BEKO:大语言模型与知识图谱的双向增强

吴信东 黄满宗 卜晨阳

(合肥工业大学大数据知识工程教育部重点实验室 合肥 230009)

摘 要 以 ChatGPT 为代表的大型语言模型(LLMs)在多种任务中展现了巨大潜力。然而,LLMs 仍然面临幻觉现象和长尾知识遗忘等问题。为了解决这些问题,现有方法通过结合知识图谱等外部知识显著增强 LLMs 的生成能力,从而提升回答的准确性和完整性。但是,这些方法存在如知识图谱构建复杂、语义丢失以及知识单向流动等问题。为此,我们提出了一种双向增强框架,不仅利用知识图谱增强 LLMs 的生成效果,而且利用 LLMs 的推理结果补充知识图谱,从而形成知识的双向流动,并最终形成知识图谱与 LLMs 之间的循环正反馈,不断优化系统效果。此外,通过设计增强知识图谱(Enhanced Knowledge Graph,EKG),我们将关系抽取任务延迟到检索阶段,降低知识图谱的构建成本,并利用向量检索技术缓解语义丢失问题。基于此框架,本文构建了双向增强系统——BEKO(Bidirectional Enhancement with a Knowledge Ocean)系统,并在关系推理应用中相比传统方法取得明显的性能提升,验证了双向增强框架的可行性和有效性。BEKO 系统目前已经部署在公开的网站——ko. zhonghuapu. com。

关键词 知识图谱;大语言模型;检索增强生成;关系推理;知识问答

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2025.01572

BEKO: Bidirectional Enhancement with a Knowledge Ocean for LLMs and KGs

WU Xin-Dong HUANG Man-Zong BU Chen-Yang

(Key Laboratory of Knowledge Engineering with Big Data (the Ministry of Education of China), Hefei University of Technology, Hefei 230009)

Abstract Large Language Models (LLMs), epitomized by advanced systems such as ChatGPT, have exhibited remarkable capabilities across a diverse range of applications, including but not limited to knowledge-based question answering, factual verification, and creative content generation. The sophisticated contextual understanding and generative prowess of these models have significantly expanded the horizons of artificial intelligence applications. However, LLMs are not without limitations; they are prone to issues such as hallucination and the oversight of long-tail knowledge, which undermine their ability to produce precise and exhaustive content. Specifically, these models may generate outputs that deviate from factual accuracy or struggle to address intricate queries that require multi-step reasoning or the integration of cross-domain knowledge. To mitigate these shortcomings, current methodologies have sought to augment the generative efficiency of LLMs by integrating external knowledge repositories, such as knowledge graphs, thereby enhancing the veracity and comprehensiveness of the outputs. Despite these advancements, the existing approaches are hampered by challenges including the intricate nature of knowledge graph construction, potential semantic degradation, and the unidirectional nature of knowledge flow, which collectively constrain their efficacy and scalability in real-world scenarios. In response to these challenges, this study

收稿日期:2024-07-27;在线发布日期:2025-03-18。本课题得到国家自然科学基金(62120106008)、安徽省科技攻关项目(202423k09020015)以及安徽省科协青年科技人才托举计划(RCTJ202420)资助。**吴信东**,博士,教授,主要研究领域为数据挖掘、大数据分析、知识工程。E-mail: xwu@hfut.edu.cn。**黄满宗**,博士研究生,主要研究方向为知识图谱推理、数据挖掘。**卜晨阳**,博士,副教授,目前研究兴趣为知识驱动的智能优化。

introduces a bidirectional enhancement framework that not only utilizes knowledge graphs to bolster the generative performance of LLMs but also employs the reasoning outcomes of LLMs to enrich the knowledge graphs, thereby establishing a reciprocal knowledge exchange. This framework fosters a synergistic relationship between knowledge graphs and LLMs, culminating in a perpetual optimization of system performance. Central to this framework is the introduction of an Enhanced Knowledge Graph (EKG), a novel knowledge representation methodology that simplifies the construction process by performing only named entity recognition initially, thereby preserving the connections between entities and their source documents and postponing relation extraction to the reasoning phase. This strategy markedly reduces the complexity and resource expenditure associated with knowledge graph construction. Moreover, by maintaining the contextual linkages of entities, the EKG facilitates access to the original text during the reasoning process, thereby minimizing semantic loss and ensuring the fidelity and accuracy of the information, which is particularly vital when processing data from diverse and heterogeneous sources. Furthermore, the study delineates a bidirectional enhancement mechanism: the EKG augments the reasoning capabilities of LLMs by providing both structured and unstructured contextual information, thereby significantly enhancing LLM performance in complex relational reasoning tasks such as those involving multi-entity relationships or cross-domain knowledge integration, where the EKG can supply extensive background information. Conversely, the reasoning outputs of LLMs are utilized to update and enrich the knowledge graph, thereby establishing a bidirectional knowledge flow that continually refines and enhances system performance. This reciprocal enhancement not only optimizes the individual components but also fortifies their collective efficacy. Implementing this bidirectional enhancement framework, the study has developed BEKO (Bidirectional Enhancement with a Knowledge Ocean), a practical system that has been successfully deployed on the public platform ko. zhonghuapu. com, demonstrating commendable outcomes in real-world relational reasoning applications. Empirical evaluations reveal that BEKO outperforms traditional methodologies in complex relational reasoning tasks, with a notable 4.9% improvement in $F1$ score over the baseline method GraphRAG, thereby substantiating the superiority of BEKO and affirming the viability and effectiveness of the bidirectional enhancement strategy.

Keywords　knowledge graph; large language models; retrieval-augmented generation; relation reasoning; question answering

1　引　言

近年来，以 ChatGPT[1] 为代表的大语言模型(Large Language Models，LLMs)在知识问答[2]、事实验证[3]以及艺术创作[4]等任务中展现了巨大的潜力。然而，尽管 LLMs 取得了显著进展，但在生成准确且完整的内容方面仍存在挑战。例如，大语言模型在生成过程中可能会出现“幻觉”现象[5-6]，即生成与事实不符的内容。因此，如何解决 LLMs 生成存在的幻觉以及知识遗忘等问题引起广泛关注。

引入知识图谱(Knowledge Graph，KG)[7]等外部知识，可显著提升 LLMs 生成内容的准确性和完整性[8-9]。LLMs 具备强大的上下文学习能力[10]，通过在模型输入中添加与问题相关的上下文信息，可有效减轻生成过程中的错误，并提高其对新知识的理解和响应能力[9,11]。而知识图谱因其结构化、准确性以及知识的不断演进等特点[8]，被广泛视为理想的外部知识库。利用知识图谱提供准确的显式知识来增强大模型生成，可以显著提升生成结果的质量[12-13]。尤其是在处理多跳问题或需要深度推理的任务时，知识图谱作为外部知识库表现出优于目前广泛使用的基于向量匹配实现检索的文档-向量知识库[14]的效果。如图 1(a)所示，推理“吴信东”与“郑磊”的关系需要结合多个文档中的信息。然而，基于文档-向量知识库的检索增强大模型(Naive RAG)

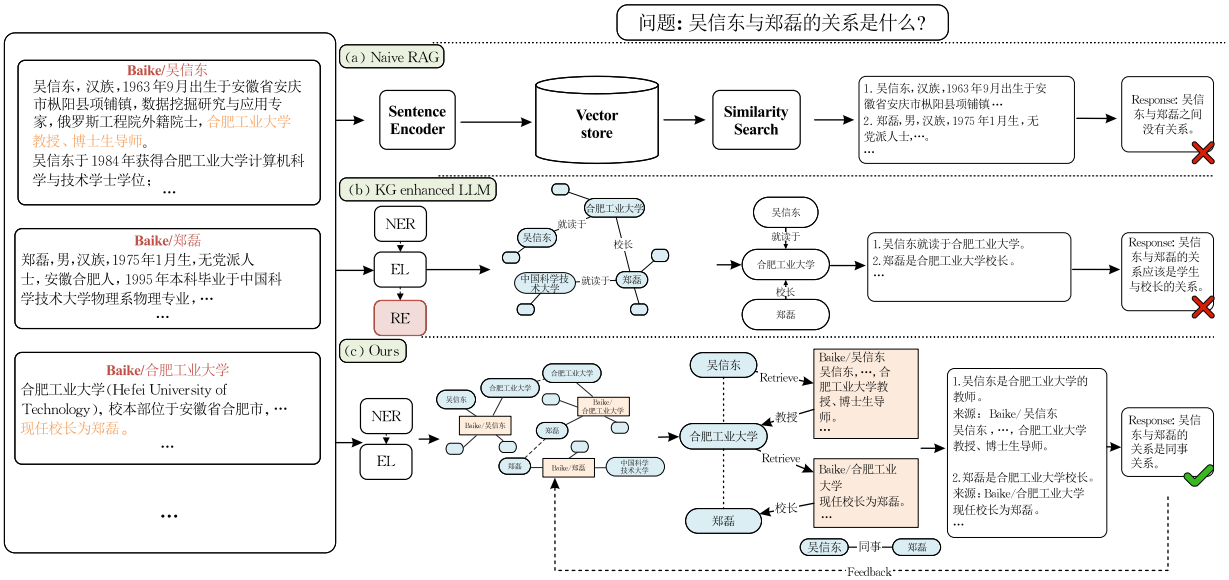


图 1 三种不同知识增强 LLMs 的方式((a) Naive RAG，利用向量匹配实现检索的文档-向量知识库作为外部知识库增强大模型方法；(b) KG enhanced LLM，利用知识图谱作为外部知识库单向增强大模型方法；(c) Ours，我们提出的双向增强方法)

方法[9,15-17]通常只能检索出与问题显性相关的信息，这些信息不足以准确推理出实体间的关系。而如图 1(b)所示，基于知识图谱作为外部知识库的方法则可以有效地检索出与问题相关的结构化信息。

然而，现有利用知识图谱增强大模型方法的成功在很大程度上取决于知识图谱的完整性和高质量。如图 1(b)所示，目前利用知识图谱增强大模型的方法遵循着一个研究范式[12-13,18]：首先将语料库构建一个完善的知识图谱，在推理时从知识图谱中检索相关信息，相应地增强问题上下文，从而增强 LLMs。这些方法仍然存在以下挑战：

(1) 知识图谱构建困难：构建知识图谱需要在精确定义的模式下，通过多个自然语言处理子任务的协作完成[19-20]，包括命名实体识别[21]、实体链接[22]和关系抽取[23-24]等。这一过程不仅消耗巨大，还极其复杂，往往难以达到预期效果。此外，由于受到数据来源多样性和信息更新频繁性的影响，导致图谱中的信息存在不完整性问题[19]。因此，在系统构建初期，很难立即构建出完善的知识图谱。

(2) 构建知识图谱过程中存在语义丢失：在知识图谱构建过程中，将非结构化文本转化为结构化知识时，不可避免地会出现语义流失[23,25]。如图 1(b)所示，在知识图谱构建中，将“吴信东”与“合肥工业大学”的关系抽取为“就读于”，而忽略了就读时间。这使得在利用知识图谱中“吴信东”至“郑磊”的图谱信息增强 LLMs 进行推理时，由于缺失了准确的就读时间，导致了其错误推理“吴信东”与“郑磊”的关系。语义流失不仅会导致信息的不完整，还可能引起误解和错误推理，进而影响基于知识图谱的推理和决策过程。

(3) 知识单向流动：现有方法主要聚焦于利用知识图谱中的知识来增强大语言模型(LLMs)的推理能力[12-14]，或是利用大模型构建知识图谱[26]。然而，这些方法都表现为知识的单向流动，即知识图谱增强 LLMs，或 LLMs 增强知识图谱，而未能实现双向互补。实际上，知识图谱不仅可以增强大模型的推理能力，同时在推理过程中也能挖掘出有价值的信息，反过来补充到知识图谱中。现有的方法并未充分利用这种双向潜力，忽视了 LLMs 和知识图谱之间的互补性。

为了应对上述挑战，我们提出了一种双向增强框架——BEKO(Bidirectional Enhancement with a Knowledge Ocean)。首先，我们设计了一种新的知识表示方式——增强知识图谱(Enhanced Knowledge Graph，EKG)。如图 1(c)所示，从原始文本构建 EKG 时，仅需进行命名实体识别，保留实体与原始文档的链接，并将关系抽取推迟到推理阶段。这一方式大幅降低了知识图谱的构建成本与复杂性，缓解了知识图谱构建困难的问题。同时，因保留了实体与上下文的关联，EKG 在推理过程中可溯源至原始文本，减少了在构建知识图谱过程中的语义丢失。此外，我们设计了一个双向增强机制：一方面，

利用 EKG 中的知识增强大语言模型(LLMs)的推理能力；另一方面，将 LLMs 的推理结果反馈至知识图谱中，补充丰富知识，实现了知识在知识图谱与大模型之间的双向流动。这一循环正反馈机制不断优化系统效果。基于该双向增强框架，我们构建了实际系统 BEKO，并已部署于公开网站，在实际关系推理应用中取得了良好效果。实验结果表明，与传统方法相比，BEKO 在处理复杂关系推理任务时表现出显著的性能提升，验证了双向增强方法的可行性和有效性。

综上所述，本文的贡献如下：

(1) 分析了现有方法的局限性，提出了知识图谱与大语言模型(LLMs)之间的双向增强框架，实现了知识的循环流动。

(2) 设计了增强知识图谱(EKG)，并基于此从知识库中检索出结构化信息和非结构化信息，显著增强了大模型的推理能力。

(3) 构建了一个实际应用系统——BEKO。实验结果表明，BEKO 系统在处理复杂关系推理任务时，性能显著优于传统方法，验证了我们提出方法的可行性和有效性。

2　BEKO 系统实现双向增强

BEKO 系统依托 Knowledge Ocean(KO)平台(ko.zhonghuapu.com)，实现知识图谱与大模型双向增强。KO 知识库目前共收录了 494 367 394 个实体节点和 2 317 974 658 个关系，且数据规模仍在持续增长。在利用 KO 知识图谱增强大模型推理实体关系时，如图 2 所示，我们不仅从 KO 知识库中检索相关图谱结构知识，还利用知识图谱的路径信息引导证据检索过程，逐步收集与问题相关的文本信息。随后，利用这些信息增强大模型的推理，并将最终的推理结果反馈到知识图谱中，从而形成一个知识图谱与大模型双向流动的循环，不断优化整个系统。

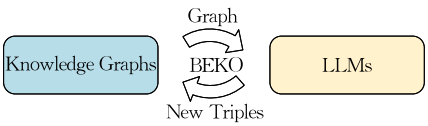


图 2　双向增强框架

2.1　KO 知识图谱

KO 知识库中包含了海量的知识图谱信息以及文本信息。依托于互联网公开的数据，我们构建了 KO 知识海洋知识库，汇集了多个来源的信息。这些信息包括文档信息，如维基百科[①]、百度百科[②]等，以及结构化信息，如 Wikidata[③] 中的数据和华谱网[④]的家谱信息等。对于文档信息，我们基于第 4 节中提出的增强知识图谱构建方法，将其抽取为增强知识图谱；而结构化信息由于其本身就具备三元组的形式，因此可以直接添加到 KO 知识库中。通过整合这些多样化的来源，KO 知识海洋知识库能够提供相当全面和丰富的知识图谱，支持 BEKO 系统的高效运行。

2.2　问答场景示例

BEKO 系统支持多种实体关系查询，包括人物关系以及人物与组织之间的关系。

2.2.1　实体关系

如图 3 所示，用户可以在输入框内输入任意两个实体，系统将从 KO 知识图谱中检索出多条从起始实体到目标实体的路径，这些路径隐含了实体间的可能关联。如果实体不存在于 KO 知识库中，系统会实时从网络中进行检索，并添加到 KO 知识库中，以确保数据的完整性和及时性。这种动态扩展能力使得 KO 系统能够不断更新和完善其知识库。

图 3　前端页面展示

用户选择头尾实体后，系统在右侧窗口显示基于向量检索的 Naive RAG 方法的推理结果。然而，通过 Naive RAG 方法推理得到的实体关系并不总是准确的，例如，通过图 3 中的检索出的路径，我们可以看到，“吴信东”和“郑磊”都受雇佣于“合肥工业大学”，他们之间具有明显的“同事”关系。但是基于 Naive RAG 实现的关系推理方法无法准确推理出他们之间的关系。这是因为这类方法依赖于文本向量实现检索增强大模型，然而基于向量匹配的检索方式在准确性和完整性上表现不佳。在面对复杂的

① zh. wikipedia. org
② baike. baidu. com
③ wikidata. org
④ zhonghuapu. com

实体关系时，可能会遗漏关键的信息从而法准确推理出实体间的关系。

2.2.2 知识图谱增强大模型

为了提高推理结果的准确性和完整性，我们采用第 4 节中设计的 BEKO 双向增强方法。如图 4 所示，用户可以选择需要进行推理的路径，随后点击“增强”，利用图谱路径信息增强大模型推理实体关系。我们利用路径信息作为证据检索的线索，引导证据检索过程，从而从知识库中检索出有助于推理实体关系的图谱信息以及文本信息，并最终利用这些信息增强 LLMs 推理出实体关系。相比于图 3 中 Naive RAG 方法推理得到实体关系，BEKO 双向增强方法准确推理出了“吴信东”和“郑磊”之间具有“同事关系”。此外，通过挖掘语料库中的潜在文本信息，我们的方法还额外推理出“吴信东”和“郑磊”之间具有的“同乡”关系。因此，基于 BEKO 双向增强方法得到的推理结果在准确性和完整性上更为优秀。这证明了我们能够提供更加可靠和全面的关系推理结果，满足用户在复杂关系推理任务中的需求。

图 4 使用知识图谱增强大模型推理实体关系

2.2.3 大模型结果反馈增强知识图谱

如图 4 所示，在使用知识图谱增强大模型推理“吴信东”和“郑磊”之间关系时，BEKO 双向增强方法挖掘出潜在的“同乡”关系。LLMs 推理结果为知识图谱提供有价值的补充信息，不断丰富知识图谱内容。如图 5 所示，推理出“吴信东”和“郑磊”之间具有的多重关系的结果后，点击“增强”按钮，系统会解析出“吴信东”和“郑磊”之间的“同乡”以及“同事”关系，以及他们与“安徽省”共同的关联，并将这些关联信息作为补充内容添加到知识图谱中。这种反馈机制可以不断丰富知识图谱内容，并优化后续推理效果，不断提升知识图谱的完整性，确保其在处理复杂关系推理任务时表现更出色。

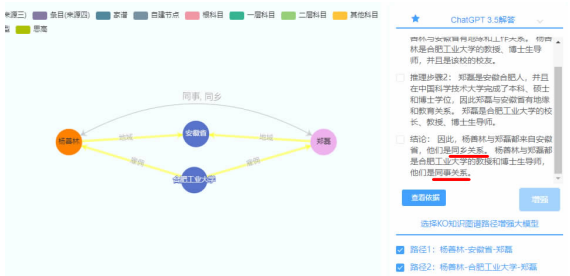
图 5 反馈增强知识图谱

2.2.4 多轮双向增强

随着上述双向增强过程的持续，知识图谱内容不断丰富，进而提升了知识图谱增强大模型的推理效果。如图 5 所示，当将大模型的推理结果反馈到知识图谱中后，图谱的内容得到了补充。这不仅增加了知识图谱的丰富度，又提升了后续利用知识图谱增强大模型的推理效果。多轮增强在这一过程中起到了关键的作用，通过多轮次的反馈和更新，知识图谱能够持续吸收新的信息，形成更为全面和细致的知识网络。

如图 6 所示，在使用“吴信东”与“郑磊”的推理结果增强知识图谱后，新的三元组信息“郑磊-地域-安徽”被添加到知识图谱中。这一新增信息不仅丰富了“杨善林”与“郑磊”之间的推理路径，还提升了关系推理的效果。在多轮增强前，“杨善林”与“郑磊”之间仅能推导出具有的“同事”关系，而在使用“吴信东”与“郑磊”推理结果增强知识图谱后，“杨善林”

(a)



(b)

图 6 多轮增强结果((a) 多轮增强前的“杨善林”与“郑磊”推理结果；(b) 在使用“吴信东”与“郑磊”推理结果增强知识图谱后，“杨善林”与“郑磊”的推理结果)

与“郑磊”的推理结果不仅包含了“同事”关系，还额外推理出了“同乡”关系。

相比于单次增强，多轮增强能够更好地捕捉复杂的关系和隐含的信息，从而不断丰富知识图谱内容，使得推理结果更加多样和精确。通过不断的迭代和优化，知识图谱能够动态调整和完善自身结构，从而在每一轮增强中获得更高的知识密度和推理能力。这种持续的增强过程不仅提升了知识图谱的整体质量，也为大模型提供了更加丰富和可靠的推理基础，使得实体之间的推理结果变得更加准确和全面。知识图谱和大模型知识的双向流动，使得系统整体能够持续优化。

3　相关工作

3.1　知识图谱增强大模型

知识图谱因其结构化、准确性以及知识的不断演进等特点，被广泛视为理想的外部知识库，许多研究致力于利用知识图谱增强大模型的能力。现有的研究中主要通过将来自 KGs 的相关结构化知识转换为 LLM 的文本提示，从而提升 LLM 的能力。例如，Li 等人[27]提出利用 LLMs 生成 SPARQL 查询的骨架，并通过知识图谱填充完整信息；Baek 等人[28]则从问题中抽取包含实体的三元组，供 LLMs 进行推理；Li 等人[29]的方法使用 LLMs 将问题分解为多个子问题，并生成相应的 SPARQL 查询以从知识图谱中检索知识；Wang 等人[30]提出了一个检索-阅读-验证的问答系统，与 LLMs 交互获取外部知识。此外，Edge 等人[26]提出的 GraphRAG 方法利用 LLMs 从非结构化文本中提取结构化数据，构建带标签的知识图谱，以支持数据集问题生成、摘要问答等多种应用场景，并通过图机器学习算法进行语义聚合和层次化分析，能够回答高层次的抽象或总结性问题。

然而，这些方法的成功在很大程度上依赖于知识图谱的完整性和高质量。目前的知识图谱构建方法在处理真实世界中不完整和动态变化的特性方面是不足够的。此外，在知识图谱构建过程中，将非结构化文本转化为结构化知识时，不可避免地会出现语义流失[23,25]。以上问题限制了现有方法在实际应用中的效果。我们提出的双向增强框架基于 EKG，不仅从图谱中检索出结构化的三元组信息，还通过对关系进行溯源，获取相关的上下文信息，缓解了知识图谱语义丢失的问题。同时，我们设计了双向增强机制，让 LLMs 不仅从 EKG 中获取知识增强推理能力，还将推理结果反馈至知识图谱中，在推理过程中动态补充新的知识至知识图谱。BEKO 框架实现了知识图谱与 LLMs 之间的互补和协同，克服了传统方法中知识单向流动的限制，不断优化系统性能。

3.2　大模型增强知识图谱

最近的研究致力于整合大型语言模型（LLMs）以增强知识图谱（KG），通过结合文本信息来提升下游任务的性能。知识图谱（KGs）存储结构化知识，在许多实际应用中发挥着重要作用[20]。然而，现有的知识图谱方法在处理不完整的知识图谱和文本语料库以构建知识图谱方面存在不足[31]。鉴于大型语言模型（LLMs）的通用性，许多研究正在尝试利用 LLMs 的强大能力来解决与知识图谱相关的任务。研究人员利用 LLMs 处理知识图谱中的文本语料库，并使用生成的文本表示来丰富知识图谱的表示[32-33]。此外，一些研究还使用 LLMs 处理原始语料，从中提取关系和实体，以构建知识图谱[34-35]。

最近的研究尝试通过设计知识图谱提示，有效地将结构化的知识图谱转换为 LLMs 可以理解的格式。通过这种方式，LLMs 可以直接应用于诸如知识图谱推理[36-37]等任务，从而进一步提升其性能和应用范围。然而，知识图谱难以构建，现有方法未能有效地对未见过的实体与关系进行建模并表示新事实。在 BEKO 框架中，我们设计了一种知识表示方式——增强知识图谱（Enhanced Knowledge Graph，EKG）。从原始文本构建 EKG 时，仅需进行命名实体识别，保留实体与原始文档的链接，并将关系抽取推迟到推理阶段，降低了知识图谱的构建成本与复杂性，缓解了知识图谱构建困难的问题。此外，通过双向增强机制，我们在推理过程中动态补全知识图谱，从而不断完善知识图谱。

3.3　协同知识图谱与大模型

LLMs 和知识图谱是两种本质上互补的技术，它们之间是可以相互促进的。LLMs 和知识图谱的协同作用也越来越受到研究者的关注。现有研究中，协同知识图谱以及大模型的方法主要有两类。第一类是协同知识表示。在知识图谱（KGs）和文本语料库中都包含了大量的知识，但前者是显式且结构化的，而后者通常是隐式且非结构化的。为了有效地表示这两种知识，研究人员提出了协同知识表示模型，通过引入额外的 KG 融合模块，与大型语言

模型(LLMs)联合训练[38-40]。第二类是协同推理。为了更好地利用文本语料库和知识图谱推理中的知识,协同推理旨在设计一种可以有效地使用 LLMs 和 KG 进行推理的协同模型。有研究通过对 LLMs 进行重新训练,以优化文本和知识图谱之间的交互[41-42]。也有研究将 LLMs 作为 agent,在知识图谱上进行推理[14,43]。

然而,这些方法要么需要重新训练大语言模型,但这一过程成本高昂;要么只是利用大模型的能力在知识图谱上进行推理,而未能实现知识在知识图谱与大模型之间的双向流动和双向增强。因此,在真正实现大模型与知识图谱之间的相互增强方面仍然存在不足。

4 双向增强

如图 7 所示,我们提出的双向增强框架分为 4 个部分:(1)增强知识图谱(EKG)构建;(2)知识检索;(3)知识增强 LLMs 推理;(4)反馈增强 EKG。首先,我们通过实体抽取以及实体链接构建初始的 EKG。随后,通过基于 EKG 的检索方法提高知识检索的准确性和完整性。之后,我们利用检索出的知识增强大模型,提高其生成的准确性和完整性。最后,将大模型推理结果作为新的知识反馈增强 EKG,形成知识图谱↔大模型间的循环反馈过程,持续优化系统的推理效果。

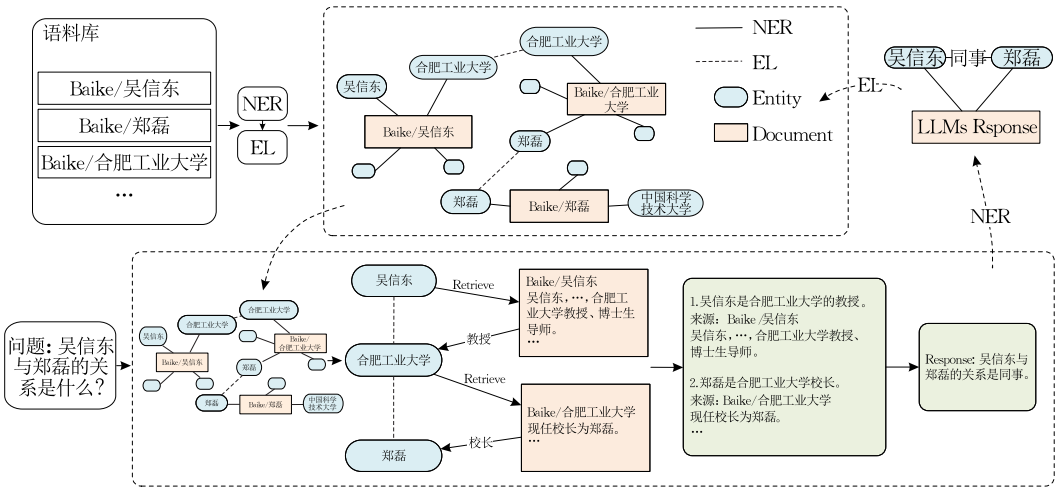


图 7 双向增强框架

4.1 增强知识图谱(EKG)及其构建

传统的知识图谱构建方法成本高昂且容易导致语义丢失,同时难以满足下游任务对动态、个性化信息的需求。如 Cohen 等人[44]所描述的,传统知识图谱通常采用"信息驱动"的方式,需在构建初期完整抽取和整合信息源。例如,现有的知识图谱构建方法需要在精确定义的模式下,通过多个自然语言处理子任务的协作完成[19-20],包括命名实体识别[21]、实体链接[22]和关系抽取[23-24]等。这一过程不仅困难,而且需要大量的时间和高昂的成本[19-20]。更为重要的是,在将非结构化信息转化为结构化信息的过程中,不可避免地会丢失部分语义信息,从而影响下游任务的性能。

此外,当前的许多自然语言处理任务,如跨文档关系抽取[45-46]、开放域知识问答[47]以及事实抽取和验证[48-49]等,其核心挑战已经从根据已有事实知识推理出最终答案转变为从庞大的知识库中检索出关键事实知识。因此,知识图谱在这些任务中的应用方式也应随之变化。我们提出的增强知识图谱(EKG)表示方式,旨在改善传统知识图谱在构建成本以及信息检索中的表现。

4.1.1 增强知识图谱(EKG)

EKG 是一种高度灵活的知识表示框架,它基于以下几个核心概念和组成部分:

节点(Node):

(1)文档节点:存储具体的文档内容。

(2)实体节点:代表图谱中的实体,如人物、地点、组织、概念等。

边(Edge):

(1)文档节点与实体节点的边:这种链接表示某个实体来源于特定的文档。

(2)实体节点与实体节点之间的边:这种边表示实体之间的具体关系。

在 EKG 中,文档节点作为文档中实体关系的

聚合，使得 EKG 无需关系抽取即可构建，并在构建初期就能获得节点间的潜在关联，在使用过程中逐渐完善和补全图谱。通过文档节点与实体节点的链接，可以实现对实体来源的快速溯源，从而缓解了传统知识图谱中语义丢失的问题。此外，EKG 在构建初期无需进行关系抽取，显著降低了初期构建的复杂性和成本，并且在使用过程中可以逐步完善和补全图谱中的关系。

4.1.2　EKG 构建

对于语料库中的文档，我们首先通过命名实体识别（NER）技术提取文档中的实体，并利用实体链接（EL）技术将不同文档中提及的相同实体关联起来。对于每个文档 d_i，我们对其进行实体抽取：

$$E_i=\mathrm{NER}(d_i),$$

其中，$\mathrm{NER}(d_i)$ 表示对文档 d_i 进行命名实体识别，提取出文档中的实体集合 E_i。实体抽取方法我们基于 LLMs 实现，具体的提示词见附录。

接下来，通过实体链接（EL）技术将这些实体进行关联，构建增强知识图谱（EKG）：

$$\mathrm{EKG}=\mathrm{EL}(\bigcup_i E_i),$$

其中，$\mathrm{EL}(\cdot)$ 表示对所有文档提取的实体集合进行实体链接，构建出增强知识图谱 EKG。

以上的构建过程将不同的文档通过共同具有的实体将他们联系起来，从而构建了一个文档以及实体相互关联的图谱。在 EKG 中，实体之间的关系潜在地蕴含在与他们通过关联的文档之中，而文档之间的关联则通过共有的实体实现。基于 EKG，我们可以有效地检索出能推理出实体之间关系的文本信息。

4.2　知识检索

在关系推理中，我们需要有效的方法来检索与实体对 (e_h,e_t) 相关的信息。为了实现这一目标，我们设计了一种从增强知识图谱（EKG）中获取结构化路径信息和非结构化文本信息的检索方法，这些信息将作为后续推理步骤的基础。

现有的基于向量的检索方法通常只能获取直接关联的文本信息。然而，当实体对之间的关系需要通过多个中间实体进行多跳推理时，这些方法难以检索到所有必要的推理证据。因此，我们需要利用图谱中的路径信息来辅助推断实体间的关系。此外，在关系抽取过程中，描述实体间关系的文本可能会被简化为更一般的关系短语，这可能导致语义丢失。例如，将“表兄”这样的具体关系抽取为“亲戚”这一更广泛的术语，可能会导致具体语义的丢失。因此，我们不仅希望获取实体间的关系路径，还希望获取推理实体间关系的原始文本，以确保语义的完整和准确。

为了克服这些挑战，我们利用关系路径上的三元组信息，从相关联的文档中抽取出与三元组相关的文本信息。通过结合知识图谱中的推理路径和原始文本，我们能够更全面地理解和推断头尾实体之间的复杂关系。

4.2.1　路径检索与筛选

我们在 EKG 上使用广度优先搜索（BFS）来检索从实体对 e_h 到 e_t 的路径。具体来说，从 e_h 出发，利用 BFS 算法找出多条路径，每条路径的形式为

$$P_i=\{e_1,d_1,e_2,d_2,\cdots,e_{m-1},d_{m-1},e_m\},$$

其中，P_i 表示从 e_h 到 e_t 的第 i 条路径，包含实体节点 e 和文档节点 d。为了控制算法的复杂度，我们限定每条路径上的实体节点数量 $m<4$。

为了进一步筛选路径，我们对所有路径进行排序。受到 TF-IDF 的启发，我们设计了 EF-IDF（Entity Frequency-Inverse Document Frequency）公式来评估路径中的实体重要性：

$$\mathrm{EF}(e,P_i)=\frac{f_{e,P_i}}{\sum_{e'\in P_i} f_{e',P_i}},$$

其中 f_{e,P_i} 是实体 e 在检索出的所有路径 P_i 中出现的次数。实体频率 $\mathrm{EF}(e,P_i)$ 是实体 e 在检索出的所有路径 P_i 中出现的频率。

$$\mathrm{IDF}(e,D)=\log\frac{N}{|\{d\in D:e\in d\}|},$$

逆文档频率（IDF）用来衡量某个实体在整个文档集合中的重要性。N 是 EKG 中的文档节点 D 总数，$|\{d\in D:e\in d\}|$ 是包含实体 e 的文档数。

$$\text{EF-IDF}(e,d,D)=\mathrm{EF}(e,P_i)\times\mathrm{IDF}(e,D),$$

EF-IDF 用于评估路径中实体的重要程度。

$$\mathrm{Score}(P_i)=\frac{1}{|E_{P_i}|}\sum_{e\in E_{P_i}}\text{EF-IDF}(e,d,D),$$

我们计算每条路径的平均分数，其中，$|E_{P_i}|$ 是路径 P_i 中实体节点的数量。

最后，根据路径总分对所有路径进行排序，选取分数最高的前 $M=10$ 条路径作为候选路径。这种方法确保了每条路径中所有实体的重要性都被综合考虑，提高了推理的准确性和信息量。

4.2.2　关系路径构建

通过以上的路径检索步骤，我们已经得到了从

e_h到e_t的多条路径，每条路径表示为

$$P_i=\{e_h,d_1,e_2,d_2,\cdots,e_{m-1},d_{m-1},e_t\}。$$

随后，我们对路径上相邻的两个实体e_i和e_{i+1}进行关系抽取。具体来说，我们利用与这两个实体直接相连的文档节点d_i中的信息来进行关系抽取以及检索关系证据，从而构建关系路径。关系抽取的过程可以表示为

$$t_i=S(d_i,e_i,e_{i+1}),$$
$$r_i=RE(t_i,e_i,e_{i+1}),$$

其中，$S(d_i,e_i,e_{i+1})$表示我们使用向量检索的方式从与中e_i或e_{i+1}相关联的文档节点d_i检索出的与e_i或e_{i+1}相关的句子t_i作为关系证据。$RE(t_i,e_i,e_{i+1})$表示利用文档节点d_i中的信息t_i，推理出实体e_i和e_{i+1}之间的关系r_i。关系推理方法基于 LLMs 实现，具体的提示词见附录。

如果路径上的两个实体之间不存在任何关系，我们则认为e_i和e_{i+1}在这条路径上没有表达任何关系，则将该路径丢弃。

最后，我们得到了多个具有明确关系以及关系证据的三元组$[e_i,(r_i,t_i),e_{i+1}]$构成推理路径：

$$P_i=\{e_h,(r_1,t_1),e_2,r_2,\cdots,e_{m-1},(r_{m-1},t_{m-1}),e_t\}。$$

4.3 关系推理

通过以上步骤，我们得到多条关系路径P_i。随后基于这些路径信息利用大模型推理出头尾实体的关系。为了降低推理复杂度以及提高推理的准确性，我们的推理分为 2 步：(1) 单路径推理；(2) 多路径融合。

对于每条路径P_i，我们首先构建 Prompt，随后将 Prompt 输入大模型，得到头尾实体在这条路径上表达的关系。随后，我们通过融合多条路径上表达的关系得到头尾实体的关系：

$$P_i=\{e_h,(r_1,t_1),e_2,r_2,\cdots,e_{m-1},(r_{m-1},t_{m-1}),e_t\},$$
$$Prompt=T(P_i,Q),$$
$$A_i=\mathrm{LLM}(Prompt),$$
$$Answer=Merge(A_i),$$

其中，T是用于生成提示的函数，它将关系路径P_i和问题Q结合起来，生成一个综合性的输入提示。LLM 表示使用大模型推理，它接收提示输入并输出每条路径上表达的关系A_i。得到每条路径上表达的关系描述后，我们综合这多个关系得到一个融合答案。$Merge(A_i)$我们同样使用大模型实现，具体的提示词见附录。

4.4 反馈增强

最后我们将大模型的推理结果以及中间推理过程为知识图谱补充有价值的知识，从而形成知识在知识图谱与大模型之间的双向流动，形成循环正反馈，从而不断优化系统的效果。在以上的多个步骤中，我们可以得到多个大模型推理出的三元组信息，我们将这些信息经过验证后反馈增强到 EKG 中，从而不断丰富图谱的内容，并不断循环以优化系统性能。反馈增强过程可以表示为

$$\mathrm{Verify}(\{(e_i,(r_i,t_i),e_{i+1})\mid r_i,t_i\in P_i\}),$$
$$\{(e_i,(r_i,t_i),e_{i+1})\mid r_i,t_i\in P_i\}\rightarrow\mathrm{EKG}。$$

在 Verify(·)的实现中，我们采用更强大的模型(如 GPT-4)来对关系及其证据进行一致性验证，以确认提取的三元组(e_i,r_i,e_{i+1})与相应的关系证据t_i的一致性。如果验证成功，则将该三元组和关系证据纳入增强知识图谱(EKG)。具体的验证方法提示词详见附录。

此外，LLM 最终推理结果也可以视为新的文档节点，经过实体抽取和实体链接后添加到原有的图谱中：

$$E_i=\mathrm{NER}(Answer),$$
$$\mathrm{EL}(\bigcup_i E_i)\rightarrow\mathrm{EKG},$$

其中 NER(·)表示对结果进行实体抽取，随后进行实体链接 EL(·)，从而将大模型生成结果作为新的三元组添加到图谱中。

通过以上方式，我们能够将经过验证的答案和三元组信息不断反馈到知识图谱中，形成一个循环正反馈机制，逐步优化和增强系统的性能。

5 实 验

为了验证 BEKO 方法的有效性，我们在两个公开的关系抽取数据集上进行了全面实验，并与当前主流方法进行了详细对比分析。实验基于跨文档关系抽取的 CocRED[45] 数据集和文档关系抽取的 DocRED[24] 数据集，评估并比较了不同方法在关系抽取任务中的性能。此外，我们通过分析不同方法针对特定问题的推理结果，定性分析了方法在提升大模型生成的准确性和完整性方面的优势，以及双向增强框架的有效性。

我们试图回答以下三个关键问题：

RQ1：BEKO 方法是否在关系推理任务中优于现有单向增强和无增强方法？

RQ2：BEKO 方法对大模型的增强效果是否具有泛化性？

RQ3：多轮双向增强机制是否能够持续优化系

统的效果？

5.1 实验设置

数据集

(1) CodRED[45] 数据集：是一个设计用于跨文档关系抽取任务的开源数据集，旨在评估模型从多个文档中检索关系证据并推理实体关系的能力。我们随机抽取了 20 个类别，每个类别选取 30 条数据，共计 600 个样本作为测试基准。

(2) DocRED[24] 数据集：是一个用于文档级别关系抽取任务的开源数据集，旨在评估模型从单一文档中推断跨句实体关系的能力。我们同样随机抽取了 20 个类别，每个类别选取 50 条数据，共计 1000 个样本作为测试基准。

案例分析 为定性分析方法的有效性，实验中使用不同方法回答两个特定问题：Q_1："阿诺德·施瓦辛格与约翰·肯尼迪之间的关系？"以及 Q_2："阿诺德·施瓦辛格与卡罗琳·肯尼迪的关系？"。我们选取了与这两个问题相关的 20 个维基百科页面作为基础语料库，要求方法从中获取相关信息并推理出问题的答案。

对比方法 在定量分析实验中，我们比较了多种方法在两个数据集上的实验结果，包括基于 gpt-4o-mini-2024-07-18 和 gpt-4o-2024-05-13 模型实现的以下几种对比方法：

Only LLM 仅使用大模型对头尾实体进行关系推理，具体的提示词见附录。

Naive RAG 检索增强生成(RAG)是一种基于用户查询搜索信息并将结果作为参考提供给 LLM 生成答案的技术。这项技术是大多数基于大模型(LLM)工具的重要组成部分，大多数 RAG 方法使用向量相似性作为搜索技术。作为对比，我们使用 LangChainQ&A① 中的实现作为对比，命名为 Naive RAG，这是当前广泛使用的这一类 RAG 工具的一个知名代表示例。

GraphRAG Edge 等人[26] 提出了 GraphRAG 方法，该方法利用大语言模型(LLM)从私有数据集中构建知识图谱，并结合图机器学习技术以增强查询提示。GraphRAG 在开放域问答中表现出显著进步。其具体实现中缺乏从图谱中检索路径并进行关系推理的机制。我们基于 GraphRAG 方法构建知识图谱，随后从中检索出多跳路径并按照 4.2.1 节中的方法对所有路径进行排序，选出前 $M=10$ 条候选路径用于提示增强，从而推理出头尾实体之间的主要关系，具体的推理提示词见附录。

在定性分析中，我们对比了以下主流大模型的直接推理结果：(1) GPT-3.5-turbo②；(2) GPT-4o②；(3) DeepSeek-V3[50]；(4) 文言一心③；(5) GLM-4④；(6) ChatGLM3-6b[51]；此外，我们还对比了目前流行的两种 RAG 方法：一种是基于 LangChainQ&A 实现的使用向量相似性作为搜索技术的文档检索增强方法(Naive RAG)；另一种是具有在线搜索功能的 RAG 平台 Kimi⑤，该平台通过联网搜索实时获取相关信息补充上下文用于生成问题的答案。

评价指标 在 CodRED 和 DocRED 数据集上的实验中，我们采用宏观指标(Macro Metrics)进行评价，包括精确率(Precision)、召回率(Recall)和 $F1$ 分数。对于案例分析，重点分析实验结果的准确性和全面性，即推理结果是否与事实相符，以及是否包含全部的事实关系。

5.2 实验结果

5.2.1 RQ1：BEKO 方法是否在关系推理任务中优于现有单向增强和无增强方法？

在基于多个基础模型以及多个数据集上，BEKO 方法的性能均优于现有主流方法。BEKO 方法通过获取从头实体到尾实体的多跳关系路径信息以及与路径中关系的关系证据文本信息，实现更全面和准确的关系推理。如表 1 所示，BEKO 方法基于不同模型在不同数据集上均表现出最佳性能。具体而言，在 DocRED 数据集上，基于 GPT-4o 的 BEKO 方法达到了 0.743 的 $F1$ 值，相比于 GraphRAG 提高了 1.2%，比 Naive RAG 提升了 5.1%，比 Only LLM 方法提升了 3.5%。而基于 GPT-4o-mini 的 BEKO 方法达到了 0.637 的 $F1$ 值，相比于 GraphRAG 提高了 5.1%，比 Naive RAG 提升了 8.0%，比 Only LLM 方法提升了 8.6%。而在需要进行复杂的跨文档关系推理的 CodRED 数据集上，基于 GPT-4o 的 BEKO 方法达到了 0.776 的 $F1$ 值，相比于 GraphRAG 提高了 2.4%，比 Naive RAG 提升了 2.8%，比 Only LLM 方法提升了 6.8%。基于 GPT-4o-mini 的 BEKO 方法达到了 0.669 的 $F1$ 值，相比于 GraphRAG 提高了 3.0%，比 Naive RAG 提升了 7.7%，比 Only LLM 方法提升了 9.2%。

① https://python.langchain.com/v0.1/docs/use_cases/question_answering
② https://openai.com/
③ https://yiyan.baidu.com/
④ https://open.bigmodel.cn/
⑤ https://kimi.moonshot.cn/

表 1 主要实验结果

数据集		DocRED			CodRED		
基础模型	方法	*Precision*	*Recall*	*F*1	*Precision*	*Recall*	*F*1
GPT-4o-mini	Only LLM	0.551	0.531	0.551	0.652	0.605	0.577
	Naive RAG	0.557	0.538	0.557	0.630	0.626	0.592
	GraphRAG	0.586	0.587	0.586	0.689	0.685	0.639
	BEKO	0.637	0.627	**0.637**	0.736	0.697	**0.669**
	BEKO+MRE	0.635	0.628	**0.635**	0.731	0.703	**0.676**
GPT-4o	Only LLM	0.754	0.718	0.708	0.754	0.718	0.708
	Naive RAG	0.739	0.694	0.692	0.771	0.775	0.748
	GraphRAG	0.776	0.735	0.731	0.747	0.777	0.752
	BEKO	0.784	0.740	**0.743**	0.793	0.798	**0.776**
	BEKO+MRE	0.788	0.741	**0.746**	0.834	0.813	**0.789**

注:Only LLM 为直接使用大模型推理的方法;Naive RAG 为采用向量相似性检索的文档增强方法;GraphRAG 为结合知识图谱增强大模型的方法;BEKO 为我们提出的双向增强方法;BEKO+MRE 表示在 BEKO 方法的基础上进一步进行了多轮双向增强。

在定性分析中,BEKO 方法得到的推理结果相较于其他方法也更加准确和完整。如表 2 中的推理案例所示,主流大模型在阿诺德·施瓦辛格与约翰·肯尼迪关系推理任务中均未能完整推理出两者之间的多重关系。具体而言,GPT-4o 和文言一心模型虽然能识别两人间的家族联姻关联,但未能推理出其政治关联;而 GPT-3.5-turbo 和 GLM-4 则完全未能推理出两者之间的任何关联。这验证了大模型在缺乏外部显性知识支撑时,难以实现复杂关系的完整推理。相比之下,BEKO 方法通过构建 EKG(实体知识图谱)从语料库中检索出分散的、细粒度的事实知识单元,成功实现了对"家族联姻-政治"双重关系的准确推断。这些对比结果有力证明了 BEKO 方法在关系证据检索和推理方面的独特优势,其通过结构化的知识组织方式突破了传统大模型的知识边界限制。

表 2 对比案例分析

方法	回答	结果
GPT-3.5-turbo	阿诺德·施瓦辛格(Arnold Schwarzenegger)与约翰·肯尼迪(John F. Kennedy)之间没有直接的家族或亲属关系	准确 × 全面 ×
ChatGLM3-6B	阿诺德·施瓦辛格和约翰·肯尼迪之间没有直接的关系。约翰·肯尼迪是美国第 35 任总统,而阿诺德·施瓦辛格是一位著名的美国演员、政治家和前州长	准确 × 全面 ×
GLM-4	阿诺德·施瓦辛格和约翰·肯尼迪之间没有关系	准确 × 全面 ×
GPT-4o	阿诺德·施瓦辛格是约翰·肯尼迪的外甥女玛丽亚·施莱佛的前夫,因此与肯尼迪家族有姻亲关系。	准确 √ 全面 ×
DeepSeek-V3	阿诺德·施瓦辛格与约翰·肯尼迪之间没有直接关系。施瓦辛格是奥地利裔美国演员、政治家,而肯尼迪是美国前总统。施瓦辛格娶了肯尼迪的外甥女玛丽亚·施莱弗,因此两人是姻亲关系。	准确 √ 全面 ×
文言一心	施瓦辛格的妻子玛利亚·施莱弗是美国前总统约翰·肯尼迪的侄女	准确 √ 全面 ×
Naive RAG	阿诺德·施瓦辛格和约翰·肯尼迪之间没有直接的关系	准确 × 全面 ×
Kimi	阿诺德·施瓦辛格与约翰·肯尼迪之间没有直接的个人关系,但施瓦辛格与肯尼迪家族有联系,因为他娶了肯尼迪家族的成员玛利亚·施莱弗为妻	准确 √ 全面 ×
Ours	阿诺德·施瓦辛格与约翰·肯尼迪之间的关系:政治上,他们属于美国政治中的两个对立的主要政党——共和党和民主党;家庭上,通过阿诺德的前妻玛丽亚·施赖弗,阿诺德是约翰·肯尼迪的前侄女婿	准确 √ 全面 √

注:表格中展示了不同的方法得到关于 Q_1:"阿诺德·施瓦辛格与约翰·肯尼迪之间的关系"的推理结果。

5.2.2 RQ2:BEKO 方法是否对大模型的增强效果具有泛化性?

BEKO 方法在不同的大模型上均表现出显著的增强效果。通过基于增强知识图谱的检索方法,从语料库中检索出与头尾实体相关的结构化路径信息以及非结构化文本信息,并将这些信息作为外部显性知识输入大模型,显著提升了大模型对复杂关系的推理能力。如表 1 所示,无论是基于当前最先进的大模型之一 GPT-4o,还是性能较弱的 GPT-4o-mini,BEKO 方法在两个数据集上的实验结果均优于其他方法。特别是在基于较弱的 GPT-4o-mini 模型的实现中,BEKO 方法在两个数据集上的指标均显著优于其他方法。例如,在 DocRED 数据集上,基于 GPT-4o 的 BEKO 方法的 *F*1 值达到了 0.743,比同样基于 GPT-4o 的 GraphRAG 方法提升了 1.2%;基于 GPT-4o-mini 的 BEKO 方法的 *F*1 值达到了 0.637,比同样基于 GPT-4o-mini 的 GraphRAG 方法提升了 5.1%。在 CodRED 数据集上,基于 GPT-4o 的 BEKO 方法的 *F*1 值达到了 0.776,比同样基于 GPT-4o 的 GraphRAG 方法提升了 2.4%;基于 GPT-4o-mini 的 BEKO 方法的 *F*1 值达到了 0.669,比同样基于 GPT-4o-mini 的 GraphRAG 方法提升了 3.0%。这些结果表明,BEKO 方法在不同的模型上都能取得泛化的增强效果。

通过分析具体的案例也可以看出 BEKO 方法具有泛化性。如表 3 所示，即使更换基础大模型，BEKO 方法仍能有效地推理出 Q_1 的准确且全面的答案。这些实验结果进一步证明了 BEKO 方法在不同模型上的泛化能力。

表 3　泛化性实验

基础模型	回答	结果
GPT-3.5-turbo	阿诺德·施瓦辛格与约翰·肯尼迪之间的关系：政治上，他们属于美国政治中的两个对立的主要政党——共和党和民主党；家庭上，通过阿诺德的前妻玛丽亚·施赖弗，阿诺德是约翰·肯尼迪的前侄女婿	准确 √ 全面 √
Spark	阿诺德·施瓦辛格与约翰·肯尼迪之间的关系包括政治竞争关系和通过家庭关联，具体如下：政治竞争关系：阿诺德·施瓦辛格是共和党成员，而约翰·肯尼迪是民主党人士，两者属于美国两大主要政党，并在政治竞争中处于对立面。家庭关联：阿诺德·施瓦辛格的前妻玛丽亚·施赖弗的母亲尤妮斯·肯尼迪·施赖弗是约翰·肯尼迪的妹妹，因此阿诺德·施瓦辛格通过其前妻及其母亲与约翰·肯尼迪有关联	准确 √ 全面 √
GLM4	阿诺德·施瓦辛格与约翰·肯尼迪之间的关系包括：阿诺德·施瓦辛格是约翰·肯尼迪的外甥女玛丽亚·施赖弗的前夫，因此与约翰·肯尼迪有姻亲关系。在政治上，阿诺德·施瓦辛格与约翰·肯尼迪分属不同政党，施瓦辛格是共和党人，而肯尼迪是民主党人，他们在政治理念和党派立场上存在根本差异	准确 √ 全面 √
ChatGLM3-6B	阿诺德·施瓦辛格是美国第 35 任总统约翰·肯尼迪的妹妹的丈夫。因为玛丽亚·施赖弗是尤妮斯·肯尼迪·施赖弗的女儿，所以阿诺德·施瓦辛格是美国第 35 任总统肯尼迪家族的成员。同时，阿诺德·施瓦辛格也是约翰·肯尼迪的亲戚，因为他是约翰·肯尼迪的孙女	准确 × 全面 ×

注：基于不同的模型实现 BEKO 方法得到 Q_1 的推理结果。

然而，当关系推理链路过长时，基于参数量较小的模型可能会出现推理错误。这是因为过长的推理链路导致最终输入到大模型的上下文过长，使得部分内容在推理过程中被遗忘[52]。这一现象表明，尽管 BEKO 方法在大多数情况下表现出色，但在处理复杂推理任务时，模型的参数量和上下文处理能力仍然是关键因素。

5.2.3　轮双向增强机制是否能不断优化系统的效果？

在不同模型和多个数据集上的实验结果表明，BEKO 方法中创新的双向增强机制显著提升了关系抽取任务的综合性能。该机制通过建立大语言模型与知识图谱之间的动态交互，将模型的推理结果以结构化三元组形式反馈至知识图谱，进而形成知识要素在知识图谱与大模型之间的双向流动。这种机制不仅持续丰富了知识图谱的内容，更重要的是通过知识反哺持续优化大模型的推理能力，形成“数据增强-知识积累-能力提升”的良性循环。随着双向增强的不断进行，推理效果得以持续优化。

表 1 展示了基于不同模型在 DocRED 数据集以及 CodRED 数据集上的 BEKO＋MRE 方法（其中 MRE 表示 Multi Round Enhancement，即多轮增强）的实验结果。如表 1 所示，经过多轮双向增强后，基于不同模型的 BEKO 方法在两个数据集上的实验结果均有不同程度的提升。例如，在 CodRED 数据集上，基于 GPT-4o 模型实现的 BEKO＋MRE 方法将 $F1$ 分数从未进行多轮双向增强前的 0.776 提升至 0.789，增幅为 1.3%；基于 GPT-4o-mini 实现的 BEKO＋MRE 方法则将 $F1$ 分数从未进行多轮双向增强前的 0.635 提升至 0.637，增幅为 0.2%。这些结果表明，多轮双向增强机制通过丰富知识图谱内容增强了 BEKO 方法的关系推理能力。

双向增强机制使得 BEKO 方法在关系推理任务上的推理效果持续提升并趋于收敛。表 4 以及图 8 展示了在多轮双向增强过程中各项指标的变化趋

表 4　在 CodRED 数据集上的多轮次双向增强实验结果以及变化趋势

数据集		CodRED			
基础模型	轮次	*Precision*	*Recall*	*F*1	*CR*
GPT-4o	R0	0.793	0.798	0.776	0.677
	R1	0.774	0.805	0.780	0.758
	R2	0.778	0.803	0.781	0.765
	R3	0.778	0.803	0.779	0.773
	R4	0.828	0.807	0.782	0.775
	R5	0.831	0.810	0.786	0.777
	R6	0.834	0.813	0.789	0.777
	R7	0.831	0.810	0.785	0.777
	R8	0.834	0.813	0.789	0.780
	R9	0.834	0.813	0.789	0.780
	R10	0.834	0.813	0.789	0.780

注：*CR*（Connectivity Ratio）表示在最后推理阶段前，检索到从头实体到尾实体的有效联通路径的比率。R* 表示双向增强轮次，R0 表示未进行双向增强前的实验结果。

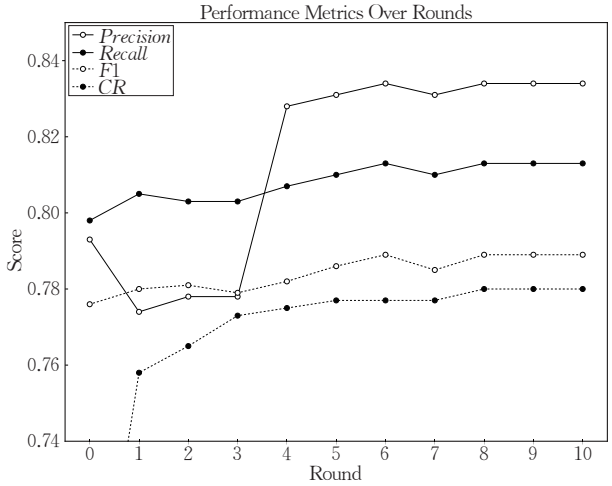


图 8　多轮增强结果折线图

势。如表 4 以及图 8 所示，随着多轮双向增强的持续进行，在 CodRED 数据集上的各项指标都在持续提升。具体而言，随着双向增强的持续进行，$F1$ 分数从未增强前的 0.776 逐渐提高，并最终收敛稳定于 0.789，增幅为 1.3%。

此外，双向增强机制使 BEKO 方法在检索阶段的探索范围持续拓展，从而检索到更多的有效信息。通过将大模型的推理结果反馈到知识图谱中，可以有效扩展检索阶段的探索范围。例如，原本需要三跳的路径可以缩短为两跳，原本无法联通的路径通过新的三元组实现联通。如表 4 以及图 8 所示，观察推理阶段前检索到的从头实体到尾实体的有效联通路径比率(Connectivity Ratio，CR)，可以看到，该比率随着双向增强轮次的增加而不断上升。未增强前，只有 66.7%的样本在最终的推理阶段前可以检索到从头实体到尾实体的有效联通路径，而在多轮的双向增强后，这一比率提升至了 78.0%。随着联通率的提升，其他各项实验指标也随之逐渐提升。此外，从表 5 也可以观察到，在使用与 Q_2 相关的 Q_1 的推理结果增强图谱前，BEKO 方法无法正确推理出 Q_2 的答案，而在使用 Q_1 的推理结果：〈玛丽亚·施赖弗，前妻，阿诺德·施瓦辛格〉，〈阿诺德·施瓦辛格，前侄女婿，约翰·肯尼迪〉增强图谱内容后，则可以准确推理出 Q_2 的答案。以上结果表明，通过多轮的双向增强机制，知识图谱的内容在不断地丰富，从而提升了检索阶段的探索范围，提高了检索的有效性，进而增强了 BEKO 方法的复杂关系推理能力。

表 5 双向增强实验案例分析结果

方法	回答	结果
$\rightarrow Q_2$	阿诺德·施瓦辛格与卡罗琳·肯尼迪之间没有直接的家族关系。	准确 ×
$Q_1 \rightarrow Q_2$	阿诺德·施瓦辛格与卡罗琳·肯尼迪之间的关联是：阿诺德·施瓦辛格是卡罗琳·肯尼迪的前外甥舅。	准确 √

注：其中$\rightarrow Q_2$表示使用 BEKO 方法获取 Q_2 回答的结果。$Q_1 \rightarrow Q_2$表示在得到 Q_1 的结果并反馈增强图谱后再推理 Q_2 得到的结果。

然而，双向增强机制在提升 BEKO 方法的推理结果方面存在局限性。从表 4 和图 8 中可以观察到，随着双向增强的持续进行，BEKO 方法在关系推理性能上的提升逐渐变小并趋于收敛，而不是持续增长。这是因为随着过程的推进，从相同样本的推理过程以及结果中获得的知识逐渐减少，导致添加到知识图谱中的知识也相应减少。图 9 展示了在 CodRED 数据集上的多轮双向增强中每个轮次抽取出的新三元组数量变化。在多轮双向增强过程中，每轮新增的三元组数量逐步减少，从第 1 轮的 2368 个减少到第 10 轮的仅 6 个。这表明随着轮次增加，图谱中新增的有价值知识显著减少。此外，大模型推理结果以及最终的验证过程中可能存在误差，导致错误知识的引入，从而为知识图谱增加了噪音。这两种因素共同作用，使得关系推理性能的提升逐渐减小并趋于稳定。综上所述，双向增强机制在提升 BEKO 方法推理性能方面的效果是显著的，但其提升幅度逐渐减小并最终收敛。

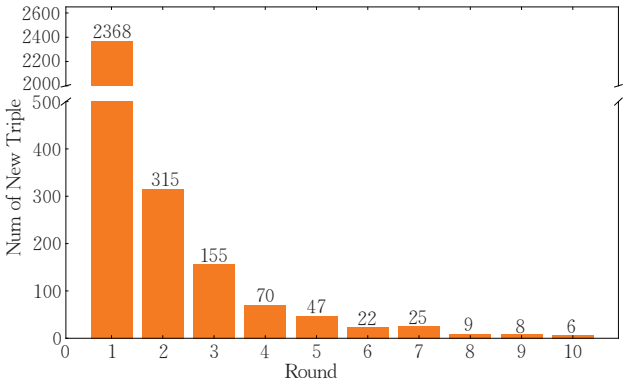


图 9 新三元组数量变化趋势(纵坐标表示每轮获取的新三元组数量；横坐标表示关系推理的轮数)

6 结 论

本文提出的 BEKO 方法在关系推理任务中表现优异，在多个数据集上取得了卓越性能，验证了其创新性和有效性。该方法基于增强知识图谱(EKG)的检索机制，能够有效提取与问题相关的细粒度分散事实知识，从而实现了比现有主流方法更强的推理能力。其多轮双向增强机制将大模型的推理结果反馈至知识图谱，不仅丰富了知识图谱内容，还提升了大模型的推理能力，形成双向知识循环，持续优化推理效果。在 CodRED 和 DocRED 数据集上的实验结果充分证明了 BEKO 方法的优越性及其双向增强机制的有效性。未来研究将致力于将该方法扩展到如通用问答和事实验证等更多任务中，并探索其实际应用潜力。

参 考 文 献

[1] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901

[2] Arefeen M A, Debnath B, Chakradhar S. LeanContext: Cost-efficient domain-specific question answering using LLMs.

Natural Language Processing Journal, 2024, 7: 100065

[3] Atanasova P, Simonsen J G, Lioma C, et al. Generating fact checking explanations//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online, 2020: 7352-7364

[4] Lin S, Warner J, Zamfirescu-Pereira J, et al. Rambler: Supporting writing with speech via LLM-assisted gist manipulation//Proceedings of the CHI Conference on Human Factors in Computing Systems. Honolulu, USA, 2024: 1-19

[5] Ji Z, Lee N, Frieske R, et al. Survey of hallucination in natural language generation. ACM Computing Surveys, 2023, 55(12): 1-38

[6] Hamed A A, Wu X. Detection of ChatGPT fake science with the xFakeSci learning algorithm. Scientific Reports, 2024, 14(1): 16231

[7] Wu Xin-Dong, Bai Ting, Zhang Jie, et al. Knowledge Graph. Beijing: Science Press, 2022(in Chinese)
(吴信东, 白婷, 张杰等. 知识图谱. 北京: 科学出版社, 2022)

[8] Pan S, Luo L, Wang Y, et al. Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7): 3580-3599

[9] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474

[10] Xu X, Liu Y, Pasupat P, et al. In-context learning with retrieved demonstrations for language models: A survey. arXiv preprint arXiv:2401.11624, 2024

[11] Ma X, Gong Y, He P, et al. Query rewriting for retrieval-augmented large language models. arXiv preprint arXiv: 2305.14283, 2023

[12] Baek J, Aji A F, Saffari A. Knowledge-augmented language model prompting for zero-shot knowledge graph question answering. arXiv preprint arXiv:2306.04136, 2023

[13] Jiang J, Zhou K, Dong Z, et al. StructGPT: A general framework for large language model to reason over structured data//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 9237-9251

[14] Sun J, Xu C, Tang L, et al. Think-on-graph: Deep and responsible reasoning of large language model with knowledge graph. arXiv preprint arXiv:2307.07697, 2023

[15] Zhao P, Zhang H, Yu Q, et al. Retrieval-augmented generation for AI-generated content: A survey. arXiv preprint arXiv:2402.19473, 2024

[16] Izacard G, Grave E. Leveraging passage retrieval with generative models for open domain question answering// Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. Online, 2021: 874-880

[17] Ram O, Levine Y, Dalmedigos I, et al. In-context retrieval-augmented language models. Transactions of the Association for Computational Linguistics, 2023, 11: 1316-1331

[18] Yang L, Chen H, Li Z, et al. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7): 3091-3110

[19] Chen Z, Wang Y, Zhao B, et al. Knowledge graph completion: A review. IEEE Access, 2020, 8: 192435-192456

[20] Ji S, Pan S, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494-514

[21] Jehangir B, Radhakrishnan S, Agarwal R. A survey on named entity recognition—datasets, tools, and methodologies. Natural Language Processing Journal, 2023, 3: 100017

[22] Sevgili Ö, Shelmanov A, Arkhipov M, et al. Neural entity linking: A survey of models based on deep learning. Semantic Web, 2022, 13(3): 527-570

[23] Detroja K, Bhensdadia C, Bhatt B S. A survey on relation extraction. Intelligent Systems with Applications, 2023, 19: 200244

[24] Han R, Peng T, Wang B, et al. Document-level relation extraction with relation correlations. Neural Networks, 2024, 171: 14-24

[25] Zhong L, Wu J, Li Q, et al. A comprehensive survey on automatic knowledge graph construction. ACM Computing Surveys, 2023, 56(4): 1-62

[26] Edge D, Trinh H, Cheng N, et al. From local to global: A graph RAG approach to query-focused summarization. arXiv preprint arXiv:2404.16130, 2024

[27] Li T, Ma X, Zhuang A, et al. Few-shot in-context learning on knowledge base question answering//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada, 2023: 6966-6980

[28] Baek J, Aji A F, Lehmann J, et al. Direct fact retrieval from knowledge graphs without entity linking//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada, 2023: 10038-10055

[29] Li X, Zhao R, Chia Y K, et al. Chain of knowledge: A framework for grounding large language models with structured knowledge bases. arXiv preprint arXiv: 2305.13269, 2023

[30] Wang K, Duan F, Wang S, et al. Knowledge-driven CoT: Exploring faithful reasoning in LLMs for knowledge-intensive question answering. arXiv preprint arXiv:2308.13259, 2023

[31] Zhao Z, Luo X, Chen M, et al. A survey of knowledge graph construction using machine learning. CMES-Computer Modeling in Engineering & Sciences, 2023, 139(1): 225-257

[32] Cheng S, Zhang N, Tian B, et al. Editing language model-based knowledge graph embeddings//Proceedings of the 38th

AAAI Conference on Artificial Intelligence: 38. Washington, USA, 2024: 17835-17843

[33] Cao J, Fang J, Meng Z, et al. Knowledge graph embedding: A survey from the perspective of representation spaces. ACM Computing Surveys, 2024, 56(6): 1-42

[34] Wan Z, Cheng F, Mao Z, et al. GPT-RE: In-context learning for relation extraction using large language models. arXiv preprint arXiv:2305.02105, 2023

[35] Efeoglu S, Paschke A. Retrieval-augmented generation-based relation extraction. arXiv preprint arXiv:2404.13397, 2024

[36] Xie X, Zhang N, Li Z, et al. From discrimination to generation: Knowledge graph completion with generative transformer//Companion Proceedings of the Web Conference 2022. Lyon, France, 2022: 162-165

[37] Chen Z, Xu C, Su F, et al. Incorporating structured sentences with time-enhanced BERT for fully-inductive temporal relation prediction//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei, China, 2023: 889-899

[38] Su Y, Han X, Zhang Z, et al. CokeBERT: Contextual knowledge selection and embedding towards enhanced pre-trained language models. AI Open, 2021, 2: 127-134

[39] Zhang Z, Han X, Liu Z, et al. ERNIE: Enhanced language representation with informative entities. arXiv preprint arXiv: 1905.07129, 2019

[40] Zhu H, Peng H, Lyu Z, et al. Pre-training language model incorporating domain-specific heterogeneous knowledge into a unified representation. Expert Systems with Applications, 2023, 215: 119369

[41] 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, 2022: 5049-5060

[42] Zhang X, Bosselut A, Yasunaga M, et al. GreaseLM: Graph reasoning enhanced language models for question answering. arXiv preprint arXiv:2201.08860, 2022

[43] Wang Y, Lipka N, Rossi R A, et al. Knowledge graph prompting for multi-document question answering. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38: 19206-19214

[44] Cohen W W, Chen W, De Jong M, et al. QA is the new KR: Question-answer pairs as knowledge bases//Proceedings of the 37th AAAI Conference on Artificial Intelligence, 2023, 37: 15385-15392

[45] Yao Y, Du J, Lin Y, et al. CodRED: Across-document relation extraction dataset for acquiring knowledge in the wild//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta Cana, The Dominican Republic, 2021: 4452-4472

[46] Lu K, Hsu I H, Zhou W, et al. Multi-hop evidence retrieval for cross-document relation extraction//Findings of the Association for Computational Linguistics: ACL 2023. Toronto, Canada, 2023: 10336-10351

[47] Siriwardhana S, Weerasekera R, Wen E, et al. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. Transactions of the Association for Computational Linguistics, 2023, 11: 1-17

[48] Guo Z, Schlichtkrull M, Vlachos A. A survey on automated fact-checking. Transactions of the Association for Computational Linguistics, 2022, 10: 178-206

[49] Schlichtkrull M, Guo Z, Vlachos A. AVeriTeC: A dataset for real-world claim verification with evidence from the Web. Advances in Neural Information Processing Systems, 2023, 36

[50] DeepSeek-AI, Liu A, Feng B, et al. DeepSeek-V3 Technical Report. CoRR, 2024, abs/2412.19437

[51] Zeng A, Xu B, Wang B, et al. ChatGLM: A family of large language models from GLM-130B to GLM-4 all tools. CoRR, 2024, abs/2406.12793

[52] Liu N F, Lin K, Hewitt J, et al. Lost in the middle: How language models use long contexts. Transactions of the Association for Computational Linguistics, 2024, 12: 157-173

附 录

实体抽取提示词

"""

-Goal-

1. Identify all named entities in the given text.

2. Format entities result as

〈entity_name〉

〈entity_name〉

〈entity_name〉

-Examples-

Text:

"On September 1,2023, Apple Inc. launched its latest iPhone model in Cupertino, California. "

Output:

September 1,2023

Apple Inc.

California

Cupertino

iPhone

-Attention-

-You should identify named entities in the text as much as possible.

-Only extract explicit named entities; avoid extracting general

terms like "actor", "person".
-Ensure entity names consistent with those in the text.
-Ensure that all extracted entities are named entities.
-Real Data-
Text:
{input_text}
Output:
"""

关系抽取提示词

"""
-Goal-
Given two entities and some textual information, please choose the most appropriate relationship between the subject and object from the Candidate Relation Options.
-Attentions-
* Please do not provide additional content such as explanations.
* The final output must choose the relationship name from the Candidate Relationship Options I provide, otherwise you will face serious penalties.
* Only return relationship name, not relation description.
* Please pay attention to distinguish between subject and object; the direction of the extracted relationship is from subject to object.
-Real Data-
Subject:
{subj}
Object:
{obj}
Textual Information:
{info}
Candidate Relation Options:
{class_of_rel}
Output:
"""

一致性验证提示词

"""
-Goal-
Given two entities and the relationship between them, as well as some sentences from different articles that may validate this relationship, please determine the correctness of the relationship based on the information in the sentences.
-Attentions-
-Please do not provide additional content such as explanations, Only return True or False.
Subject:
{subj}
Object:
{obj}
Relationship:
{rel}
Textual Information:
{triple_info}
"""

Only LLM 方法推理提示词

"""
-Goal-
Given two entities and some textual information, please choose the most appropriate relationship from the options I provide.
-Attentions-
-Please do not provide additional content such as explanations.
-Please strictly choose according to the relationships I provided.
-The final output must choose the relationship name from the Candidate Relationship Options I provide, otherwise you will face serious penalties.
-Only return relationship name, not relation description.
-Real Data-
Subject:
{subj}
Object:
{obj}
Candidate Relation Options:
{class_of_rel}
Output:
"""

GraphRAG 方法推理提示词

"""
-Goal-
Given two entities and some graph path information, please choose the most appropriate relationship from the options I provide based on the graph path information and your knowledge.
-Attentions-
-Please strictly follow the format in the example to answer, do not provide additional content such as explanations.
-Please strictly choose according to the relationships I provided.
-The final output must choose the relationship name from the Candidate Relationship Options I provide, otherwise you will face serious penalties.
-Only return relationship name, not relation description.
-Real Data-
Subject:
{subj}
Object:
{obj}
Graph Path Information:
{info}
Candidate Relation Options:
{class_of_rel}
Output:
"""

WU Xin-Dong, Ph. D., professor. His main research interests include data mining, big data analytics, and knowledge engineering.

HUANG Man-Zong, Ph. D. candidate. His primary research areas are knowledge graphs and data mining.

BU Chen-Yang, Ph. D., associate professor. His main research interest is knowledge-driven optimization.

Background

The research presented in this paper delves into the critical intersection of large language models (LLMs) and knowledge graphs (KGs), a topic of paramount importance within the rapidly evolving fields of artificial intelligence (AI) and natural language processing (NLP). Despite the remarkable advancements in LLMs, as exemplified by state-of-the-art models like ChatGPT, significant challenges remain unresolved, particularly concerning the accuracy, reliability, and completeness of the content generated by these models. These challenges are especially critical given the increasing reliance on LLMs across a wide range of domains, including economics, politics, cultural studies, and beyond. The potential for misinformation, semantic inconsistencies, and incomplete knowledge representation underscores the urgent need for innovative solutions to enhance the performance of these models.

In recent years, researchers have made substantial progress in understanding how to integrate external knowledge into LLMs, with a particular focus on leveraging the structured and semantically rich nature of KGs. These efforts have explored various dimensions of knowledge integration, including the generation, evolution, and propagation of structured knowledge, as well as its impact on improving the contextual understanding and reasoning capabilities of LLMs. However, the unprecedented explosion of online data sources over the past decade presents a new and unique opportunity to validate and refine these methods using massive, real-world datasets derived from diverse interactions. This wealth of data provides a fertile ground for testing the robustness and scalability of existing approaches, while also uncovering new challenges and opportunities for innovation.

This paper aims to address several outstanding issues in the field by proposing a novel bidirectional enhancement framework that synergistically leverages the strengths of both KGs and LLMs. Our proposed system, named Bidirectional Enhancement with Knowledge Ocean (BEKO), is designed to create a continuous positive feedback loop for system optimization. BEKO seeks to address two critical challenges: (1) reducing the initial costs associated with KG construction, which often require significant human effort and computational resources, and (2) minimizing semantic loss during the transformation of unstructured data into structured knowledge representations. By fostering a bidirectional interaction between KGs and LLMs, BEKO enables more accurate extraction, representation, and application of knowledge, thereby enhancing the contextual understanding and reasoning capabilities of LLMs.

The bidirectional nature of BEKO allows for a dynamic and iterative process of knowledge enrichment. On one hand, KGs provide LLMs with structured, high-quality knowledge that can be used to ground their outputs in factual information. On the other hand, LLMs contribute to the evolution and expansion of KGs by extracting new knowledge from unstructured text and refining existing knowledge through advanced reasoning and inference. This symbiotic relationship not only improves the performance of both components but also ensures that the system remains adaptive to new information and evolving contexts.

This work was supported by the National Natural Science Foundation of China under Grant No. 62120106008, the Anhui Province Science and Technology Fortification Plan under Grant No. 202423k09020015, and the Youth Talent Support Program of Anhui Association for Science and Technology (No. RCTJ202420). Our research is deeply rooted in the construction and applications of knowledge graphs, with a focus on advancing the state-of-the-art in AI and NLP through innovative methodologies and practical solutions. By addressing the challenges of knowledge integration and semantic representation, we aim to contribute to the development of more reliable, accurate, and contextually aware AI systems that can be deployed across a wide range of real-world applications.