

控制与决策

Control and Decision

大语言模型时代下的知识图谱构建综述

和婷婷, 张强, 郑冠彧, 赵铁军, 王浩畅, 王颖

引用本文:

和婷婷, 张强, 郑冠彧, 等. 大语言模型时代下的知识图谱构建综述[J]. *控制与决策*, 2025, 40(12): 3509–3527.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0413>

您可能感兴趣的其他文章

Articles you may be interested in

基于云模型和多层权重求解的多粒度语言大群体决策方法

Multi-granularity linguistic large group decision-making based on cloud model and multi-layer weight determination

控制与决策. 2021, 36(9): 2257–2266 <https://doi.org/10.13195/j.kzyjc.2020.0102>

实体消歧综述

Entity disambiguation: A review

控制与决策. 2021, 36(5): 1025–1039 <https://doi.org/10.13195/j.kzyjc.2020.0388>

融合稀疏编码与深度学习的草图特征表示

A feature representation of sketch based on fusion of sparse coding and deep learning

控制与决策. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

基于深度学习的行人轨迹预测方法综述

Survey of pedestrian trajectory prediction methods based on deep learning

控制与决策. 2021, 36(12): 2841–2850 <https://doi.org/10.13195/j.kzyjc.2020.1841>

基于知识粒度特征的多目标粗糙集属性约简算法

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

大语言模型时代下的知识图谱构建综述

和婷婷¹, 张强^{1†}, 郑冠彧², 赵铁军³, 王浩畅¹, 王颖¹

1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318;
2. 华南理工大学 软件学院, 广州 510006;
3. 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 知识图谱旨在将现实世界中的对象、概念及其相互关系, 以图的形式进行有效组织. 面对不断演变的真实世界信息, 传统静态知识图谱在数据质量、准确性、复杂性、动态更新维护等方面面临挑战. 近期, 大语言模型在语义理解和文本生成方面取得了卓越进展, 其跨领域、跨模态、跨任务的强泛化能力为知识图谱构建提供了新机遇. 鉴于此, 针对大语言模型在知识图谱构建中的研究进展进行综述. 首先, 介绍知识图谱和大语言模型的基本概念以及两者融合的通用计算框架; 然后, 围绕知识图谱的生命周期, 深入分析和讨论大语言模型在知识抽取、知识融合和知识推理这3大关键环节的进展和挑战; 接着, 从知识问答和检索增强生成系统剖析大语言模型驱动的知识图谱的应用成效; 最后, 总结和展望大语言模型时代知识图谱构建的发展趋势, 多角度探讨亟待研究的问题, 为未来研究提供启示.

关键词: 大语言模型; 知识图谱构建; 知识抽取; 知识融合; 知识推理; 知识问答; 检索增强生成

中图分类号: TP309 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2025.0413

引用格式: 和婷婷, 张强, 郑冠彧, 等. 大语言模型时代下的知识图谱构建综述 [J]. 控制与决策, 2025, 40(12): 3509-3527.

A survey of knowledge graph construction in the era of large language models

HE Ting-ting¹, ZHANG Qiang^{1†}, ZHENG Guan-yu², ZHAO Tie-jun³, WANG Hao-chang¹, WANG Ying¹

1. School of Computer & Information Technology, Northeast Petroleum University, Daqing 163318, China;
2. School of Software Engineering, South China University of Technology, Guangzhou 510006, China;
3. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Knowledge graphs aim to organize real-world entities, concepts, and their relations in a structured graph form. Traditional static knowledge graphs face challenges in data quality, accuracy, complexity, and dynamic updates. Real-world information evolves constantly, increasing the difficulty of maintenance. Recently, large language models achieve remarkable progress in semantic understanding and text generation. Their strong generalization ability across domains, modalities, and tasks brings new opportunities for knowledge graphs construction. This paper surveys recent advances in using large language models for building knowledge graphs. First, it introduces the basic concepts of knowledge graphs and large language models, and outlines a general framework for their integration. Then, it analyzes the progress and challenges of large language models in three key phases: knowledge extraction, knowledge fusion, and knowledge reasoning. Next, it discusses practical applications in knowledge-based question answering and retrieval-augmented generation systems. Finally, it summarizes development trends and open problems, providing insights for future research.

Keywords: large language models; knowledge graphs construction; knowledge extraction; knowledge fusion; knowledge reasoning; knowledge-based question answering; retrieval-augmented generation

收稿日期: 2025-04-18; 录用日期: 2025-08-19.

基金项目: 国家自然科学基金项目 (42002138); 黑龙江省优秀青年教师基础研究支持计划项目 (YQJH2023073); 黑龙江省自然科学基金项目 (LH2024F005); 黑龙江省研究生课程思政建设项目计算智能及石油大数据分析导学思政团队项目 (YJSKCSZ_202303); 河北省自然科学基金项目 (D2023107002).

责任编委: 阳春华.

†通信作者. E-mail: nepu_zq@163.com.

0 引言

知识图谱 (KG)^[1] 是以图形式表示客观世界中的实体及其关系的知识库^[2], 由实体和关系组成. 实体作为节点, 表示真实世界中的物体或抽象的概念; 关系作为边, 表示实体间的联系. 2012 年谷歌首次提出了知识图谱的概念^[3], 随后广泛将其应用于问答^[4]、推荐^[5] 和搜索^[6] 等多个下游任务. 知识图谱的构建质量是智能系统效果的基础, 直接影响推理的准确性、决策的可靠性、下游任务的应用效率等.

依赖于深度学习 (DL) 技术, 从结构化和非结构化数据中提取三元组并形成图谱的构建方法^[7-9], 是大语言模型兴起前的主流手段, 但是其结果多为静态图谱, 泛化性、推理性、实时性和解释性均具有局限性, 不足以支持知识的动态更新和跨场景迁移.

近期, 以 DeepSeek^[10]、GPT-4^[11] 为代表的大语言模型 (LLMs) 在语义理解、上下文推理和文本生成方面取得了显著突破, 展现出强大的跨领域、跨模态、跨任务的泛化能力, 从而衍生了一系列融合大语言模型的知识图谱构建研究成果^[12]. 本质上, LLMs 和 KG 均是存储和利用知识的一种方式. 形式上, 前者通过网络参数隐式地存储知识, 后者通过三元组显式地存储知识. 其中: LLMs 既能够完成场景复杂的生成任务, 又可应用于开放域理解任务, 在少样本和零样本场景内, 知识表示能力甚至超过有监督模型^[13]. 将 LLMs 用于知识图谱构建, 首先, 能够有效从多源异构数据中提取结构化知识并更新至 KG 中; 其次, 通过表征实体及其关系来推理补全或校正图谱内容, 从而提升构建质量; 再次, 可从多种形式的用户查询中准确识别其意图, 缩小召回范围, 以提高知识检索的准确性^[14-15]. LLMs 以其出色的语言建模和推理能力, 正在推动知识图谱构建进入新阶段. 值得注意的是, 本文提到的 LLMs, 涵盖“预训练+微调”范式的预训练语言模型 (PLMs)^[16] 以及“预训练+提示”范式的大语言模型^[17-18].

近两年进入 LLMs 时代后, KG 迎来新的发展机遇, 学者们纷纷发表了综述文章. 如: 曾泽凡等^[19] 从知识图谱的基本构成要素出发, 梳理了基于 PLMs 的相关算法模型和方法的核心思路; 李紫宣等^[20] 综述了代码大语言模型在 KG 中的应用, 主要介绍了代码化的知识表示、知识抽取和知识推理; 杨观赐等^[21] 重点分析了 KG 在搜索引擎优化、智能问答系统、具身智能系统等典型应用领域. 尽管这些综述均为近两年在大语言模型背景下发表, 然而, 鲜有从文档级、多模态、低资源、跨领域等具体任务不同场景

的角度进行深入细致分析. 鉴于此, 在对已有综述进行深入剖析和思考的基础上, 本文将视角拓展至整个大语言模型时代, 围绕“大语言模型如何赋能知识图谱构建”这一核心问题, 系统地梳理该方向的最新研究进展. 与已有综述相比, 本文能够更加精准地呼应实际应用中知识图谱构建多样化的任务需求, 更加深入地挖掘 LLMs 对知识图谱构建的潜在优势. 本文主要内容如下:

1) 以知识图谱生命周期为视角, 系统地梳理 LLMs 在知识抽取、知识融合、知识推理 3 个核心阶段中的应用进展, 并深入剖析各阶段具体任务场景所面临的关键挑战;

2) 探讨 LLMs 时代下 KG 的应用价值, 首先分析 KG 在问答系统中与 LLMs 的协同构建和集成应用, 然后聚焦于 KG 在检索增强生成框架中的关键作用和实践路径;

3) 总结当前知识图谱构建所面临的主要挑战, 进一步提出若干具有前瞻性的研究方向.

1 知识图谱定义及其构建任务

知识图谱通常采用三元组的形式, 将现实世界的实体及其相互关系以结构化的方式进行存储, 实体和关系分别视作节点和连接节点的有向边^[22], 可表示为 $G = (E, R, S)$. 其中: E 为实体集合, R 为关系集合, S 为包含在 $R \times E \times E$ 中的三元组集合. 每个三元组通常表示一条事实, 包括头实体、尾实体和关系. 在特定情况下, “关系”亦称为“属性”, 对应尾实体被视为属性值. 目前, 公开的大规模知识图谱包括 Freebase^[23]、DBpedia^[24]、Knowledge Vault^[25] 等. 这些图谱广泛应用于多种人工智能任务中, 成为认知智能的重要基础设施. 知识图谱的构建流程包括知识获取、知识融合和知识推理^[26] 这几个核心阶段. 在传统方法中, 这一流程依赖人工标注或深度学习模型对语料进行信息抽取, 存在知识更新滞后、泛化能力有限以及构建成本高等问题.

伴随 LLMs 的飞速发展, 融合 LLMs 的知识图谱构建方法受到了广泛关注. LLMs 具备强大的语言生成能力、上下文理解能力、零样本/少样本学习、跨领域知识整合等能力, 面对海量非结构化文本数据, 能够自动化地完成语言学知识^[27]、世界知识^[28] 和常识知识^[29] 的实体识别以及关系抽取任务, 同时, 具备消除歧义、推理和更新功能, 为高质量知识图谱的自动化构建提供了新路径.

图 1 为融合 LLMs 的知识图谱构建通用框架.

该框架包括数据层、模型层和应用层: 数据层是 KG 构建的基础, 涵盖表格等结构化数据, 百科等半结构化数据以及文本、音频等多模态的非结构化数据; 模型层充分利用 LLMs 的优势, 将 LLMs 深度融合 KG 构建的全流程中, 完成知识抽取、知识融合、知识推

理等任务, 显著提升了知识图谱的构建质量和自动化程度; 应用层侧重于将融合后的 LLMs 和 KG 应用于更复杂的下游任务, 如检索增强生成、智能问答、语义检索等, 进一步实现知识驱动的智能决策和信息服务。

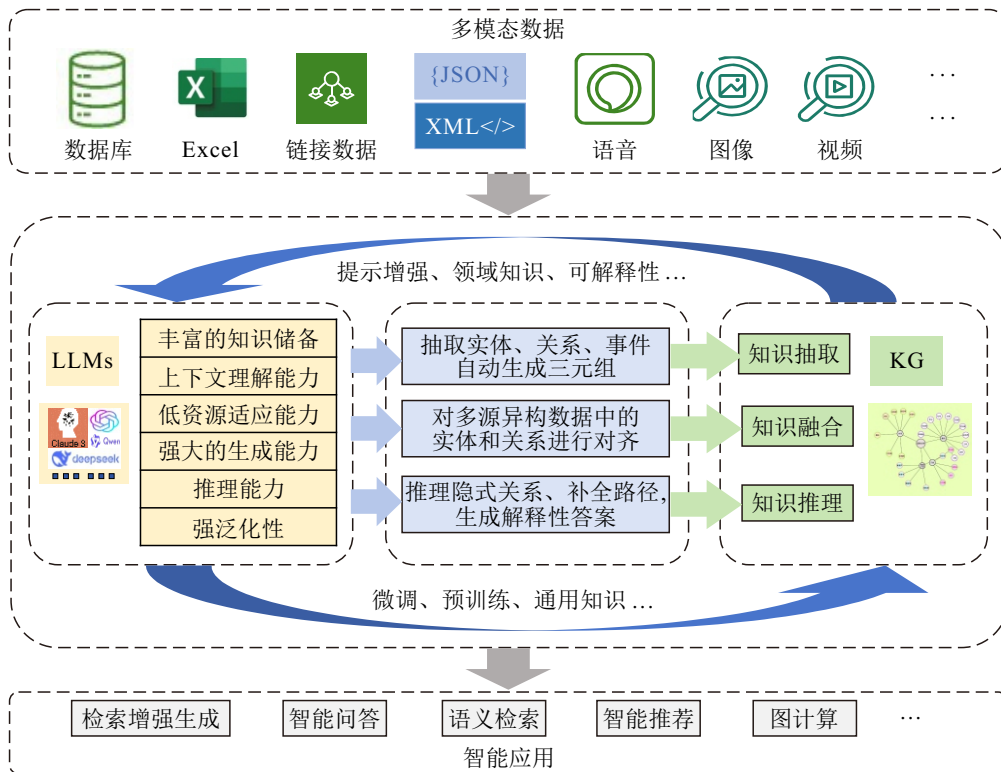


图1 LLMs 和 KG 构建结合的通用框架^[18]

2 大语言模型时代下知识图谱构建关键技术

本节将从知识抽取、知识融合和知识推理 3 个关键任务对融合 LLMs 的知识图谱构建技术研究成果进行梳理、剖析、总结和分析。

2.1 知识抽取遇上大语言模型

对于包含文本、图片、音频、视频等非结构化数据的互联网信息, 从文本模态中抽取知识被称为信息抽取^[30]. 信息抽取包括 4 个子任务, 即实体识别、属性抽取、关系抽取和事件抽取^[31], 其中属性抽取可看作实体与属性值间的一种名词性关系, 能够转化为关系抽取。

LLMs 时代下的实体抽取、关系抽取、事件抽取 3 个子任务范式渐渐由判别式过渡到生成式^[32]. 核心思想为给定一段输入文本, 包含 n 个标记词序列 $X = [x_1, x_2, \dots, x_n]$ 、一个提示 P 、目标序列 $Y = [y_1, y_2, \dots, y_m]$, 目标是最大化条件概率, 如下所示:

$$p_{\theta}(Y|X, P) = \prod_{i=1}^m p_{\theta}(y_i|X, P). \quad (1)$$

其中: θ 为 LLMs 中的参数, 允许冻结或训练; 在 X 上附加额外提示 P , 以增强 LLMs 对特定任务的理解能力. 整体任务框架如图 2 所示。

2.1.1 实体抽取

实体抽取, 亦称为命名实体识别 (NER), 旨在从文本中识别出实体的指称项, 并标记类别, 是知识图谱构建和信息抽取的基础. 最早期的实体抽取方法多依赖于规则和统计模型^[33], 随后为提升实体边界定位的精度和复杂文本的识别效果, 如 CNN-BiLSTM-CRF^[34]、ALL-CNN^[35]、BiLSTM-ATT-CRF^[36] 等深度学习模型及其改进模型架构^[37] 被提出. 尽管如此, 面对高准确度要求和复杂情境下, 传统深度学习方法依然面临着一定的局限性, 尤其是在精细化的实体识别和复杂的上下文理解方面, 表现尚未达到理想状态. 伴随 LLMs 的广泛应用, 实体抽取技术迎来了新一轮的创新和突破。

对于实体类型预测任务, Yuan 等^[38] 提出了生成式实体分类范式, 引入了课程学习思想, 依靠课程对实体类型粒度以及异构数据的理解自我调整进度, 从而使得 PLMs 生成高质量的细粒度实体类型; Feng

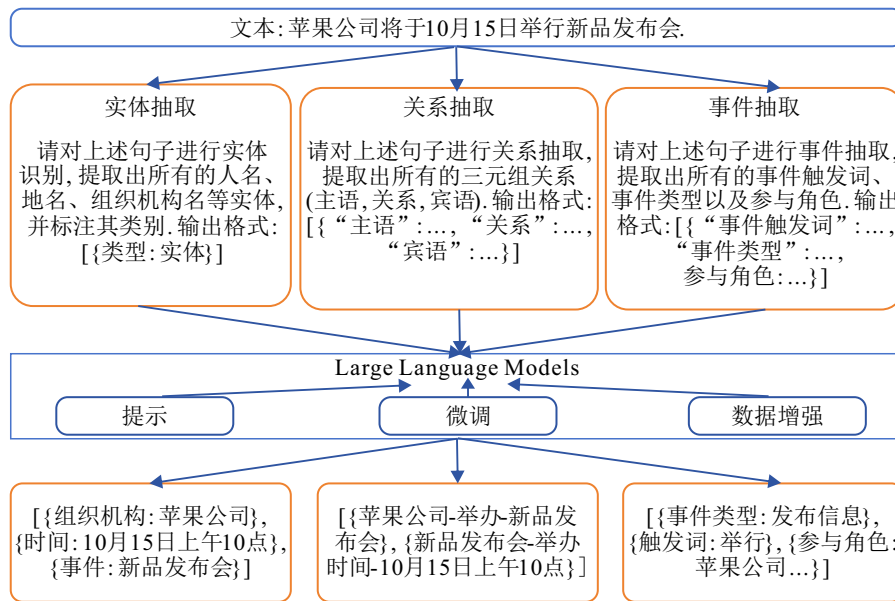


图2 知识抽取任务框架

等^[39]提出了CASENT方法,通过约束搜索生成多种候选实体类型,采用校准方法将预测类型的序列概率转化为置信度分数;此外,GPT-NER框架^[40]以及开放域信息抽取的学生模型^[41]等,均利用语言模型学习丰富的知识,减少了实体类型预测误差。

对于少样本或零样本任务,Zhang等^[42]提出的2INER框架通过引入指令微调和辅助任务,有效增强了实体类型的表征能力;同时,少样本多模态NER结构^[43]、OADA方法^[44]、PromptNER方法^[45]以及基于实体到文本的EnTDA增强方法^[46],进一步优化了少样本和零样本任务的处理。

对于文档级任务,Amalvy等^[47]针对Transformer结构在处理长文本时的局限性,通过检索相关上下文增强了文档级NER能力,并采用指令微调的LLMs生成了上下文数据集。

对于多模态场景,Li等^[48]提出的PGIM框架利用ChatGPT作为隐式知识库,解决了多模态场景下文本预测的知识冗余问题,显著提高了实体预测的效率。

对于跨领域场景,Chen等^[49]基于文本到文本生成式PLMs,提出了CP-NER以调整协作域前缀,解决了跨领域实体类型间可能存在的不匹配问题。

对于低资源场景,Chen等^[50]利用轻量级微调范式;Cui等^[51]通过知识迁移有效提升了实体识别性能;Wang等^[52]提出的多任务指令生成框架,通过引入实体抽取和实体分类辅助任务,进一步增强了实体识别的效果。

对于嵌套和非连续实体的场景,Yan等^[53]将NER任务统一表示为实体跨度序列生成问题,提出了新

的实体表示法;Zhang等^[54]通过去噪增强方法进一步优化了实体识别;Xia等^[55]发现生成式NER结构易产生偏差,从而引入了对比损失来解决这一问题;最新的研究中,DiFiNet方法^[56]从捕捉实体差异和减少非实体噪声两方面提升了嵌套实体边界的敏感度。在ACE04、ACE05和GENIA三个基准数据集上的实验结果表明, F_1 分别达到了88.45%、88.94%和81.87%,为解决嵌套NER任务中的边界不敏感问题提供了新思路。

综上,表1对近4年实体抽取领域的代表性工作进行了详细对比。借助LLMs的生成范式,实体抽取任务在处理复杂任务时取得了显著进展。与此同时,LLMs仍然存在一些局限性,如可解释性不足^[13],模型决策过程难以透明化;其训练数据可能存在偏差,需要通过预处理或数据增强来缓解噪声影响。随着LLMs的进一步应用和优化,其在实体抽取领域的潜力仍然有待深入挖掘和扩展。

2.1.2 关系抽取

关系抽取(RE)旨在识别两个或多个实体间存在的特定语义联系。基于深度学习的方法^[57-58]分为流水线和联合学习,克服了传统特征工程的局限,但是,未有效解决关系多义性和重叠问题。随着LLMs的引入,关系抽取迎来了突破性进展。

针对流水线方式可能导致错误传播问题,Cabot等^[59]将关系三元组表示为文本序列,引入了自回归生成式模型,以端到端的形式有效抽取了200余种关系类型。

针对文档级场景中长距离关系表征困难的问题,Huang等^[60]将其转化为模板生成问题,有效捕捉了

表1 实体抽取代表性工作及其分析

任务	发表时间	名称/作者	数据集	$P/R/F_1/\%$	主要贡献	局限性	难点	LLMs优势
实体类型预测	2022	GET ^[38]	多语言NER	66.20 / 63.52 / 64.83	能够生成未见的开放类型,增加实体多样性;提高类型生成质量	生成时可能忽略正确类型;生成类型选择的阈值可能影响精确度与召回率的平衡		通过预训练和微调能够学习到丰富的语言模式和知识,以预测超出预定义集合的实体类型
	2023	CASENT ^[39]	UFET	53.30 / 49.50 / 51.30	训练过程简易;高效推理;对未见领域和类型的泛化性强	对其他语言的数据表现未知;自回归解码可能降速	无法将实体分配给预定义类型集合外的类型,导致类型扩展误差	
	2023	GPT-NER ^[40]	CoNLL2003	94.73 / 96.97 / 95.85	在低资源和少样本下表现良好;通过自我验证策略有效减少了LLMs的幻觉问题	难区分相似实体类型;生成格式可能与预定义不符		
少样本零样本	2023	2INER ^[42]	CoNLL2003	— / — / 93.93	在少样本或标注成本高昂的场景中表现出色	面对大规模实体类型,示例和候选类型字段易变得冗长,影响模型性能		
	2023	FewMner ^[43]	Twitter2015 Twitter2017	51.24 / 67.20 / 58.14 69.00 / 78.09 / 73.26	利用LLMs的上下文学习能力,有效处理多模态数据,提升了少样本情况下的性能	采用全部数据量,效果弱于微调方法	少样本或零样本任务中,训练数据有限,难以充分学习到实体与关系间的复杂模式	强大泛化能力和扩展能力,能够借助已有的知识模式,扩展到更大的数据规模,对新的实体有效抽取
	2024	OADA ^[44]	CoNLL2003 ACE2005	— / — / 81.47 — / — / 67.08	从顺序无关的角度首次探索数据增强,提出了使用排序指令和创新的OADA-XE损失函数来联合微调预训练语言模型	生成所有可能的实体排列组合在计算上是不现实的;可能无法完全捕捉实体间的复杂嵌套关系		
文档级	2023	Amalvy等 ^[47]	English NER literary dataset	— / — / 92.50	相比于无监督检索方法或引入人工标注数据训练,性能均有所提升	可能引入冗余信息;计算成本增加;LLMs可能会复制训练数据中的偏见,引入偏差	一词多义和歧义词、长距离依赖的问题	深层网络结构利于捕捉语篇内上下文信息和语义特征,以理解文档特征
多模态	2023	PGIM ^[48]	Twitter2015 Twitter2017	79.21 / 79.45 / 79.33 90.86 / 92.01 / 91.43	基于多模态数据生成的辅助知识更加精炼;能够正确预测未标记实体	将图像转换为标题的方法难以完全捕捉图像所有细节	实体类型不匹配问题	整合来自不同模态的信息,同时对实体类型不匹配问题进行校正
跨领域	2023	CP-NER ^[49]	CoNLL2003	— / — / 72.30	无需为每个领域设计模型和微调所有参数;利用多源知识增强目标领域的表现	对特定领域适应性较弱;依赖大量无标注数据	领域间词汇差异大,标注数据稀缺	统一的模型架构可适配多个领域任务,能够跨领域迁移,无需为每个新领域训练独立模型
低资源	2022	LightNER ^[50]	CoNLL2003	92.39 / 93.48 / 92.93	无需针对目标领域构建新模型;通过指导模块来实现跨领域和跨任务的知识迁移	更高资源场景下仍然有改进空间;需要进一步引入KG来增强知识迁移能力	难以处理长尾类型且识别效率较低	利用上下文捕捉实体类型的概率分布,基于少样本和零样本学习能力,生成合理的提示
	2022	Instructionner ^[52]	AirlineTravcl data	— / — / 95.40	利用LLMs,提高泛化能力;辅助任务增强边界定位和类型识别精度	当辅助任务数据量超过主任务时,影响主任务性能		
嵌套和非连续	2023	RerankNER ^[55]	CoNLL2003 OntoNotes5 ACE2004 ACE2005	93.26 / 93.69 / 93.48 90.03 / 91.24 / 90.63 87.64 / 87.61 / 87.63 85.01 / 87.47 / 86.22	对实体边界和类型进行校正,减少候选生成预测偏差	增加额外计算开销;与理论上最佳分数(oracle score)仍然有差距,需要进一步缩小	边界定位错误问题	将现有问题转换为生成式任务,利用LLMs提示学习可提高实体边界定位精度
	2024	DiFiNet ^[56]	ACE2004 ACE2005 GENIA	88.57 / 88.43 / 88.45 88.16 / 88.74 / 88.94 83.01 / 80.80 / 81.87	提出新的边界感知语义差异化和过滤网络,构建了一个新颖的端到端框架	需要遍历所有跨度,带来了较高的计算成本;预测连续实体较好,对于非连续实体识别任务的适用性有限		

跨实体依赖关系,并利用标签语义避免识别 N 元关系,降低了计算复杂度. 2024年发表的 REIC 方法^[61]首次提出了基于强化学习的句子选择模块,在长文档中通过迭代选择关键句子并结合强化学习优化过

程,更好地捕捉了实体对之间的复杂关系.

针对数据中 NULL 示例关系定义归类错误的问题, Wan 等^[62]提出了 GPT-RE 模型,其在演示中添加了特定任务的实体表示和真实样本诱导的逻辑推

理,以弥补 LLMs 与全监督模型的性能差距.

针对无监督关系抽取任务难以挖掘出潜在的、未知的关系类型问题, Wang 等^[63]提出了对比学习策略; Zhang 等^[64]提出了 LLMaAA 方法,探索 LLMs 生成训练数据的可能性,提升了关系类型的预测准确性.

针对零样本/少样本关系抽取中大量预定义类型的关系类型干扰 LLMs 识别未定义类型的问题, Li 等^[65]引入了思维链来改进 RE 框架; Zhang 等^[66]提出了 QA4RE 框架,将 RE 问题与问题解答(QA)相结合; Xu 等^[67]基于 GPT-3.5 进一步提出了任务相关指令和模式约束数据生成.

表 2 详尽展示了近 4 年相关代表性方法的贡献和局限性,揭示了 LLMs 克服深度学习时代面临的难题,并在常用数据集上实现了性能的显著提升.得益于 LLMs 在语义理解和上下文感知方面的进步, RE 任务获得新的推动力,但是同时也面临着新的挑战,特别是在实体间弱相关性问题上.此外,其黑盒特性易造成生成结果的可解释性较差,这可能会导致特

定领域关系抽取的性能下降.在资源有限的情况下, LLMs 缺乏对于领域的深层理解.

2.1.3 事件抽取

事件作为现实世界中的动态过程,具备时间、地点以及参与者等固有属性,其发生于特定动作的执行或系统状态的转变.事件抽取(EE)旨在精准识别并提取用户关注的事件信息,并以结构化的形式呈现.此过程不仅涉及对文本中事件要素的识别,还包括对事件类型、属性以及参与者的抽取.通过该过程,能够系统理解和描述事件的结构和内涵,为文本挖掘、知识图谱构建等方向提供有力支持.如今,随着 LLMs 的突破,事件抽取已进入一个更先进的发展阶段.

在事件检测(ED)中, LLMs 与小语言模型(SLM)的结合^[68],充分利用提示策略,解决 SLM 难以处理的样本,对于构建事件抽取器有较大优势.面对事件类型多样和标注数据稀缺的挑战, Cai 等^[69]提出了自动化数据生成方式,构建了涵盖 3 000 多种事件类型的数据集,微调后的 LLaMA-2-7B 模型在零样本

表 2 关系抽取代表性工作及其分析

任务	发表时间	名称/作者	数据集	$P/R/F_1/\%$	主要贡献	局限	难点	LLMs 优势
传统 RE	2021	REBEL ^[59]	CONLL04	75.59 / 75.12 / 75.35	自适应新领域或任务;模型结构简单,且易于高效训练	参数量较大,需要较多的计算资源;推广到关系分类,可能需要额外微调	传统关系抽取依赖于 NLP 工具,引入噪声易导致错误传递,影响抽取性能	泛化能力: LLMs 在多项 NLP 任务上表现出色,有助于减少因使用性能较差的工具产生的误差
			DocRED	45.89 / 48.37 / 47.10				
			NYT	91.71 / 92.21 / 91.96				
			ADE	81.45 / 83.07 / 82.21				
文档级场景	2021	Huang 等 ^[60]	MUC-4	68.55 / 49.90 / 57.76	利用生成模型的自回归特性,有效地捕捉实体间的依赖关系,避免传统序列标注方法的局限性	生成模型,参数数量较多,训练和推理过程需要较高的计算资源	文本中可能存在隐式关系信息,缺乏明确的标识,增加了抽取的难度	隐式关系识别: LLMs 通过深度上下文理解,挖掘文本中隐含的浅层关系,即使得缺乏明确标识或直接表述,也能为 RE 任务提供提示
			SciREX	17.11 / 13.56 / 14.47				
无监督 RE	2024	REIC ^[61]	CodRED	— / — / 65.02	提出了基于强化学习的句子选择模块,通过最大化关系预测分数作为奖励,有效地选择了包含关系证据的句子	引入网络增加了模型的运行时间,训练时间和推理时间均成倍增加		
			NULL 示例 RE	2023	GPT-RE ^[62]	SemEval	— / — / 91.90	显著降低 LLMs, 错误地将 NULL 示例归类为其他预定义标签的可能性
TACRED	— / — / 72.14							
SciERC	— / — / 69.00							
ACE05	— / — / 68.73							
无监督 RE	2023	LLMaAA ^[64]	OntoNotes4.0	72.66 / 75.49 / 74.00	通过主动学习策略,能够高效利用未标注数据,并优化训练过程	面对复杂任务,若 LLMs 的零样本推理错误,则需要引入演示,这可能会增加成本	缺乏标注数据,难以处理实体歧义、数据不均衡以及模型的可解释性问题	通过强大的语义理解和推理能力,可以在没有人工标注的情况下,自动从大量文本中挖掘潜在关系
			CoNLL03	82.91 / 82.78 / 82.84				
零样本/少样本 RE	2023	Li 等 ^[65]	NYT	55.70 / 78.30 / 65.10	增加中间推理步骤提升 LLMs 性能;其在处理重叠关系抽取时表现良好	受预算限制,提示格式种类有限,存在更有效的提示设计	同一实体对可能存在多重关系,难以抽取全部关系	多重关系识别: LLMs 综合考虑上下文信息和全局语境,准确捕捉复杂关系
			TACRED	47.70 / 78.60 / 59.40				
	RETACRD	56.20 / 67.20 / 61.20	与 QA 任务对齐,提升 LLMs 在 RE 中的性能;减少对高质量模板的依赖,提高对模板变化的鲁棒性	对未经指令微调的 LLMs,可能无法提升性能;少样本条件下,缺乏 LLMs 性能与样本量的关联探讨				
	TACREV	46.00 / 83.60 / 59.40						
SemEval10	41.70 / 45.00 / 43.30							

事件检测中超越了 GPT-3.5-Turbo 等大语言模型,取得了显著的性能提升.这验证了所提出方法为知识图谱构建中事件检测提供了高效解决方案.

在事件论元抽取 (EAE) 中,现有的深度学习方法易忽视不同论元间的相关性.由此, Ma 等^[70]提出了 PAIE 模型,利用对抽取目标的及时调整来发挥 PLMs 的最大优势; Luo 等^[71]基于扩散模型创建了首个端到端的 Zero-shot 事件论据生成提示方法; Wang 等^[72]发现结构化代码能够与语义结构形成互译,因此,将事件论元抽取重构为代码生成问题,以充分利用基于文本和代码混合训练的 LLMs;此外, Zhang 等^[73]引入了 HMPEAE 的新模型,解决了统一类型论据间的语义差异和不同类型论据差异,在 WikiEvents 数据集上 F_1 值达到 72.1%,结果达到了以往研究的最好效果.

然而,低资源情况会限制事件抽取的效果.对此, Zhao 等^[74]提出了基于演示增强的模式引导生成方法,从两方面提升抽取效果:1) 针对事件抽取提出了基于演示的学习范式,将其转换为演示来揭示抽取过程,进而充分利用标注数据;2) 将事件抽取表述为基于模式提示的文本生成任务,从而利用标签语义来推动知识迁移.为降低成本, Ma 等^[75]提出了一种名为 STAR 的数据生成方法,进而发现 LLMs 对于低资源事件抽取的性能提升显著.

伴随研究深入,一些学者逐步将事件抽取视为一种提示驱动的顺序生成问题.如: Lu 等^[76]提出的生成范式 Text2Event, Liu 等^[77]提出的基于动态前缀的生成式事件抽取方法, Pang 等^[78]提出的指南学习框架.最为典型的是 Liu 等^[77]提出的 GTEE-DynPref,旨在解决生成式方法所面临的次优提示和静态事件类型信息的挑战,该方法通过集成上下文信息和类型特定的前缀,以学习上下文特定前缀.

综上,表 3 整理了近期事件抽取的代表性工作,分析得出:

1) LLMs 在事件检测、事件论元提取、低资源事件抽取等子任务中展现出更好的性能;

2) 将事件抽取视为生成问题,对抽取效果有显著提升.然而,尽管 LLMs 拥有丰富的先验知识,其直接应用仍然面临挑战,如必须符合严格的输出格式规范^[79].详细挑战以及未来展望将在第 4 节进行阐释.

2.2 知识融合遇上大语言模型

知识融合是对不同来源的数据、不同表示结构的知识进行融合,是对知识图谱进行补充、去重、更

新的有效手段.由于数据来源非常广泛,质量也会参差不齐,关注重点也不尽相同.因此,不同知识图谱存在多样性和异构性,而知识融合是解决知识图谱异构问题的有效途径,任务示例如图 3 所示. LLMs 在知识融合任务中扮演着重要角色,采用多领域知识整合、信息汇总和提炼、实时知识更新等能力促进信息的互通和创新.根据融合的对象,将从 LLMs 时代下的本体融合和实例融合两方面进行详细评述.

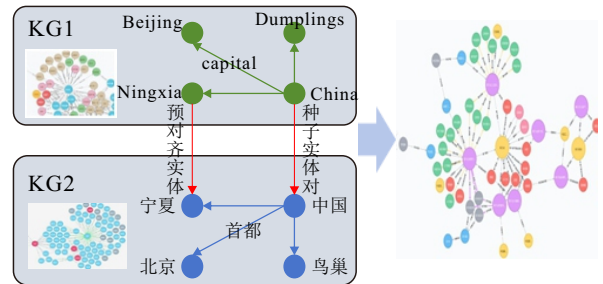


图3 知识融合任务示例

2.2.1 本体融合

本体映射和本体集成是解决本体异构的通用方法^[80].来自不同信息源的异构本体可用两种方式进行融合:1) 本体间建立映射规则,即本体映射;2) 多个本体合成统一本体间交互,即本体集成.最新研究中本体融合主要利用 LLMs 获取本体的向量表示,并将语义相似度计算结果作为本体融合的依据.

He 等^[81]针对传统本体对齐系统性能较低的问题,提出了一种名为 BERTMap 的新型本体对齐系统,通过比较不同知识图谱中的全部文本概念,复杂度达到了平方级别,成本较高;随之, Hertling 等^[82]首先使用预训练的句子编码器,生成匹配的候选句子,然后使用经过校准的基于 Transformer 的交叉编码器生成了最佳候选句; Amir 等^[83]提出了 Truveta Mapper (TM) 框架,利用多任务模型,以零样本、统一式、端到端的方式在多个本体间进行对齐;随后, Wang^[84]发现前人工作中并未考虑运行时的延迟以及端到端框架内不同图结构的问题,进而引入了一种新颖的自监督学习本体匹配框架 LaKERMap,充分利用了概念的上下文和结构信息; Schneider 等^[85]提出了自然语言导向的本体对齐模型,以处理本体的语义和图形结构;此外,事件本体为现实世界中发生的事情提供了一个共享的、正式规范, Guo 等^[86]基于 FrameNet 和 Wikidata 构建了一个事件本体对齐数据集,并提出了一种多视角事件本体对齐方法,利用描述信息和邻接信息以获得丰富的事件本体表示.

现有研究表明,利用 LLMs 的语义理解能力、丰

表3 事件抽取代表性工作及其分析

任务	发表时间	名称/作者	数据集	$P/R/F_1/\%$	主要贡献	局限	难点	LLMs优势
事件检测	2023	Ma等 ^[68]	ACE2005	— / — / 67.80	优化LLMs的调用, 显著降低预算和延迟; LLMs有效纠正SLM的错误预测	样本难度评估方式可能不完善	事件表达多样性涉及不同的句法结构、语义组合和上下文关系, 导致其复杂性较高	全面理解: LLMs凭借其上下文理解能力, 能够有效识别事件描述的句法结构、语义组合、上下文关系, 提升抽取的准确性
	2024	Cai等 ^[69]	ACE2005 M2E2 MEE	46.7 / 26.9 / 34.2 21.2 / 26.1 / 23.4 70.9 / 16.7 / 27.1	设计了一个包含多样化事件类型和定义的自动生成数据集; 发现事件类型数量和定义多样性对于零样本事件检测性能有显著影响	依赖于自动生成数据集, 可能无法完全捕捉现实事件的复杂性; 方法的有效性依赖于详细事件本体和硬负样本		
事件论元抽取	2022	Code4struct ^[70]	ACE2005	— / — / 62.3	将自然语言处理任务转化为代码生成问题, 利用编程语言的结构化特性, 提高模型对于事件结构的理解能力, 有效提高事件论元抽取的性能	对事件检测系统的依赖, 对复杂事件层次的覆盖不足, 以及相对于传统监督学习方法较高的推理成本	需要综合考虑上下文, 进行广泛地文本理解; 事件间可能存在嵌套关系, 需要识别其层级结构	歧义消除: 利用长文本处理能力, LLMs能够有效捕捉事件描述的相关上下文, 增强抽取的鲁棒性
	2024	HMPEAE ^[73]	RAMS WikiEvents	— / — / 58.6 — / — / 72.1	提出了针对类内差异和类间差异的原型网络方法, 以及最优传输方法解决论元和原型的分配问题, 增强了模型对不同类别论元的识别能力	模型对不同类别的子簇数量假设有一限制, 可能某些事件类型的论元抽取不够精准		
低资源场景	2023	DemoSG ^[74]	ACE2005	— / — / 25.5	低资源事件抽取中表现出色, 能够在不需要微调模型参数的情况下适应新领域和事件类型	需要进一步探索减少训练和推理阶段暴露偏差的方法, 以增强鲁棒性	相关领域知识匮乏, 标注数据缺少, 有用信息难以挖掘	提示学习: LLMs基于海量语料训练, 获取丰富语言知识和语境信息, 一定程度上可以进行自我学习
	2023	STAR ^[75]	ACE2005	— / — / 54.56	能够在有限的种子示例下生成高质量的训练数据; 提出一种自我学习机制, 迭代改进生成数据	在某些情况下仍然需要人工干预		
统一为生成方式	2021	Text2Event ^[76]	ACE2005	69.60 / 74.40 / 71.90	增强决策交互和信息共享; 在推理过程中注入事件知识	对复杂事件结构, 可能需要特定优化; 在不同语言或领域中的泛化能力未知	事件的多样性、非结构化描述和跨文档推理需求	其强大的生成能力、上下文推理能力和跨文档整合能力, 具有研究前景
	2022	GTEE-DynPref ^[77]	ACE2005	63.7 / 84.4 / 72.60	有效利用上下文信息与事件类型间的关联; 通过迁移学习提升模型对新事件类型的适应性	在处理特定类型的错误或歧义时面临挑战; 复杂事件结构需要针对性优化		

富的语境知识, 能够进一步提升本体融合的效果. 然而, 本体融合的语法、逻辑、表达能力不匹配、元语异构和概念化异构等问题并不能完全消解^[87], 仍需持续研究.

2.2.2 实例融合

实例融合是指将一个知识图谱内部或多个不同知识图谱间的实体实例、关系实例进行融合的过程, 实体对齐 (EA) 是其主要手段. 任务演示如图 3 所示. 传统的实体对齐方法其有效性易受到有限的知识图谱数据和表示学习能力的限制. 融合 LLMs 的实体对齐技术通常采用基于语义匹配的方法, 将知识图

谱中的名称、属性、描述等多维度的实例压缩为低维向量, 计算实例向量间的相似度以判断两个实例间的语义关系.

在跨知识图谱的实体对齐中, Zhao 等^[88] 提出了两种基于 PLMs 的实体对齐器, 将序列间的关联概率表征为实体间的相似性; Jiang 等^[89] 提出了框架 ChatEA, 将知识图谱结构翻译为 LLMs 可理解的格式, 创新地结合 LLMs 来改进实体对齐.

在跨语言实体对齐中, Yang 等^[90] 探究了如何应用深度学习中的图卷积网络结合实体的多方面信息学习实体嵌入, 然而, 结果显示性能仅有 87.1; 而 Tang

等^[91]采用了两种预训练多语言 BERT 模型来弥补跨语言差距; Zhang 等^[92]构建了用于跨语言实体对齐的新型关系感知图神经网络; Jiang 等^[93]提出了无监督跨语言实体对齐方法, 结果达到了 96.6, 比 Yang 等^[90]的方法提升了 9.5.

在多语境实体对齐中, 多数方法缺乏实体与关系间的全局交互, 造成关键语义信息丢失. 对此, Xin 等^[94]提出了一种知情多语境实体对齐模型, 以全面捕捉关系、路径和邻域上下文, 同时设计了推理方法, 以通过嵌入相似度和实体关系交互来估计对齐概率.

在多模态信息的实体对齐中, Zhu 等^[95]提出了 PathFusion, 通过构建连接实体与模态节点的路径来

表示多种模态以简化配准过程, 以路径为信息载体, 有效融合了来自不同模态的信息; Huang 等^[96]提出了 PMF 策略, 通过为模态分配对齐相关性分数, 逐步对齐不相关的特征, 采用对比学习解决多个模态间的一致性问题的.

表 4 对比了近期实体对齐的典型方法. 进一步分析发现, 利用 LLMs 识别不同语言中描述相同实体的等价关系, 归纳实体间的语义模式和关系, 从而获取更丰富的实体特征, 以减少人工成本, 对实体对齐的研究产生巨大影响. 然而, LLMs 的局限性也较为显著, 即对大量的计算资源进行训练和推理, 增加了实体对齐的复杂度.

表4 实体对齐典型方法及其分析

任务	发表时间	名称/作者	数据集	H@1	主要贡献	局限	难点	LLMs优势
跨知识图谱的实体对齐	2023	Zhao等 ^[88]	DBP15K	95.0	通过文本蕴含框架统一建模关系和属性; 考虑实体间直接交互	重排阶段计算成本过高; 候选选择可能在某些极端情况下受限	知识图谱异构性, 实体的描述和属性在不同图谱中的表达方式不一致, 增加了对齐的复杂性	通过上下文推理来理解不同知识图谱中的实体语义, 进而帮助识别与对齐相同的实体
	2024	ChatEA ^[89]	DBP15K	99.0	提出了将KG结构转换为LLMs可理解的格式, 采用两阶段策略, 利用LLMs在对话格式中的多步推理能力, 增强准确性的同时保持效率	在时间敏感或资源受限的环境中, 由于现有 LLMs 的推理速度存在固有限制, 其性能受限		
跨语言实体对齐	2019	Yang等 ^[90]	DBP15K	87.1	通过两种策略整合GCN和BERT模型, 提高对齐准确性, 是深度学习时期较优方式	受限于图结构和特征选择; 处理大规模知识图谱所需计算资源庞大	某些语言中的实体或概念可能没有直接对应的翻译或等价物, 造成对齐困难	结合上下文信息, 不仅基于单一的翻译, 还能通过对不同语言文本的整体理解来进行准确的实体对齐
	2023	Jiang等 ^[93]	DBP15K	96.6	减少对标注数据的依赖; 能够适应不同的数据规模, 保持较高的对齐精度	特殊类型的实体对齐(如缩写、节目名称或特殊字符), 可能需要额外预处理		
多语境实体对齐	2022	IMEA ^[94]	EN-DN-15K	63.9	推理过程增强嵌入学习, 提升性能; 实现模型迭代优化	处理大规模知识图谱可能造成推理的复杂度增加	语境可能包括时间、地点、事件等多种因素, 难以在多变的语境下准确对齐实体	LLMs可以学习不同语境下的实体表现, 做到更高效的跨语境对齐
多模态实体对齐	2023	PathFusion ^[95]	FB15K-DB15K	87.7	迭代融合来自不同模态的信息, 统一建模不同模态	计算复杂度较高; 要求两知识图谱具有近似的模态分布	不同形式的数据在表达同一实体时的方式有很大不同	通过对信息的联合建模, 生成更丰富的语义表达, 进而有效地进行多模态实体对齐
	2024	PMF ^[96]	DBP15K	87.9	通过为模态分配对齐相关性分数, 逐步冻结对齐不相关的特征, 同时整合有益的多模态信息	在跨知识图谱中识别相似特征的有用性时, 有时可靠性不一致		

2.3 知识推理遇上大语言模型

知识推理的目标是通过推理的方式揭示出已有知识中的潜在信息. 现实世界中的数据量是无比庞大的, 知识图谱明显缺乏完整性, 而对于知识图谱中某些缺失的知识难以直接进行填补. 因此, 通过已有知识推断出未知知识是重要途径, 是知识推理的核

心.

传统的知识推理可分为命题逻辑、描述逻辑、一阶逻辑以及高阶逻辑. Chen 等^[97]提出了 KGDL 方法, 联合 TransE 与 LSTM 模型将实体描述的句子嵌入与三元组编码为实体描述, 从而实现知识推理, 这种深度学习推理本质上是一种映射, 并不具备推理

性, 缺乏“归纳推广”的能力, 难以实现更高维度的知识推理; 随后, Zhou 等^[98]提出的 L2M 将语言模型迈向了更复杂的推理; 思维链 (CoT) 的开山之作^[99]指出 LLMs 的常识推理能力赶超人类, 数学逻辑推理能力显著提升, 具备更佳的可解释性和可信性, 进而 LLMs 的推理能力得到了更加深入的研究^[100-101].

最早利用 LLMs 处理文本问题并指导 KG 中的推理过程出现在 Yasunaga 等^[102]提出的 QA-GNN 模型, 通过弥合文本与结构信息间的差距, 从而为推理过程提供可解释性; 面对复杂的知识图谱推理任务, Choudhary 等^[103]提出了 LARK 方法, 利用 LLMs 引导逻辑推理, 实现了 KG 的推理任务; 同样地, Wang 等^[104]将结构推理和语言模型统一在一个学习框架中, 采用 LLMs 学习上下文表征来初始化结构空间, 沿着语义表征空间对 KG 执行逐步推理从而确定答案. 为缓解 KG 中关系的不确定以及不同关系可能包含不同语义的局限, 2024 年最新的工作^[105]在多链路查询方面进行了改进, 在两个基准数据集上的实验验证了该方法超越了目前最先进的方法.

通过思维链技术, LLMs 具备通用且灵活的推理能力, 对更加复杂的 KG 逻辑推理任务有较明显的助力. 然而, 在多模态、多语言、垂直领域场景中, 逻辑歧义、专业知识缺乏均可能会导致错误的提示^[106]. 因此, 结合 LLMs 的知识推理研究是一个难点问题.

3 大语言模型时代下的知识图谱应用

知识图谱的出现显著提升了网络智能化水平, 提供了高效的方式来表达、组织、管理和利用海量异构的动态数据, 使得信息处理更贴近人类思维. 通过构建知识图谱, 可以更好地挖掘和利用网络中的信息, 为智能系统的发展奠定基础. LLMs 的引入进一步推动了知识图谱的应用发展^[107-108]. 本节将总结融合 LLMs 的知识图谱的代表性应用, 重点综述目前应用最广的知识问答和检索增强生成.

3.1 知识问答

知识问答是知识图谱的核心应用之一, 旨在通过解析用户的自然语言问题, 准确识别其询问意图并检索相关信息, 最终生成恰当的答案. 如图 4 所示, LLMs 作为实体关系提取器和答案推理器, 与知识图谱相融合, 搭建了知识问答框架.

LLMs 作为实体关系提取器: LLMs 在语言理解方面具有出色的能力, 利用 LLMs 识别自然语言问题中提及的实体和关系, 并在 KG 中检索相关事实, 是一个具有前景的应用方向. 如 Yasunaga 等^[102]的 QA-GNN 方法、Zhang 等^[109]提出的基于 LLMs 的路

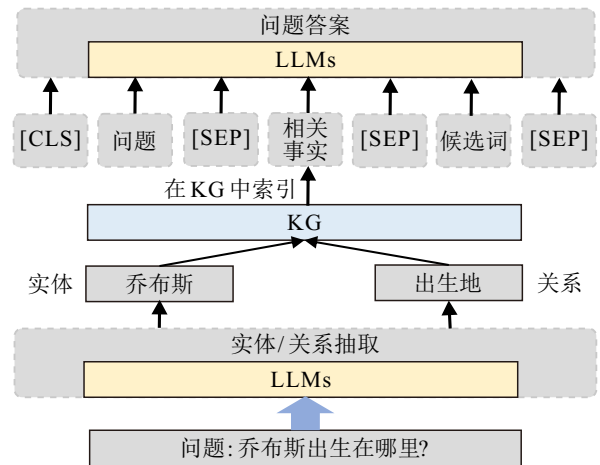


图4 融合 LLMs 和 KG 的知识问答框架

径检索器以及 Hu 等^[110]采用 LLMs 检测实体和关系在知识图谱中查询答案等, 均表明利用 LLMs 进行实体链接和关系检测已显著提高了问答系统的性能.

LLMs 作为答案推理器: LLMs 具备直接生成答案的能力, 在知识问答系统中能够对检索到的事实进行推理并直接生成理想答案. Xu 等^[111]和 Yan 等^[112]利用外部知识库提取相关事实, 解决了搜索答案时推理路径不一致问题. 面对多跳问答任务和复杂问答任务, 具有代表性的方法有 UniKGQA 模型^[113]和 RELMKG 模型^[114]. 研究发现, LLMs 作为答案推理器, 在复杂知识问答结果的准确性上有了很大的进展.

在知识问答应用中, LLMs 利用语言理解和生成能力, 对简单问题直接检索生成答案, 对复杂问题采用提示工程, 结合 KG 中的关键信息, 进行推理从而生成结果, 具有出色表现. 总体而言, 融合 LLMs 的 KG 在知识问答中的应用展现了强大的语言理解和生成能力, 但是仍然需要解决与多轮交互、一致性和生成内容安全性相关的挑战, 以确保其广泛应用的可行性和安全性^[18].

3.2 检索增强生成

大语言模型的一大优势是能够从无标注数据中学习知识, 并将这些事实以参数形式进行存储, 然而, 其存在下述问题: 1) 可能过时的知识; 2) 无法追踪的推理过程; 3) 缺乏现实的知识存储结构; 4) 易产生“幻觉”; 5) 缺乏可解释性. 上述问题均易导致其在知识检索和推理的过程中产生偏差, 降低结果的精确性. 针对上述问题, 将 KG 应用于 LLMs 中, 成为一种很有前景的解决方案.

当前比较具有代表性的方法为检索增强生成 (RAG), 旨在通过语义相似度计算从外部知识库中检索到相关文本块, 并将其与输入信息组合并反馈给

LLMs, 以增强 LLMs^[115]. 通过引入外部知识, RAG 系统可有效减少生成与事实不一致的内容, 提高生成内容的可信度, 解决幻觉^[116-117]、知识时效性^[118]等问题. 经多方验证的 KG 能够提供更精确的信息^[119], 将其与 LLMs 集成, 能够显著增强 RAG 过程. 这种协作方式有助于对齐现实世界的知识和文本向量空间. KG 可作为检索数据源, 不仅能够提高 LLMs 生成信息的质量, 也能有效提升索引构建的质量^[120]. 具体而言, 通过 KG 来构建文档的层次结构, 以保持一致性. 由 KG 构建的索引具有两个优点: 1) 能够描述不同概念与实体间的关系, 有效减少模型产生“幻觉”的可能性; 2) 将语义检索过程转换为适应 LLMs 理解的指令, 提高知识检索的准确性, 促进 LLMs 生成上下文一致的内容, 从而增强 RAG 的整体性能. 如 Wang 等^[121]引入 KG 在多个文档间构建索引, 以挖掘文档内容与结构间的逻辑关系, 有效解决了多文档背景下知识检索和推理的问题. 因此, 将 KG 集成到 RAG 系统具有较大的研究潜力. 通过利用知识图谱中结构化和相互连通的数据, 能够显著提高现有 RAG 系统的推理能力^[121].

4 挑战及未来展望

4.1 挑战

知识图谱作为重要的知识表征手段, 自从 LLMs 涌现后与其进行了更深入地融合研究. 然而, 在展现出巨大潜力和应用价值的同时, 仍然面临诸多挑战.

1) 知识更新难.

从知识表示角度看, LLMs 和 KG 本质上均为知识载体, 面临相似的时效性和更新挑战. 当前, 剑桥大学团队指出现有技术仍然难以处理 50 万条规模的真实世界知识, 正在研究采用知识编辑技术突破这一瓶颈. 在 ICLR 2025 的杰出论文中, Fang 等^[122]针对大语言模型存在的知识遗忘、文本生成质量下降等问题, 提出了 AlphaEdit 方法, 为解决知识更新中的模型稳定性问题提供了新思路. 然而, 面对训练过程中的海量数据, 知识更新过程中存在明显的时间差, 幻觉、过时、领域信息缺少等问题, 均可能影响知识图谱构建质量. 因此, 确保知识图谱持续反映最新的知识状态具有挑战性.

2) 知识冲突.

针对知识图谱构建面临多源数据融合引发的知识冲突问题, 目前, 最新研究从不同角度提出了创新性的解决方案, 如由香港大学、北京智源研究院、厦门大学和深圳中文大学联合提出的 Micro-Act 框架^[123], 将复杂知识分解为细粒度的推理步骤, 在解决

时序冲突和语义冲突方面表现出色. 然而, 知识冲突的复杂性仍然需要从不同维度进一步探索, 如: ① 事实一致性: 新知识应与现实世界相符, 避免引入虚假或过时的信息; ② 逻辑一致性: KG 中的实体及其关系应遵循逻辑规则, 避免出现推理错误; ③ 语义一致性: 新知识应在概念上与已有知识保持一致, 避免引入歧义. 尽管 LLMs 在处理非结构化文本方面表现卓越, 但是在处理结构化知识时, 仍然面临上述问题.

3) 缺乏可解释性.

LLMs 产生的知识可能缺乏解释性, 降低了 KG 的可信度. 其原因在于 LLMs 属于黑盒模型, 推理过程往往是不透明的.

在将 LLMs 嵌入 KG 构建的过程中, 可能因其黑盒特性而导致知识缺乏可解释性, 降低 KG 的可信度和准确性. 具体而言, LLMs 的推理过程不透明, 使得用户难以清晰理解知识图谱中实体、关系以及推理的逻辑. 此外, 研究指出 LLMs 易出现“谄媚”行为^[124], 即为迎合用户偏好而牺牲知识的真实性, 如模仿用户的错误或提供有偏见的反馈. 尽管基于人类反馈的强化学习^[125]尝试缓解此类问题, 但是, 其高昂的标注成本和人类标注者的主观性仍然制约着 KG 知识的解释性提升.

4) 多语言多模态场景的 KG 构建.

传统知识图谱通常基于纯文本和图结构构建, 而现实世界中的知识多源于图像、音频、视频等非文本模态数据, 如何有效表征和融合这些数据成为知识图谱构建研究的重要挑战^[126]. 多模态大模型^[127]的兴起为模态对齐提供了潜在方案, 但是, 如何有效弥合其与知识图谱间的差距仍是亟待解决的问题. 此外, 不同文化和语言背景下的知识可为知识图谱提供更丰富的信息, 但是, 跨语言知识表达的差异也给知识的对齐和融合带来了新的难题.

5) 垂直领域 KG 构建.

从应用角度而言, 通用 KG 构建和垂直领域 KG 构建在数据来源、知识粒度、建模方法以及与 LLMs 的融合方式方面存在显著差异, 如表 5 所示. 这些差异化的融合路径表明, 垂直领域 KG 构建在未来交叉研究领域仍是一个挑战.

综上所述, 大语言模型时代下的知识图谱构建和应用仍然面临诸多挑战, 在工业界和学术界的研究和实践中需要持续探索, 以解决这些挑战.

4.2 未来展望

剖析上述挑战, 本文展望未来, 认为以下 7 个研究方向具有较大潜力和价值.

表5 通用 KG 与垂直领域 KG 及其与 LLMs 融合差异

对比维度	通用KG构建	垂直领域KG构建	与LLMs融合差异
数据来源	开放网络文本(如维基百科、新闻、论坛)	行业专有数据(如病历、地质报告、财报、油藏数据)	通用KG偏向大规模非结构化数据挖掘; 领域KG需结构化+语义理解混合处理
术语体系	通用词汇、跨领域术语	领域术语、行业标准、上下文相关概念	领域KG需领域术语消歧和归一化, 依赖术语词典或知识本体辅助LLMs
知识结构	面向事实三元组(如实体-关系-实体), 以广覆盖为目标	强结构化、包含属性、类型、层次分类、事件演化等复杂结构	领域LLMs需具备图结构感知能力, 支持层次/因果建模, 融合更依赖结构先验
实体类型和范围	实体数量庞大、类型繁多、泛化性强	实体种类少但是要求精细区分(如症状VS疾病, 油藏VS储层)	领域LLMs需进行高精度NER和上下位概念识别, 往往需要提示/微调
构建方式	大规模自动抽取为主, 容忍一定错误率	质量优先, 需人工校验与规则支持	通用KG侧重抽取效率, 垂直KG更重视精度和可控性
KG应用	QA系统、搜索引擎、推荐系统	决策支持、故障诊断、科研辅助	领域LLMs生成需具备推理能力、可解释性、遵循行业规则
LLMs使用方式	Prompt工程+零样本抽取; 通用模型迁移能力较强	微调或领域适配(如LoRA、指令微调); 需要更具体更强结构感知	领域对LLMs参数量控制、性能稳定性、可解释性提出更高要求
示例	Freebase ^[23] 、Knowledge Vault ^[25] 等	医学图谱UMLS ^[28] 、金融FinDKG ^[129] 等	—

1) 端到端的 KG 构建.

如图 5 所示, 当前研究^[130] 尝试利用 LLMs 实现端到端的 KG 构建, 这一方法不同于传统的多阶段管道模式. 目前, Melnyk 等^[131] 提出的端到端多阶段提取器, 通过语言模型生成图节点, 随后构建边关系, 实现了从文本到 KG 的高效转换. LLMs 具备隐式知识编码能力^[28], 使得直接从模型中提取知识构建 KG 成为可能, 尤其是在常识性 KG 构建中表现突出^[29]. 未来发展将向其他类型的通用知识图谱和领域知识图谱的构建进行扩增. 此外, LLMs 可用于数据标注和增强^[132], 提升 KG 构建效率, 特别适用于数据敏感或隔离的场景. 尽管面临幻觉、知识准确性、知识更新的挑战, 端到端的 KG 构建方法仍然在通用领域和特定领域知识图谱构建中展现出广阔前景.

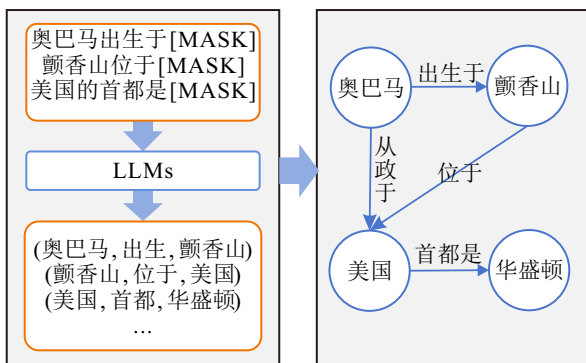


图5 端到端 KG 构建框架示例

2) LLMs 理解 KG 结构.

当前, LLMs 主要以非结构化文本为训练语料, 缺乏对图结构数据的有效编码能力, 限制其对知识图谱中显式知识和隐式语义的理解. 此外, 知识图谱规模庞大, 难以将整个图谱线性化为适配于 LLMs 的输入指令, 导致图谱结构以及关键语义信息的损失, 给 LLMs 的图数据推理造成困难. 另一方面, 利

用 LLMs 优化知识图谱结构具有显著潜力^[133]: ① 扩展实体描述; ② 提炼图谱的结构信息; ③ 深层理解实体关系. 因此, 开发具备直接理解知识图谱结构能力的 LLMs^[134], 对于推动图谱数据的深度语义解析具有重要意义.

3) KG 增强 LLMs.

Pan 等^[18] 对 KG 增强 LLMs 的预训练能力、推理能力和可解释性的经典方法进行了较为全面的总结. KG 通过其结构化的存储机制, 提供对知识来源、关系和更新过程的透明性和可追溯性, 令其能够有效作为 LLMs 的补充训练资源, 增强 LLMs 对复杂关系和深层次知识的解析能力, 进而在精确性、透明度、专业性、一致性等重要方面弥补 LLMs 的潜在缺陷. 未来探索 KG 对 LLMs 的知识推理、语义理解和文本生成等方面的增强机制将是一个潜在的研究方向.

4) KG 与 LLMs 协同.

已有研究^[135] 提出了 GraphEval 评估框架, 旨在衡量 LLMs 与 KG 间的关联度. 尽管二者具有明显的互补优势, 但是, 现有工作对其协同机制的探索较为有限.

利用 KG 的结构化知识表示信息和 LLMs 的上下文学习能力, 构成高效协同系统. 然而, 为充分发挥该协同效应, 需要引入多模态学习、图神经网络、智能体等前沿技术^[18]. 具体应用中, 协同系统既可通过 KG 以知识驱动的方式对潜在数据进行检索, 又能采用 LLMs 以文本驱动的推理对可能出现的新数据进行生成. 二者相互验证, 增强结果的可解释性. 未来深入融合 KG 与 LLMs, 有望实现技术突破, 并广泛应用于多种下游任务.

5) 多模态 KG 构建.

多模态知识图谱通过对齐融合文本、图像、视频、声音等异构数据, 具备更丰富的语义表示能力, 是实现人机智能的重要路径. 目前, 构建方法主要采用两种策略: ① 从图像 (images) 到符号 (symbols), 采用知识图谱中的结构符号 (如概念、实体) 以标记图像; ② 从符号到图像, 即符号接地 (symbol grounding), 将 KG 中的符号信息链接到对应的多模态数据^[136]. 近年来, 多模态大语言模型的发展为此提供强大支持, 提升了多模态知识图谱构建的准确性和效率, 并推动了智能问答、内容生成以及推荐系统等场景的深入应用. 此外, 多模态知识图谱也面临动态更新、语义理解和推理、跨领域融合、开放式构建以及可解释性等关键挑战^[137].

6) 垂直领域 KG 构建.

特定领域的知识图谱包括结构化数据、半结构化数据以及大量非结构化文本、图像、音视频等, 如何从这些日益多样化的异构数据中提取知识构建领域 KG, 是未来仍然需要考虑的基本问题. 尽管在医学^[128]和金融领域^[129] KG 构建小有进展, 然而, 垂直领域 KG 要与实际业务场景和应用需求紧密结合, 距可扩展、可持续和可解释的领域 KG 还有一定距离. 因此, 未来 KG 构建仍然需要考虑与实际业务场景和应用需求紧密结合, 持续探索特定领域的语义规则、本体模型和算法以及高效利用融合领域知识的 LLMs 的概括推理能力.

7) KG 和具身智能.

LLMs 在智能“涌现”的巨大突破, 具身智能这一前沿领域逐渐受到广泛关注. 该范式旨在研究于环境中具有实体的智能体如何通过与环境的交互来取得认知能力, 学习并掌握新技能新知识. 将 LLMs 作为智能体的核心决策模块, 能够利用其强大的场景理解和推理能力辅助智能体认知世界、执行决策, 如 Voxposer^[138] 和 PaLM-E^[139] 已在机器人路径规划以及多模态推理方面取得进展. 然而, LLMs 在具身智能领域还未达到非具身智能领域的性能表现, 可能原因有两个: ① 基座模型基于概率驱动的黑盒式参数化知识, 更新慢且难以动态维护; ② 训练过程中缺乏现实世界链接, 依赖高质量提示工程和去偏学习^[140]. 而基于场景驱动的多模态知识图谱融合符号知识和参数知识的优势, 在场景信息、可解释性方面与 LLMs 相辅相成^[141], 辅助智能体理解环境并做出合理决策, 有望成为具身智能应用的重要支撑.

5 结论

本文聚焦于大语言模型时代知识图谱构建技术及其应用. 首先, 概述了大语言模型与知识图谱融合的统一框架; 然后, 围绕知识图谱生命周期, 深入剖析了大语言模型在知识抽取、知识融合以及知识推理关键技术环节的研究进展; 接着, 从知识问答系统和检索增强生成两个典型应用场景出发, 系统地梳理了知识图谱在新范式下的实际应用价值; 最后, 全面审视了当前研究所面临的挑战, 并从多个视角探讨了未来的发展方向.

当前, 基于大语言模型的知识图谱构建正处于快速发展阶段, 相关平台和应用持续涌现, 展现出强劲的研究和落地活力. 随着知识图谱在提升大语言模型可解释性、增强知识可控性方面的作用日益凸显, 其在迈向通用人工智能过程中的战略地位也愈发重要. 综上, 本文旨在帮助读者系统把握大语言模型与知识图谱构建融合的国内外研究进展, 分析其关键技术和应用趋势, 并为未来研究提供方向指引, 期望为学术界和工业界的相关从业者提供实质性的参考和启发.

参考文献 (References)

- [1] Peng C Y, Xia F, Naseriparsa M, et al. Knowledge graphs: Opportunities and challenges[J]. *Artificial Intelligence Review*, 2023, 56(11): 13071-13102.
- [2] Vrandečić D, Krötzsch M. Wikidata: A free collaborative knowledgebase[J]. *Communications of the ACM*, 2014, 57(10): 78-85.
- [3] Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(2): 494-514.
- [4] Deng C Y, Zeng G F, Cai Z P, et al. A survey of knowledge based question answering with deep learning[J]. *Journal on Artificial Intelligence*, 2020, 2(4): 157-166.
- [5] Guo Q Y, Zhuang F Z, Qin C, et al. A survey on knowledge graph-based recommender systems[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(8): 3549-3568.
- [6] Zhong L F, Wu J, Li Q, et al. A comprehensive survey on automatic knowledge graph construction[J]. *ACM Computing Surveys*, 2023, 56(4): 1-62.
- [7] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582-600. (Liu Q, Li Y, Duan H, et al. Knowledge graph construction techniques[J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600.)
- [8] Hogan A, Blomqvist E, Cochez M, et al. Knowledge graphs[J]. *ACM Computing Surveys*, 2021, 54(4): 1-37.

- [9] 王萌, 王昊奋, 李博涵, 等. 新一代知识图谱关键技术综述[J]. *计算机研究与发展*, 2022, 59(9): 1947-1965. (Wang M, Wang H F, Li B H, et al. Survey on key technologies of new generation knowledge graph[J]. *Journal of Computer Research and Development*, 2022, 59(9): 1947-1965.)
- [10] Liu A, Feng B, Xue B, et al. Deepseek-v3 technical report[J/OL]. 2024, arXiv: 2412.19437.
- [11] OpenAI R. GPT-4 technical report[J/OL]. 2023, arXiv: 2303.08774.
- [12] 刘挺. 从 ChatGPT 谈大语言模型及其应用[J]. *语言战略研究*, 2023, 8(5): 14-18. (Liu T, Talking about large language models and applications from ChatGPT[J]. *Chinese Journal of Language Policy and Planning*, 2023, 8(5): 14-18.)
- [13] Zhao W X, Zhou K, Li J, et al. A survey of large language models[J]. *ACM Transactions on Intelligent Systems and Technology*, 2024, 15(3): 1-45.
- [14] Hadi M U, Qureshi R, Shah A, et al. A survey on large language models: Applications, challenges, limitations, and practical usage[EB/OL]. (2023-07-10)[2025-08-31]. <https://doi.org/10.36227/techrxiv.23589741.v1>.
- [15] Kaddour J, Harris J, Mozes M, et al. Challenges and applications of large language models[J]. *IEEE Access*, 2024, 12: 26839-26874.
- [16] Cui Y M, Che W X, Liu T, et al. Revisiting pre-trained models for Chinese natural language processing[C]. *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online, 2020: 657-668.
- [17] Chowdhery A, Narang S, Devlin J, et al. PaLM: Scaling language modeling with pathways[J/OL]. 2022, arXiv: 2204.02311.
- [18] 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.
- [19] 曾泽凡, 胡星辰, 成清, 等. 基于预训练语言模型的知识图谱研究综述[J]. *计算机科学*, 2025, 52(1): 1-33. (Zeng Z F, Hu X C, Cheng Q, et al. Survey of research on knowledge graph based on pre-trained language models[J]. *Computer Science*, 2025, 52(1): 1-33.)
- [20] 李紫宣, 白龙, 任韦澄, 等. 代码大语言模型赋能的知识图谱关键技术综述[J]. *大数据*, 2025, 11(2): 19-28. (Li Z X, Bai L, Ren W C, et al. Review of key technologies in knowledge graphs powered by code large language models[J]. *Big Data Research*, 2025, 11(2): 19-28.)
- [21] 杨观赐, 许彪, 罗可欣, 等. 知识图谱技术综述: 构建、推理及典型应用[J]. *贵州大学学报: 自然科学版*, 2025, 42(2): 1-10. (Yang G C, Xu B, Luo K X, et al. A review of knowledge graph techniques: Construction, reasoning and typical applications[J]. *Journal of Guizhou University: Natural Sciences*, 2025, 42(2): 1-10.)
- [22] 陈伯谦, 王坚. 融合概念和属性信息的领域知识图谱补充方法[J]. *控制与决策*, 2024, 39(7): 2325-2333. (Chen B Q, Wang J. Domain knowledge graph completion method incorporating concept and attribute information[J]. *Control and Decision*, 2024, 39(7): 2325-2333.)
- [23] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge[C]. *Proceedings of the ACM SIGMOD International Conference on Management of Data*. Vancouver, 2008: 1247-1250.
- [24] Bizer C, Lehmann J, Kobilarov G, et al. DBpedia — A crystallization point for the web of data[J]. *Journal of Web Semantics*, 2009, 7(3): 154-165.
- [25] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion[C]. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, 2014: 601-610.
- [26] Fensel D, Şimşek U, Angele K, et al. Introduction: What is a knowledge graph?[M]. Cham: Springer International Publishing, 2020.
- [27] Liu N F, Gardner M, Belinkov Y, et al. Linguistic knowledge and transferability of contextual representations[C]. *Proceedings of the Conference of the North. Minneapolis*, 2019: 1073-1094.
- [28] Petroni F, Rocktäschel T, Riedel S, et al. Language models as knowledge bases?[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Hong Kong, 2019: 2463-2473.
- [29] Bosselut A, Rashkin H, Sap M, et al. COMET: Commonsense transformers for automatic knowledge graph construction[C]. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, 2019: 4762-4779.
- [30] 黄河燕, 袁长森, 冯冲. 融合实体和上下文信息的篇章关系抽取研究[J]. *自动化学报*, 2024, 50(10): 1953-1962. (Huang H Y, Yuan C S, Feng C. Document-level relation extraction with entity and context information[J]. *Acta Automatica Sinica*, 2024, 50(10): 1953-1962.)
- [31] 王浩畅, 周郴莲, Marius Gabriel Petrescu. 基于深度学习的事件抽取研究综述[J]. *软件学报*, 2023, 34(8): 3905-3923. (Wang H C, Zhou C L, Petrescu M G. Survey on event extraction based on deep learning[J]. *Journal of Software*, 2023, 34(8): 3905-3923.)
- [32] Wang X, Zhou W, Zu C, et al. InstructUIE: Multi-task instruction tuning for unified information extraction[J/OL]. 2023, arXiv: 2304.08085.
- [33] 肖蕾, 陈镇家. 数据驱动的中文实体抽取方法综述[J]. *计算机工程与应用*, 2024, 60(16): 34-48. (Xiao L, Chen Z J. Review of data-driven approaches to Chinese named entity recognition[J]. *Computer Engineering and Applications*, 2024, 60(16): 34-48.)

- [34] Ronran C, Lee S. Effect of character and word features in bidirectional LSTM-CRF for NER[C]. IEEE International Conference on Big Data and Smart Computing. Busan, 2020: 613-616.
- [35] Kong J, Zhang L X, Jiang M, et al. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition[J]. *Journal of Biomedical Informatics*, 2021, 116: 103737.
- [36] Zhong Q, Tang Y. An attention-based BILSTM-CRF for Chinese named entity recognition[C]. Proceedings of the 5th IEEE International Conference on Cloud Computing and Big Data Analytics. Chengdu, 2020: 550-555.
- [37] Peters M E, Ammar W, Bhagavatula C, et al. Semi-supervised sequence tagging with bidirectional language models[C]. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, 2017: 1756-1765.
- [38] Yuan S Y, Yang D Q, Liang J Q, et al. Generative entity typing with curriculum learning[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, 2022: 3061-3073.
- [39] Feng Y L, Pratapa A, Mortensen D. Calibrated seq2seq models for efficient and generalizable ultra-fine entity typing[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 15550-15560.
- [40] Wang S, Sun X, Li X, et al. GPT-NER: Named entity recognition via large language models[J/OL]. 2023, arXiv: 2304.10428.
- [41] Zhou W, Zhang S, Gu Y, et al. Universalner: Targeted distillation from large language models for open named entity recognition[J/OL]. 2023, arXiv: 2308.03279.
- [42] Zhang J S, Liu X K, Lai X Y, et al. 2INER: Instructive and in-context learning on few-shot named entity recognition[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 3940-3951.
- [43] Cai C R, Wang Q L, Liang B, et al. In-context learning for few-shot multimodal named entity recognition[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 2969-2979.
- [44] Wang H M, Cheng L Y, Zhang W X, et al. Order-agnostic data augmentation for few-shot named entity recognition[C]. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, 2024: 7792-7807.
- [45] Ashok D, Lipton Z C. Promptner: Prompting for named entity recognition[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 12492-12507.
- [46] Hu X M, Jiang Y, Liu A W, et al. Entity-to-text based data augmentation for various named entity recognition tasks[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 9072-9087.
- [47] Amalvy A, Labatut V, Dufour R. Learning to rank context for named entity recognition using a synthetic dataset[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 10372-10382.
- [48] Li J Y, Li H, Pan Z, et al. Prompting ChatGPT in MNER: Enhanced multimodal named entity recognition with auxiliary refined knowledge[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 2787-2802.
- [49] Chen X, Li L, Qiao S F, et al. One model for all domains: Collaborative domain-prefix tuning for cross-domain NER[C]. Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Macau, 2023: 5030-5038.
- [50] Chen X, Li L, Deng S, et al. LightNER: A lightweight tuning paradigm for low-resource NER via pluggable prompting[C]. Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, 2022: 2374-2387.
- [51] Cui L Y, Wu Y, Liu J, et al. Template-based named entity recognition using BART[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics. Online, 2021: 1835-1845.
- [52] Wang L, Li R, Yan Y, et al. Instructionner: A multi-task instruction-based generative framework for few-shot ner[J/OL]. 2022, arXiv: 2203.03903.
- [53] Yan H, Gui T, Dai J Q, et al. A unified generative framework for various NER subtasks[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Online, 2021: 5808-5822.
- [54] Zhang S, Shen Y L, Tan Z Q, et al. De-bias for generative extraction in unified NER task[C]. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, 2022: 808-818.
- [55] Xia Y, Zhao Y W, Wu W H, et al. Debiasing generative named entity recognition by calibrating sequence likelihood[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 1137-1148.
- [56] Cai Y X, Liu Q, Gan Y L, et al. DiFiNet: Boundary-aware semantic differentiation and filtration network for nested named entity recognition[C]. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, 2024: 6455-6471.
- [57] Wang Y C, Yu B W, Zhang Y Y, et al. TPLinker: Single-stage joint extraction of entities and relations through token pair linking[C]. Proceedings of the 28th International Conference on Computational Linguistics. Online, 2020: 1572-1582.
- [58] Zhou M J, Ji D H, Li F. Relation extraction in

- dialogues: A deep learning model based on the generality and specialty of dialogue text[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2021, 29: 2015-2026.
- [59] Cabot P L H, Navigli R. REBEL: Relation extraction by end-to-end language generation[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, 2021: 2370-2381.
- [60] Huang K H, Tang S, Peng N Y. Document-level entity-based extraction as template generation[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, 2021: 5257-5269.
- [61] Na B, Jo S, Kim Y, et al. Reward-based input construction for cross-document relation extraction[C]. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Bangkok, 2024: 9254-9270.
- [62] Wan Z, Cheng F, Mao Z Y, et al. GPT-RE: In-context learning for relation extraction using large language models[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 3534-3547.
- [63] Wang Q, Zhou K, Qiao Q, et al. Improving unsupervised relation extraction by augmenting diverse sentence pairs[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 12136-12147.
- [64] Zhang R Y, Li Y Z, Ma Y L, et al. LLMaAA: Making large language models as active annotators[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 13088-13103.
- [65] Li G Z, Wang P, Ke W J. Revisiting large language models as zero-shot relation extractors[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 6877-6892.
- [66] Zhang K, Gutierrez B J, Su Y. Aligning instruction tasks unlocks large language models as zero-shot relation extractors[C]. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. Toronto, 2023: 794-812.
- [67] Xu X, Zhu Y Q, Wang X H, et al. How to unleash the power of large language models for few-shot relation extraction?[C]. *Proceedings of the 4th Workshop on Simple and Efficient Natural Language Processing*. Toronto, 2023: 190-200.
- [68] Ma Y B, Cao Y X, Hong Y, et al. Large language model is not a good few-shot information extractor, but a good reranker for hard samples![C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 10572-10601.
- [69] Cai Z F, Kung P N, Suvarna A, et al. Improving event definition following for zero-shot event detection[C]. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Bangkok, 2024: 2842-2863.
- [70] Ma Y B, Wang Z H, Cao Y X, et al. Prompt for extraction? PAIE: Prompting argument interaction for event argument extraction[C]. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin, 2022: 6759-6774.
- [71] Luo L, Xu Y J. Context-aware prompt for generation-based event argument extraction with diffusion models[C]. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. Birmingham, 2023: 1717-1725.
- [72] Wang X Y, Li S, Ji H. Code4Struct: Code generation for few-shot event structure prediction[C]. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*. Toronto, 2023: 3640-3663.
- [73] Zhang G J, Zhang H, Wang Y J, et al. Hyperspherical multi-prototype with optimal transport for event argument extraction[C]. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*. Bangkok, 2024: 9271-9284.
- [74] Zhao G, Gong X C, Yang X J, et al. DemoSG: Demonstration-enhanced schema-guided generation for low-resource event extraction[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 1805-1816.
- [75] Ma M D, Wang X, Kung P N, et al. STAR: Improving low-resource information extraction by structure-to-text data generation with large language models[C]. *NeurIPS Workshop on Synthetic Data Generation with Generative AI*. New Orleans, 2023: Poster.
- [76] Lu Y J, Lin H Y, Xu J, et al. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction[C]. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Online, 2021: 2795-2806.
- [77] Liu X, Huang H Y, Shi G, et al. Dynamic prefix-tuning for generative template-based event extraction[C]. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin, 2022: 5216-5228.
- [78] Pang C X, Cao Y X, Ding Q, et al. Guideline learning for in-context information extraction[C]. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore, 2023: 15372-15389.
- [79] Yang J F, Jin H Y, Tang R X, et al. Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond[J]. *ACM Transactions on Knowledge Discovery from Data*, 2024, 18(6): 1-32.
- [80] Gutiérrez R, Rampérez V, Paggi H, et al. On the use of information fusion techniques to improve information quality: Taxonomy, opportunities and challenges[J]. *Information Fusion*, 2022, 78: 102-137.
- [81] He Y, Chen J Y, Antonyrajah D, et al. BERTMap: A BERT-based ontology alignment system[C].

- Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, 2022: 5684-5691.
- [82] Hertling S, Portisch J, Paulheim H. KERMIT — A transformer-based approach for knowledge graph matching[J/OL]. 2022, arXiv: 2204.13931.
- [83] Amir M, Baruah M, Eslamialishah M, et al. Truveta mapper: A zero-shot ontology alignment framework[J/OL]. 2023, arXiv: 2301.09767.
- [84] Wang Z. Contextualized structural self-supervised learning for ontology matching[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, 2021: 8754-8769.
- [85] Schneider F, Dash S, Bagchi S, et al. NLFOA: Natural language focused ontology alignment[C]. Proceedings of the 12th Knowledge Capture Conference. Pensacola, 2023: 114-121.
- [86] Guo S R, Wang C H, Chen Y B, et al. EventOA: An event ontology alignment benchmark based on FrameNet and wikidata[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 10038-10052.
- [87] Zhu R B, Luo X K, Ma M, et al. Adaptive graph convolutional network for knowledge graph entity alignment[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Abu Dhabi, 2022: 6011-6021.
- [88] Zhao Y, Wu Y K, Cai X R, et al. From alignment to entailment: A unified textual entailment framework for entity alignment[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 8795-8806.
- [89] Jiang X H, Shen Y H, Shi Z C, et al. Unlocking the power of large language models for entity alignment[C]. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, 2024: 7566-7583.
- [90] Yang H W, Zou Y Y, Shi P, et al. Aligning cross-lingual entities with multi-aspect information[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, 2019: 4431-4441.
- [91] Tang X B, Zhang J, Chen B, et al. BERT-INT: A BERT-based interaction model for knowledge graph alignment[C]. Proceedings of the 29th International Joint Conference on Artificial Intelligence. Yokohama, 2021: 3174-3180.
- [92] Zhang Y H, Wu J Q, Yu K, et al. Independent relation representation with line graph for cross-lingual entity alignment[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(11): 11503-11514.
- [93] Jiang C Y, Qian Y M, Chen L J, et al. Unsupervised deep cross-language entity alignment[C]. Proceedings of the Machine Learning and Knowledge Discovery in Databases: Research Track. Turin, 2023: 3-19.
- [94] Xin K X, Sun Z Q, Hua W, et al. Informed multi-context entity alignment[C]. Proceedings of the 15th ACM International Conference on Web Search and Data Mining. Online, 2022: 1197-1205.
- [95] Zhu B, Liu X, Mao X, et al. Universal multi-modal entity alignment via iteratively fusing modality similarity paths[J/OL]. 2023, arXiv: 2310.05364.
- [96] Huang Y N, Zhang X F, Zhang R C, et al. Progressively modality freezing for multi-modal entity alignment[C]. Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Bangkok, 2024: 3477-3489.
- [97] Chen W R, Hong D P, Zheng C. Learning knowledge graph embedding with entity descriptions based on LSTM networks[C]. IEEE International Symposium on Product Compliance Engineering-Asia. Chongqing, 2020: 1-7.
- [98] Zhou D, Schärli N, Hou L, et al. Least-to-most prompting enables complex reasoning in large language models[J/OL]. 2022, arXiv: 2205.10625.
- [99] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 11548-11559.
- [100] Ho N, Schmid L, Yun S Y. Large language models are reasoning teachers[C]. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Toronto, 2023: 14852-14882.
- [101] Yao S, Yu D, Zhao J, et al. Tree of thoughts: Deliberate problem solving with large language models[J/OL]. 2023, arXiv: 2305.10601.
- [102] Yasunaga M, Ren H Y, Bosselut A, et al. QA-GNN: Reasoning with language models and knowledge graphs for question answering[C]. Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online, 2021: 535-546.
- [103] Choudhary N, Reddy C K. Complex logical reasoning over knowledge graphs using large language models[C]. IEEE International Conference on Knowledge Graph. Abu Dhabi, 2024: 432-439.
- [104] Wang S Y, Wei Z Y, Xu J R, et al. Unifying structure reasoning and language pre-training for complex reasoning tasks[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2023, 32: 1586-1595.
- [105] Liu Y, Cao Y N, Wang S, et al. Generative models for complex logical reasoning over knowledge graphs[C]. Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Merida, 2024: 492-500.
- [106] 车万翔, 窦志成, 冯岩松, 等. 大模型时代的自然语言处理: 挑战、机遇与发展[J]. *中国科学: 信息科学*, 2023, 53(9): 1645-1687.
(Che W X, Dou Z C, Feng Y S, et al. Towards a comprehensive understanding of the impact of large

- language models on natural language processing: Challenges, opportunities and future directions[J]. *Scientia Sinica: Informationis*, 2023, 53(9): 1645-1687.)
- [107] Youssef P, Koraş O, Li M J, et al. Give me the facts! A survey on factual knowledge probing in pre-trained language models[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 15588-15605.
- [108] Chen L H, Razniewski S, Weikum G. Knowledge base completion for long-tail entities[C]. Proceedings of the First Workshop on Matching From Unstructured and Structured Data. Toronto, 2023: 99-108.
- [109] Zhang J, Zhang X K, Yu J F, et al. Subgraph retrieval enhanced model for multi-hop knowledge base question answering[C]. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Dublin, 2022: 5773-5784.
- [110] Hu N, Wu Y K, Qi G L, et al. An empirical study of pre-trained language models in simple knowledge graph question answering[J]. *World Wide Web*, 2023, 26(5): 2855-2886.
- [111] Xu Y C, Zhu C G, Xu R C, et al. Fusing context into knowledge graph for commonsense question answering[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Online, 2021: 1201-1207.
- [112] Yan Y M, Li R M, Wang S R, et al. Large-scale relation learning for question answering over knowledge bases with pre-trained language models[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Online and Punta Cana, 2021: 3653-3660.
- [113] Jiang J, Zhou K, Zhao W X, et al. Unikgqa: Unified retrieval and reasoning for solving multi-hop question answering over knowledge graph[J/OL]. 2022, arXiv: 2212.00959.
- [114] Cao X, Liu Y. ReLMKG: Reasoning with pre-trained language models and knowledge graphs for complex question answering[J]. *Applied Intelligence*, 2023, 53(10): 12032-12046.
- [115] Lyu Y J, Li Z Y, Niu S M, et al. CRUD-RAG: A comprehensive Chinese benchmark for retrieval-augmented generation of large language models[J]. *ACM Transactions on Information Systems*, 2025, 43(2): 1-32.
- [116] Chen Y Y, Fu Q, Yuan Y C, et al. Hallucination detection: Robustly discerning reliable answers in large language models[C]. Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. Birmingham, 2023: 245-255.
- [117] Lyu Y J, Zhu C, Xu T, et al. Faithful abstractive summarization via fact-aware consistency-constrained transformer[C]. Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta, 2022: 1410-1419.
- [118] He H, Zhang H, Roth D. Rethinking with retrieval: Faithful large language model inference[J/OL]. 2023, arXiv: 2301.00303.
- [119] Gaur M, Gunaratna K, Srinivasan V, et al. ISEEQ: Information seeking question generation using dynamic meta-information retrieval and knowledge graphs[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Online, 2022: 10672-10680.
- [120] Yang Y, Rao Y L, Yu M H, et al. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation[J]. *Neural Networks*, 2022, 146: 1-10.
- [121] Wang Y, Lipka N, Rossi R A, et al. Knowledge graph prompting for multi-document question answering[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, 2024: 19206-19214.
- [122] Fang J, Jiang H, Wang K, et al. Alphaedit: Null-space constrained model editing for language models[J/OL]. 2024, arXiv: 2410.02355.
- [123] Huo N, Li J Y, Qin B W, et al. Micro-act: Mitigate knowledge conflict in question answering via actionable self-reasoning[C]. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. Vienna, 2025: 18550-18574.
- [124] Sharma M, Tong M, Korbak T, et al. Towards understanding sycophancy in language models[C]. The 12th International Conference on Learning Representations. Vienna, 2024: 1-46.
- [125] Lambert N, Castricato L, von Werra L, et al. Illustrating reinforcement learning from human feedback (RLHF)[EB/OL]. (2022-12-09)[2025-08-31]. <https://huggingface.co/blog/rlhf>.
- [126] Chen Z, Huang Y F, Chen J Y, et al. LaKo: Knowledge-driven visual question answering via late knowledge-to-text injection[C]. Proceedings of the 11th International Joint Conference on Knowledge Graphs. Hangzhou, 2022: 20-29.
- [127] Girdhar R, El-Nouby A, Liu Z, et al. ImageBind one embedding space to bind them all[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, 2023: 15180-15190.
- [128] National Library of Medicine. Unified medical language system[EB/OL]. (2025-05-05)[2025-08-31]. <https://www.nlm.nih.gov/research/umls/>.
- [129] Li X V, Sanna P F. FinDKG: Dynamic knowledge graphs with large language models for detecting global trends in financial markets[C]. Proceedings of the 5th ACM International Conference on AI in Finance. Brooklyn, 2024: 573-581.
- [130] Chen B H, Bertozzi A L. AutoKG: Efficient automated knowledge graph generation for language models[C]. IEEE International Conference on Big Data. Sorrento, 2023: 3117-3126.
- [131] Melnyk I, Dognin P, Das P. Grapher: Multi-stage knowledge graph construction using pretrained

- language models[EB/OL]. (2023-05-06)[2025-08-31]. <https://openreview.net/pdf?id=N2CFXG8-pRd>.
- [132] Ji Y, Gong Y, Peng Y, et al. Exploring ChatGPT's ability to rank content: A preliminary study on consistency with human preferences[J/OL]. 2023, arXiv: 2303.07610.
- [133] Xu D R, Zhang Z H, Lin Z X, et al. Multi-perspective improvement of knowledge graph completion with large language models[C]. Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, 2022: 11956-11968.
- [134] Jiang J H, Zhou K, Dong Z C, et al. StructGPT: A general framework for large language model to reason over structured data[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Singapore, 2023: 9237-9251.
- [135] Liu X, Wu F, Xu T, et al. Evaluating the factuality of large language models using large-scale knowledge graphs[J]. IEEE Data Engineering Bulletin, 2024, 47(1): 87-108.
- [136] 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.
- [137] Chen Z, Zhang Y C, Fang Y, et al. Knowledge graphs for multi-modal learning: Survey and perspective[J]. Information Fusion, 2025, 121: 103124.
- [138] Huang W, Wang C, Zhang R, et al. VoxPoser: Composable 3D value maps for robotic manipulation with language models[C]. Proceedings of the 7th Conference on Robot Learning. Atlanta, 2023: 540-562.
- [139] Driess D, Xia F, Sajjadi M S M, et al. PaLM-E: An embodied multimodal language model[C]. Proceedings of the 40th International Conference on Machine Learning. Honolulu, 2023: 8469-8488.
- [140] Bian N, Han X P, Sun L, et al. ChatGPT is a knowledgeable but inexperienced solver: An investigation of commonsense problem in large language models[C]. Proceedings of the Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, 2024: 3098-3110.
- [141] Song Y X, Sun P L, Liu H Y, et al. Scene-driven multimodal knowledge graph construction for embodied AI[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 6962-6976.

作者简介

和婷婷 (1998-), 女, 博士生, 主要研究方向为自然语言处理、大语言模型、知识图谱, E-mail: hetingtingjiayou@163.com;

张强 (1982-), 男, 教授, 博士, 博士生导师, 主要研究方向为智能计算与应用、知识表示, E-mail: nepu_zq@163.com;

郑冠彧 (1997-), 男, 博士生, 主要研究方向为文档智能、大语言模型和知识图谱, E-mail: breadhunter@yeah.net;

赵铁军 (1962-), 男, 教授, 博士, 博士生导师, 主要研究方向为人工智能、机器翻译和文本生成, E-mail: tjzhao@hit.edu.cn;

王浩畅 (1974-), 女, 教授, 博士, 主要研究方向为信息抽取、生物信息学和知识图谱, E-mail: kinghaosing@gmail.com;

王颖 (1986-), 女, 副教授, 博士生, 主要研究方向为机器人工程、人工智能, E-mail: wangying@nepu.edu.cn.