本
Qwen3-8B	80 亿	阿里最新推出的 80 亿参数开源大模型，在多语言能力、代码、数学等方面表现出色，并使用了 GQA 等技术来提升效率，可免费商用
DeepSeek-VL-7B-base	70 亿	使用 SigLIP-L 和 SAM-B 作为支持 1 024×1 024 图像输入的混合视觉编码器，并基于在约 2 T 文本 token 语料上训练的 DeepSeek-LLM-7b-base 构建
LLama-3-8B	80 亿	Meta 推出的新一代开源模型，基于超过 15 T tokens 的高质量公开数据训练，在推理、代码生成和指令遵循方面表现优异

效率与实验严谨性需求。

不同 LLM 性能对比见表 3，展示了 LLMKG 在不同 LLM 下的性能对比结果，最好的分数用粗体表示，1hop、2hop 和 3hop 分别表示 COKG-DATA 中的 1 跳、2 跳和 3 跳数据集。

表 3 不同 LLM 性能对比

LLM	1hop	2hop	3hop
Moss-moon-003-sft-int4	88.6	83.4	79.1
Chinese-llma2-7B	96.4	93.1	90.3
ChatGLM-6B	95.6	92.7	91.3
ChatGLM2-6B	98.7	94.6	93.9
Baichuan-7B	95.3	93.7	92.7
Baichuan2-7B	99.2	97.6	97.4
Qwen3-8B	99.7	98.3	98.5
DeepSeek-VL-7B-base	98.8	97.2	96.6
LLama-3-8B	99.4	97.9	97.6

在表 3 中，可以观察到 LLMKG 方法在不同 LLM 模型上的性能表现。由于它们在训练时所用数据不同，且参数及参数量存在差异，不同 LLM 模型可能具有不同的特征提取和表示能力。因此，它们在相同任务上也可能产生不同效果。通过对比不同 LLM 下的性能指标，发现在 LLMKG 方法中，Qwen3-8B 模型展现出了最佳效果，且基于其推理复杂问题的能力，在 3hop 的表现上也很出色。在接下来的实验中，将使用基于 Qwen3-8B 的 LLMKG 方法与其他基线方法进行对比。

6.4 对比实验

本节在 COKG-DATA 数据集上设计实验，将 LLMKG 在与最先进的 KGQA 模型进行对比，融合 LLM 与知识图谱的多跳问答模型与基线模型对比见表 4，最好的分数用粗体表示。

(1) EmbedKGQA: 首次将知识图谱嵌入方法用于执行多跳 KGQA 任务，并且通过执行缺失链接预测来减少知识图谱的稀疏性，增强模型在稀疏知识图谱上回答问题的能力。

(2) TransferNet: 是一个可解释性较强的多

表 4 融合 LLM 与知识图谱的多跳问答模型与基线模型对比

模型	1hop	2hop	3hop	平均值
EmbedKGQA ^[35]	72.9	80.3	86.6	79.9
TransferNet ^[36]	98.8	96.5	13.3	69.5
COKG-QA ^[19]	95.8	92.9	97.3	95.3
LLMKG	99.7	98.3	98.5	98.8

跳问答模型，支持标签和文本关系，通过关系分数传输实体分数从而推断答案。

(3) COKG-QA: 通过增加嵌入投影和模式感知模块来扩展 EmbedKGQA，采用嵌入投影机制，将不同空间的问题和知识嵌入的重要特征融合到一个共同的语义空间中。

分析表 4 中的结果可知，LLMKG 在 KGQA 任务中相较于其他基线模型表现较好，特别是在解决多跳问题方面表现出色，相较于最好的基线模型 COKG-QA、LLMKG 的平均 Hits@1 值提升了 3.5%。LLMKG 方法通过检索知识图谱中的有用信息作为 LLM 的输入，使 LLM 更好地理解问题并获取与问题相关的知识，进而生成更准确的答案。这种知识增强的方法有效地减少了 LLM 生成不准确或错误答案，以及产生幻觉的情况。借助知识图谱中的结构化信息，LLMKG 方法能够更精准地进行推理和回答复杂问题，尤其是那些涉及多跳推理的问题。

6.5 消融实验

为了验证本文提出基于 Prompt 的关系筛选策略、基于简单连接的知识库生成方法和知识判断策略的对模型效果的影响，在此处进行消融，融合 LLM 与知识图谱的多跳问答模型消融实验结果如图 4 所示。下面是对实验设置的解释。

(1) NK (no knowledge): 表示仅使用 LLM 生成答案，而不使用知识图谱中的信息进行知识增强。在这种情况下，LLM 只依靠输入问题进行答案生成，没有额外的知识注入。

(2) RRF (remove relation filtering): 表示去

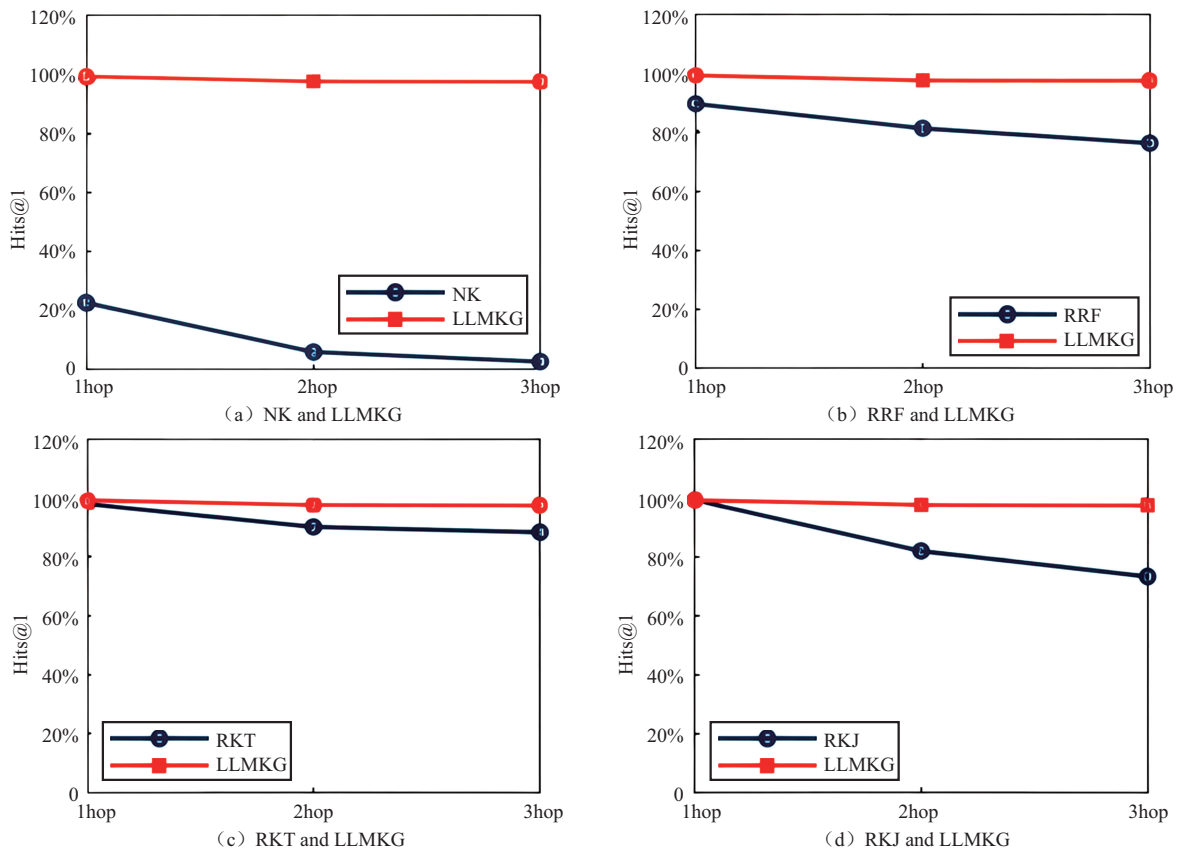


图4 融合LLM与知识图谱的多跳问答模型消融实验结果

除了关系筛选策略，不再提取top-K个相关关系，而是将在知识图谱中检索到的所有关系都视为相关关系进行后续的知识生成。这种方法的问题在于，知识图谱中的关系通常包含大量无关甚至冗余的信息，不进行筛选会导致LLM处理的知识过于复杂，从而影响问答准确性。王得强等^[37]提出了一种结合知识图谱的行业知识库构建方法，该方法主要涉及关系抽取、知识融合和信息筛选等技术，以确保知识图谱的质量和有效性。这与消融实验中RRF的问题类似，即如何从庞大的知识库中筛选出最有用的信息，从而提升问答系统的精度。

(3) RKT (remove KG2 text): 表示去除了基于简单连接的知识库生成方法，不再对检索到的三元组进行语言化处理，而是直接将结构化信息传递给LLM。这一修改对1hop(单跳)问题

影响较小，但在2hop(双跳)及以上的多跳问题中，直接提供结构化知识会降低模型的理解能力。Ma等^[38]总结了强化学习在知识图谱构建和推理中的应用，尤其是在关系抽取、知识推理和结构化数据转换方面的研究进展。这表明强化学习等方法可用于优化知识图谱数据的传递方式，从而提高LLM的理解能力，而不是直接将未经处理的结构化数据输入LLM。

(4) RKJ (remove knowledge judgment): 表示去除了知识判断策略，将仅根据主题实体检索三元组改为检索与问题相关的所有三元组，将检索到的信息注入LLM中进行问答。由于多跳问答任务通常需要正确选择和组合多个实体及其关系，去除知识判断策略后，模型可能检索到无关或错误的关系，导致回答错误。Pan等^[4]探讨了将LLM与知识图谱相结合的方法，包括知识图

谱到文本转换 (KG-to-text)、实体关系提取 (entity-relation extraction) 和多跳推理等关键技术。研究表明, 利用知识判断策略 (如主题实体的层级筛选) 能够有效提高多跳问答的准确性, 从而减少误检信息的影响。

分析图4中的结果可知, 仅使用LLM进行问答时, 准确度相对较低, 几乎无法正确回答问题。这表明LLM中内化的知识不足以生成事实答案, 使用知识图谱结合LLM从而提升问答模型的性能是非常有必要的。首先, 只有提高信息检索模块在LLMKG中的表现, 才能更好地执行知识增强任务, 即通过引入外部知识来丰富知识库。因此需要进一步测量信息检索模块的性能。实验表明, 在去除关系筛选策略的情况下, 将检索到的所有与问题相关的信息传递给LLM, 这种方式无法过滤掉与问题无关的信息, LLM在回答问题时需要理解和分析大量知识, 导致模型准确回答问题的能力下降, 特别是在问题跳数增加的情况下影响更为显著。其次, 去除基于简单连接的知识库生成方法对1跳问题的影响相对较小, 主要因为LLM本身可以理解一些简单的结构化知识。然而, 随着问题跳数的增加, LLM难以一次性理解大量结构化知识, 成为问答准确度降低的主要原因。最后, 去除知识判断策略对多跳问答的影响最为显著, 因为多跳问答通常需要链接多个实体和关系, 如果不根据主题实体迭代检索相关信息, 可能引入大量无关的关系并导致关键知识的遗漏。综合分析上述实验结果可知, 本文提出的基于Prompt的关系筛选策略、基于简单连接的知识库生成方法和知识判断策略均是有效的。

6.6 信息检索模块性能分析

为了评估信息检索模块的性能, 在本节中分析了信息检索模块检索到的top-K个关系与答案覆盖率和检索时间之间的关系。检索准确率性能分析如图5所示, 答案覆盖率是指检索到的top-K个

相关三元组中正确推理路径的覆盖率。分析实验结果可知, 检索到的top-K个相关三元组随着K值的增大而增加, 但是答案覆盖率却在降低。这是因为随着K值增大, 引入了更多与问题无关的知识。

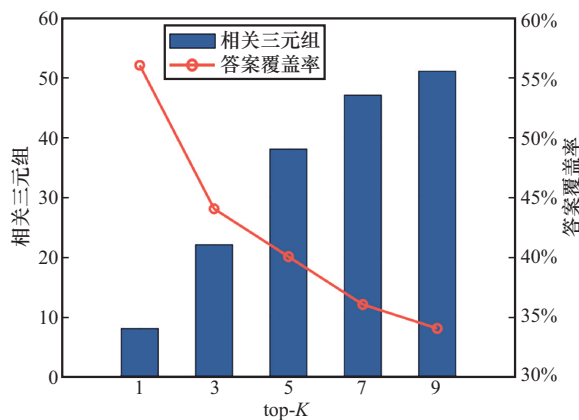


图5 检索准确率性能分析

此外, 检索效率性能分析如图6所示, 信息检索模块的检索时间随着K值的增加而增加。综合检索时间与答案覆盖率考虑, 文本将top-K的值设置为3, 即保证了信息检索模块的结果包含了正确的推理路径, 也不会引入过多的噪声, 并且合理平衡了检索时间。

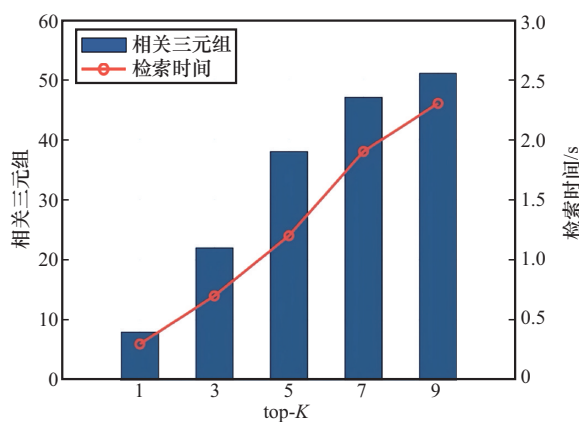


图6 检索效率性能分析

6.7 知识库生成方法的有效性

本文提出的基于简单连接的知识库生成方法通过连接词组合将结构化三元组转化为自然语言



文本，从而实现了语言化的过程。与现有使用神经网络的方法不同，该方法采用了简单的语言化技术，但效果却非常好。在本节中对比了该方法与现有的知识图谱语言化技术之间的性能差异。

本节采用了 Ribeiro 等^[39]提出的结合局部知识与全局知识的文本生成方法，将结构化知识转换为自然语言文本，并将该模型用 LLMKG-CGE 表示。为了验证基于简单连接的知识库生成方法的有效性，本节对 LLMKG-CGE 与仅使用 LLM 生成答案的 LLMKG-NK 和采用简单连接的 LLMKG 进行对比，验证知识库生成方法有效性的实验结果见表 5。

表 5 验证知识库生成方法有效性的实验结果

模型	1hop	2hop	3hop	平均值
LLMKG-NK	22.5	5.8	2.6	10.3
LLMKG-CGE	97.8	91.4	88.7	92.6
LLMKG	99.7	98.3	98.5	98.8

通过分析表 5 中的结果，可以得出以下结论。在回答 1 跳问题时，将结构化知识转化为自然语言文本与直接使用简单连接对结构化文本进行编码在问题回答效果上差别不大。然而，在回答 2hop 和 3hop 问题时，基于简单连接的知识库生成方法展现出了其优势。由于采用将知识图谱转换为自然语言文本的预训练模型存在固有误差，可能生成与原始三元组在语义上不同的错误输出。如果生成的自然语言文本不准确，将直接影响下游的问答任务效果。与此不同，简单连接方法并不改变三元组的实际含义，因此在某种程度上能够减小由于文本生成误差引起的负面影响，并且可以提高转换的效率与准确性。

本文方法的性能受多个关键因素的影响。其中，知识图谱的实体覆盖率和事实准确性决定了推理路径的可用性；Prompt 模板的结构清晰性与引导能力影响大模型对外部知识的响应效率；路径构建与子图裁剪策略决定了输入内容的信息密

度与相关性。上述因素在多跳推理场景中相互耦合，对模型性能具有显著影响。因此，未来工作可进一步从 Prompt 自动生成、路径优化、图谱补全与知识可信度控制等角度提升整体问答系统的效果与稳定性。

6.8 时序与多模态知识增强的未来研究展望

在真实世界的问答任务中，信息复杂性和语境多样性日益凸显，传统依赖文本或结构化三元组构建的静态知识图谱在语义覆盖和推理能力方面已难以满足各类应用需求。特别是当问题涉及时间约束或异构模态时，仅靠静态图谱往往存在表达受限、更新滞后、信息割裂等问题，影响了问答系统的准确性与上下文适应性。

首先，针对具有时效性需求的问题，例如“当前美国总统是谁”“某项政策于何时生效”等，模型不仅需识别实体与关系间的静态语义，还应具备对事实时序演化的理解能力。为此，近年来提出的时间性知识图谱（temporal knowledge graph, TKG）通过引入时间戳、事件序列与历史状态建模等机制，显著提升了模型对动态事实的识别与区分能力。在未来工作中，可继续拓展时间性感知能力。例如，TwiRGCN 通过在图卷积网络中引入时间权重机制和门控结构，实现了在多跳时序问答（如 TimeQuestions 数据集）中精度提升 9~10 个百分点^[40]。另一重要工作是 Chen 等^[41]提出的 Multi-granularity TKGQA 方法，它创新性处理了多粒度时间约束下的问题，并建立了 CronQuestions 等高复杂度时间问答数据集。这些成果证明，在时间问答场景下引入显式时间机制能够显著提升模型对动态事实的推理能力与准确性。

其次，在医疗诊断、电商推荐、社交媒体分析等多模态应用场景中，文本信息常常难以独立表达完整语义，而图像、音频、视频等感知层数据中蕴含着大量关键线索。为了应对这一挑战，多模态知识图谱（multimodal knowledge graph，

MMKG) 被提出, 作为一种将图像、文本、音频等异构模态与结构化知识联结的统一语义框架, 不仅继承了传统知识图谱的组合性与可推理性, 还显著增强了语义表达能力和上下文感知能力^[42-43]。Lee 等^[44]提出的 MR-MKG、He 和 Wang^[45]提出的 Multimodal Graph Transformer 等方法, 已在多模态问答与视觉推理任务中展现出优异性能, 验证了 MMKG 在增强模型语义理解与推理能力方面的潜力。然而, 现有方法大多依赖于全图输入, 缺乏针对问题的结构裁剪与上下文适配。

未来, 本文提出的“任务驱动的动态子图构建机制”可为上述两类方向提供可扩展的统一框架: 在用户提出问题时, 根据其语义信息裁剪与之强相关的时间敏感子图或多模态子图, 并以结构化形式注入大语言模型, 实现知识的高效调度与语义增强。该策略不仅能够提升问答系统对时序逻辑与感知信息的响应能力, 也有望推动系统从“感知驱动”向“知识驱动”转变, 进一步拓展其在政务、金融、医学、教育等复杂场景中的应用边界。

7 结束语

本文提出的 LLMKG 方法通过将结构化知识图谱以 Prompt 形式注入大语言模型, 有效提升了模型在多跳问答任务中的准确性与可解释性。该方法具备良好的零样本适应能力, 能够在无须微调的前提下适配多种主流 LLM。引入的时间性知识增强策略与多模态融合构想, 为复杂语义和动态场景下的问答系统提供了潜在路径。然而, 当前方法仍存在 Prompt 构造依赖人工、图谱注入、计算开销较大等问题。未来可进一步探索自动化 Prompt 生成、任务驱动的动态子图构建及多模态知识融合等方向, 以构建更高效、灵活且具知识可控性的问答系统。

参考文献:

- [1] CHEN R R, JIANG W F, QIN C W, et al. LLM-based multi-hop question answering with knowledge graph integration in evolving environments[C]//Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024. Stroudsburg: ACL, 2024: 14438-14451.
- [2] TAN X Y, WANG X Y, LIU Q, et al. Paths-over-graph: knowledge graph empowered large language model reasoning[C]//Proceedings of the ACM on Web Conference 2025. New York: ACM Press, 2025: 3505-3522.
- [3] ZHOUT, CHENYB, LIUK, et al. CogMG: collaborative augmentation between large language model and knowledge graph[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). Stroudsburg: ACL, 2024: 365-373.
- [4] PAN S R, LUO L H, WANG Y F, et al. Unifying large language models and knowledge graphs: a roadmap[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7): 3580-3599.
- [5] HUANG M L, DUAN L, ZHANG Y H, et al. Prompt learning based unsupervised relation extraction model[J]. Journal of Computer Applications, 2023, 43(7): 2010-2016.
- [6] LINDERS J, TOMCZAK J M. Knowledge graph-extended retrieval augmented generation for question answering[EB]. 2025.
- [7] LU Y F, ZHOU Y G, LI J, et al. Knowledge editing with dynamic knowledge graphs for multi-hop question answering[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2025, 39(23): 24741-24749.
- [8] XU Y, HE S Z, CHEN J B, et al. Generate-on-graph: treat LLM as both agent and KG for incomplete knowledge graph question answering[C]//Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: ACL, 2024: 18410-18430.
- [9] YANG Z K, ZHU Z X, ZHU J. CuriousLLM: elevating multi-document question answering with LLM-enhanced knowledge graph reasoning[C]//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track). Stroudsburg: ACL, 2025: 274-286.
- [10] LI Y D, SONG D D, ZHOU C Z, et al. A framework of knowledge graph-enhanced large language model based on question decomposition and atomic retrieval[C]//Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024. Stroudsburg: ACL, 2024: 11472-11485.
- [11] SHU D, CHEN T, JIN M, et al. Knowledge graph large language model (KG-LLM) for link prediction[C]//Proceedings of the 16th Asian Conference on Machine Learning Research Ox-



- ford: Pergamon Press, 2025: 143-158.
- [12] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing[J]. *ACM Computing Surveys*, 2023, 55(9): 1-35.
- [13] CHEN X, ZHANG N Y, XIE X, et al. KnowPrompt: knowledge-aware Prompt-tuning with synergistic optimization for relation extraction[C]//*Proceedings of the ACM Web Conference 2022*. New York: ACM Press, 2022: 2778-2788.
- [14] 张春, 张宁, 许亮. 一种基于 Prompt 多模板融合的零样本关系抽取方法: CN115687609A[P]. 2023.
- ZHANG C, ZHANG N, XU L. A zero-shot relation extraction method based on Prompt multi-template fusion: CN115687609A[P]. 2023.
- [15] LUO L, LI Y F, HAFFARI G, et al. Reasoning on graphs: faithful and interpretable large language model reasoning[C]//*Proceedings of the Twelfth International Conference on Learning Representations (ICLR 2024)*. Vienna: ICLR, 2024.
- [16] ZHAO H Y, CHEN H J, YANG F, et al. Explainability for large language models: a survey[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(2): 1-38.
- [17] WU Y, HU N, QI G, et al. Retrieve-rewrite-answer: a KG-to-text enhanced LLM framework for knowledge graph question answering[C]//*Proceedings of the 12th International Joint Conference on Knowledge Graphs (IJCKG' 23)*. Tokyo: IJCKG, 2023: 10.
- [18] CADEDDU A, CHESSA A, DE LEO V, et al. A comparative analysis of knowledge injection strategies for large language models in the scholarly domain[J]. *Engineering Applications of Artificial Intelligence*, 2024, 133: 108166.
- [19] DU H F, LE Z W, WANG H F, et al. COKG-QA: multi-hop question answering over COVID-19 KnowledgeGraphs[J]. *Data Intelligence*, 2022, 4(3): 471-492.
- [20] WANG J B, LEI J, SUN S N, et al. Embeddings based on relation-specific constraints for open world knowledge graph completion[J]. *Applied Intelligence*, 2023, 53(12): 16192-16204.
- [21] CHEN Z Y, ZHAO X, LIAO J Z, et al. Temporal knowledge graph question answering via subgraph reasoning[J]. *Knowledge-Based Systems*, 2022, 251: 109134.
- [22] ZHENG R, DOU S H, GAO S Y, et al. Secrets of RLHF in large language models part I: PPO[EB]. 2023.
- [23] CUI Y M, YANG Z Q, YAO X, et al. Efficient and effective text encoding for Chinese LLaMA and alpaca[EB]. 2023.
- [24] DU Z X, QIAN Y J, LIU X, et al. GLM: general language model pretraining with autoregressive blank infilling[C]//*Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: ACL, 2022: 320-335.
- [25] YANG A Y, XIAO B, WANG B N, et al. Baichuan 2: open large-scale language models[EB]. 2023.
- [26] BAI J Z, BAI S, CHU Y F, et al. Qwen technical report[EB]. 2023.
- [27] YANG A, LI A F, YANG B S, et al. Qwen3 technical report[EB]. 2025.
- [28] LU H Y, LIU W, ZHANG B, et al. DeepSeek-VL: towards real-world vision-language understanding[EB]. 2024.
- [29] DUBEY A, JAUHRI A, PANDEY A, et al. The llama 3 herd of models[EB]. 2024.
- [30] QWEN TEAM. Qwen1.5 technical report[R]. 2024.
- [31] ZHANG Y, LI H, WANG X, et al. Efficiency analysis of large Chinese LLM[EB]. 2024.
- [32] RAJBHANDARI S, LI C L, YAO Z W, et al. DeepSpeed-MoE: advancing mixture-of-experts inference and training to power next-generation AI scale[EB]. 2022.
- [33] BAI C, ZHOU Y, ZHANG Z, et al. Long context: beyond 100 K tokens[C]//*Proceedings of the 32th Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Stroudsburg: ACL, 2023: 4980-4995.
- [34] ZHOU Y Q, LEI T, LIU H X, et al. Mixture-of-experts with expert choice routing[C]//*Proceedings of the 36th Conference on Neural Information Processing Systems*. West Chester: Curran Associates, Inc., 2022.
- [35] 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*. Stroudsburg: ACL, 2020: 4498-4507.
- [36] SHI J X, CAO S L, HOU L, et al. TransferNet: an effective and transparent framework for multi-hop question answering over relation graph[C]//*Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: ACL, 2021: 4149-4158.
- [37] WANG D Q, WU J, GUAN L W. Industry knowledge base construction based on knowledge graph[J]. *Manufacturing Technology & Machine Tool* (8): 74-80.
- [38] MA A, YU Y H, YANG S L, et al. Survey of knowledge graph based on reinforcement learning[J]. *Computer Research and Development*, 2022, 59(8).
- [39] RIBEIRO L F R, ZHANG Y, GARDENT C, et al. Modeling global and local node contexts for text generation from knowl-

edge graphs[J]. Transactions of the Association for Computational Linguistics, 2020, 8: 589-604.

- [40] SHARMA A, SAXENA A, GUPTA C, et al. TwiRGCN: temporally weighted graph convolution for question answering over temporal knowledge graphs[C]//Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 2023: 2049-2060.
- [41] CHEN Z Y, LIAO J Z, ZHAO X. Multi-granularity temporal question answering over knowledge graphs[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2023: 11378-11392.
- [42] ZHU X R, LI Z X, WANG X D, et al. Multi-modal knowledge graph construction and application: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(2): 715-735.
- [43] KEJRIWAL M, KNOBLOCK C A, SZEKELY P. Knowledge graphs: fundamentals, techniques, and applications[M]. Cambridge: The MIT Press, 2021.
- [44] LEE J L, WANG Y Q, LI J, et al. Multimodal reasoning with multimodal knowledge graph[C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg: ACL, 2024: 10767-10782.
- [45] HE X H, WANG X. Multimodal graph transformer for multimodal question answering[C]//Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 2023: 189-200.

[作者简介]



蒋献 (1988-), 男, 浙江工商大学统计与数据科学学院博士生, 主要研究方向为智慧教育与智慧网络。



王涵亦 (2004-), 女, 浙江工商大学在读, 主要研究方向为通信网络和自然语言处理。



杨诗婷 (2004-), 女, 浙江工商大学在读, 主要研究方向为智能信息处理与多模态知识融合。



陈星妤 (1999-), 女, 浙江工商大学硕士生, 主要研究方向为智慧教育和自然语言处理。



周梦瑶 (2002-), 女, 浙江工商大学硕士生, 主要研究方向为智慧教育和智能体。



董黎刚 (1973-), 男, 博士, 浙江工商大学教授, 主要研究方向为智慧教育与智慧网络。