

DOI: 10.3969/j.issn.1671-0673. 2026. 04. 013

融合本体与提示学习的目标领域知识抽取方法

林思念, 陈晓慧, 刘建辉, 张继轩, 郑明琪

(信息工程大学, 河南 郑州 450001)

摘要: 针对目标领域知识获取面临的数据稀疏、语义复杂和动态演化快等挑战, 以及现有知识抽取方法存在的领域适应性差、依赖大规模标注等问题, 提出一种融合本体与提示学习的目标领域知识抽取方法。首先, 为应对目标领域概念体系的专业性和动态性, 通过“专家定义—大语言模型扩展—规则校验”的闭环迭代机制, 构建动态演化的领域本体, 为知识抽取提供精确的语义约束; 其次, 为解决目标领域的抽取难题, 设计一种基于结构化提示的混合抽取框架, 通过动态模式注入与并行 Transformer 架构, 实现实体与关系的高效联合抽取; 最后, 利用课程学习策略优化模型在小样本条件下的收敛性与泛化能力。实验结果表明, 该方法有效融合大语言模型的推理能力与领域知识的精确性, 显著提升目标领域知识抽取的准确性与鲁棒性。

关键词: 大语言模型; 本体构建; 提示学习; 知识抽取; 目标领域

中图分类号: TP391.1

文献标识码: A

文章编号: 1671-0673(2026)04-0473-07

Target Domain Knowledge Extraction Method Fusing Ontology and Prompt Learning

LIN Sinian, CHEN Xiaohui, LIU Jianhui, ZHANG Jixuan, ZHENG Mingqi

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To address the challenges in target-domain knowledge acquisition, such as data sparsity, semantic complexity and rapid dynamic evolution, and the problems of poor domain adaptability and dependence on large-scale annotation in existing knowledge extraction methods, a target domain knowledge extraction method fusing ontology and prompt learning is proposed. Firstly, in order to deal with the professionalism and dynamics of the target domain concept system, a dynamically evolving domain ontology is constructed through a closed-loop iterative mechanism of “expert definition-large language model (LLM) expansion-rule validation”, which provides precise semantic constraints for the knowledge extraction. Secondly, in order to solve the extraction difficulties in the target domain, a hybrid extraction framework based on structured prompts is designed. Through dynamic pattern injection and a parallel Transformer architecture, the efficient joint extraction of entities and relations is achieved. Finally, a curriculum learning strategy is employed to enhance the model’s convergence and generalization capabilities under few-shot conditions. Experimental results demonstrate that the reasoning capabilities of LLM are effectively fused with the precision of domain knowledge by the proposed method, significantly improving the accuracy and robustness of knowledge extraction in the target domain.

Key words: LLM; ontology construction; prompt learning; knowledge extraction; target domain

在当前大数据驱动的时代背景下, 信息技术呈指数级发展态势, 深刻影响人类社会运行模式。

其中, 目标领域, 即针对特定对手、系统或任务相关的网络空间情报信息, 因其高度对抗性、时效性

收稿日期: 2025-10-30; 修回日期: 2026-01-21

作者简介: 林思念(2003—), 男, 硕士生, 主要研究方向为信息抽取和规则推理。E-mail: 3183301413@qq.com

和保密性,成为国家安全的关键领域。该领域知识具有3个特殊性。1)数据稀疏与保密:相关情报数据通常以非结构化文本形式存在,获取困难且样本量小,导致依赖大规模标注数据的传统模型难以适用。2)概念动态演化:网络攻击手段、防御策略、基础设施等知识更新迭代极快,静态知识库无法满足实时分析需求。3)语义复杂与隐含:实体间的关联往往是间接和隐含的,需要深度语义推理能力才能发掘。

大语言模型(以下简称大模型)具备超大规模参数化知识库,已成为通用人工智能的重要知识基础^[1],其涌现出的零样本/少样本学习能力^[2],为解决上述低资源抽取问题提供了新范式。相比于需要大量标注数据进行监督训练的传统深度学习模型,如BERT、BiLSTM-CRF,大模型能够直接利用其在海量数据预训练阶段习得的语言知识和推理能力,通过精心设计的提示(prompt)理解并完成特定领域的抽取任务,极大降低了对领域标注数据的依赖。然而,通用大模型在应用于高专业性领域时,仍面临知识时效性不足^[3-4]、事实冲突型幻觉^[5]以及缺乏显式的结构化表达^[6]等挑战,严重制约了其应用的可靠性与精确性。

知识图谱作为一种结构化知识表示范式,能够将零散的知识组织成语义明确的关联网络,为大模型提供可信信息支撑^[7],并提升其推理透明度^[8]。然而,传统知识图谱高度依赖人工,尤其在本体设计和知识抽取环节,不仅成本高昂而且主观性较强。

因此,如何有效结合大模型的强大推理能力与领域知识图谱的结构化优势,是解决目标领域知识自动化、精准化构建的关键。本文提出一种融合本体与提示学习的目标领域知识抽取方法。该方法以Qwen2.5-72B模型为核心,通过构建本体驱动的知识抽取框架,解决了大模型在专业领域

应用中的“幻觉”和“不知”问题;通过设计结构化提示与课程学习策略,克服了低资源场景下的数据稀疏性挑战,推动目标领域知识抽取从“专家经验驱动”向“数据智能驱动”范式转变。

1 相关工作

1.1 大模型驱动的知识抽取方法

知识抽取作为构建知识图谱的核心环节,其技术范式正随着大模型的发展而深刻变革。传统知识抽取方法主要依赖人工规则或浅层机器学习模型,如早期基于正则表达式和词典匹配的方法,通过预定义语法规则实现实体与关系的提取。叶光辉等^[9]通过“人工标注—规则归纳—机器识别—规则补充”流程构建规则库,从文献中抽取知识单元,但该方法依赖人工设计且领域迁移能力有限。深度学习技术的引入推动了端到端知识抽取范式的发展,BERT、BiLSTM等模型通过分布式表征自动捕捉文本语义模式,显著提升了抽取效率。朱闻达^[10]针对长文本实体边界模糊问题,设计多尺度局部感知模块,将长实体识别准确率从78.3%提升至86.7%。

大模型的涌现为低资源场景知识抽取提供了新路径。Tang等^[11]提出多任务BERT-BiLSTM-AM-CRF模型,通过BERT预训练能力实现深度语义编码,结合注意力机制聚合关键信息;Wei等^[12]利用ChatGPT的多轮问答实现零样本知识抽取,首轮识别实体类型,后续通过链式模板提取细粒度知识。然而,现有研究主要集中于通用领域,专业领域知识挖掘仍存在3个方面的挑战:领域适应性不足、标注数据依赖性强和动态更新机制缺失。由于目标领域数据稀疏且存在保密要求,需定制特征提取策略,导致该场景下模型泛化能力显著受限。

1.2 知识图谱构建方法

知识图谱技术架构如图1所示。

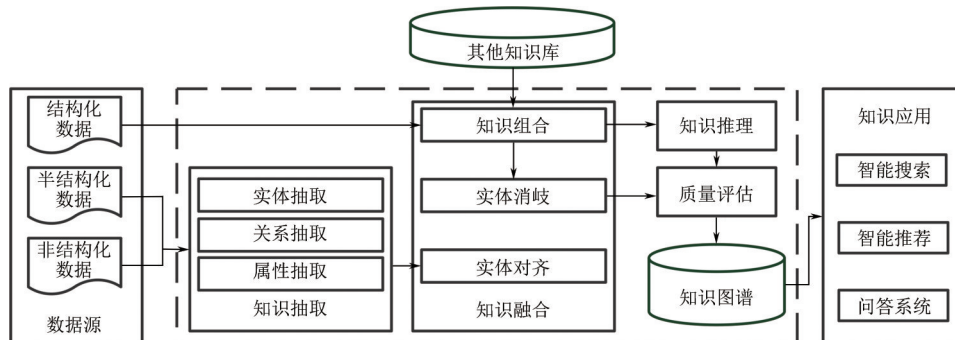


图1 知识图谱技术架构

知识图谱作为结构化知识表征的核心载体,由实体及其相互关系所构成的三元组作为核心组成单元。知识图谱的构建是将海量结构化、半结构化以及非结构化数据,经由知识抽取、知识融合和知识推理等一系列复杂操作后,有序整合至知识库中,并经过多次迭代优化,直至知识体系得以完备。大模型凭借零样本或少样本学习能力,显著提升了知识图谱构建效率。Polak等^[13]通过构建结构化提示词,实现大模型在零样本下的自动数据提取。但这类方法未针对领域特性优化,在复杂自然语言处理任务中难以保证抽取准确性。

1.3 低资源场景下的知识工程优化方法

低资源场景通常指标记数据匮乏、无标记文本稀缺或数据获取困难的情形,其核心挑战在于如何高效利用有限资源实现知识的准确提取^[14-15]。迁移学习作为应对低资源问题的主流策略,通过从其他领域迁移通用知识来提升目标任务性能,包括元学习、多任务学习和提示学习3类方法。

1)元学习通过在多任务上学习元知识,使模型快速适应新任务。Fritzier等^[16]提出的Proto模型将实体类型平均值作为原型,并基于相似度进行分类;Hou等^[17]设计的L-TapNet+CDT模型,通过分数转移机制解决源域与目标域实体类型不一致问题;基于优化的元学习方法如MAML,通过区分任务相关与无关参数,实现快速迁移^[18]。然而,元学习效果受目标域样本质量限制,模型普适性与可解释性仍需提升。

2)多任务学习通过引入辅助任务,挖掘任务关联,为模型提供丰富的监督信号。Bapna等^[19]将低资源实体获取建模为词语匹配任务,通过LSTM编码句子表示;Liu等^[20]提出的MTL+MoEE模型联合实体边界检测与分类任务,引入混合专家模型平衡实体表示。但该方法面临辅助任务筛选困难、任务平衡复杂等挑战,需通过实验确定最优任务组合。

3)提示学习通过将下游任务转换为预训练阶段形式,最大化利用预训练模型的泛化能力。Cui等^[21]提出Template-Based BART模型,通过提示模板生成预测结果;Ma等^[22]设计EntLM模型,利用实体类型高频词作为标签词优化文本表示。传统提示学习需枚举句子片段与实体类型的组合,搜索空间过大,而连续提示学习的设计质量直接影响效果。最新研究如PTE将低资源实体获取转化为文本蕴含预测任务,通过预训练模型对候选实体打分^[23],但小样本场景下监督信号不足问题仍待解决。

当前低资源知识工程研究已取得显著进展,但在专业领域应用中仍存在3大局限:1)领域特定的语义歧义消解机制不足;2)跨模态数据融合技术缺失;3)动态知识持续优化体系尚未完善。如何将大模型推理能力与领域知识体系深度耦合,构建可解释、可迭代的知识工程框架,成为亟待突破的关键方向。

2 融合本体与提示学习的目标领域知识抽取框架

为解决目标领域知识抽取的挑战,本文设计一个以大模型为核心、深度融合领域本体与提示学习的知识抽取框架。该框架包含两大核心模块:专家知识引导的本体动态构建、基于提示学习的领域知识混合抽取。

2.1 专家知识引导的本体动态构建框架

该框架旨在通过“专家定义—大模型扩展—规则校验”的闭环迭代机制,构建符合领域规范、具备动态适配能力的本体模型,为后续知识抽取提供精准、可靠的语义约束基准。

2.1.1 核心本体初始化

本体作为领域知识的结构化表示,为领域知识抽取提供了精确的语义约束和领域知识引导,使提示工程能够更精准地适应目标领域,避免了通用大模型在专业领域应用中的“幻觉”和“不知”问题。

针对目标领域体系构建中领域知识结构化表征需求,设计专家知识驱动的本体初始化机制。领域专家通过基于WebProtégé API扩展开发的可视化标注工具,定义领域核心概念层级体系。通过语义约束规则定义实体类型互斥性,形成符合领域规范的初始本体模型 $O_0 = (C_0, R_0)$,为后续知识抽取提供语义基准。其中: C_0 表示实体类型集合; R_0 表示实体之间允许出现的关系类型集合。

2.1.2 大模型驱动的关系扩展

采用“提示模板—语义筛选”机制实现本体自动化扩展。基于Qwen2.5-72B模型设计领域适配提示模板。通过动态模式注入引导模型生成候选三元组。引入语义相似度置信度 $\text{conf}(t)$ 进行评估:

$$\text{conf}(t) = \frac{1}{n} \sum_{i=1}^n \cos(C_i, C_{D_i}), t \in T_{\text{candidate}} \quad (1)$$

式中: C_i 为三元组 t 的CLS嵌入; C_{D_i} 为文档 D 中第 i 句的嵌入表示; n 为文档 D 的句子数; $T_{\text{candidate}}$ 为候选三元组集合。只有当 $\text{conf}(t)$ 大于阈值时,相应的

三元组才会被保留并进入下一阶段。

2.1.3 规则引擎逻辑校验

设计混合规则引擎实现本体一致性维护。该规则引擎包含两层校验逻辑。1)基于网络本体语言(Web Ontology Language, OWL)本体公理的逻辑校验,针对实体类型互斥性、关系传递性、属性约束等语义规则进行一致性检查;2)基于TransE关系嵌入模型计算合理性得分:

$$s(h,r,t) = \|H + R - T\|_2 \quad (2)$$

式中: h 为头实体(head entity); r 为关系类型(relation); t 为尾实体(tail entity); $H, R, T \in \mathbb{R}^d$ 分别为其各自的嵌入向量表示, d 为嵌入维度; $\|\cdot\|_2$ 为 L_2 范数; $s(h,r,t)$ 的值越小,表示三元组 (h,r,t) 越符合已学习到的结构模式。

若 $s(h,r,t)$ 大于阈值,则将该关系判定为异常关系,并触发专家复审机制。通过多轮迭代优化,实现“专家定义—大语言模型扩展—规则校验”的闭环迭代,使本体能够动态适应领域知识的演化。

2.2 基于提示学习的领域知识混合抽取方法

该方法旨在通过结构化提示设计和双通道联合抽取模型,实现低资源场景下的高效知识抽取。

2.2.1 结构化提示模板设计

构建“领域锚点—语义模式—本体约束”3层提示体系。领域锚点是指从领域本体中提取的关键实体类型,作为提示中的硬性约束;语义模式是指基于抽象语法树生成的动态提示模板,用于引导模型理解特定关系;本体约束是指通过本体定义的实体类型和关系类型之间的逻辑关系,用于保证抽取结果的语义一致性。从本体库提取关键实体类型作为硬提示输入,如“卫星通信链路”“加密协议”等专业术语。基于抽象语法树生成动态提示模板:

$$p = \text{Concat}(s_{\text{prefix}}, E_{\text{type}}, s_{\text{infix}}, R_{\text{schema}}, s_{\text{suffix}}) \quad (3)$$

式中: $\text{Concat}(\cdot)$ 为字符串拼接函数; s_{prefix} 为引导语句; E_{type} 为实体类型标注信息(由领域本体给出,用于硬约束); s_{infix} 为中间连接语句; R_{schema} 为关系模式; s_{suffix} 为后缀约束说明。

2.2.2 双通道联合抽取模型

为实现实体识别与关系分类任务的协同优化,避免传统流水线方法的误差累积,本文设计了一种基于共享编码器的并行双通道Transformer架构。该架构将结构化提示模板输入到大模型中,通过双通道并行处理,同时完成实体识别和关系分类任务,使基于提示学习的领域知识混合抽取

方法得以高效、准确地实现。模型具体组成如下。

1)共享语义编码器。使用Qwen2.5-72B作为底座编码器,负责将输入文本序列转化为富含上下文信息的分布式向量表示。该编码器强大的语义理解能力为后续两个并行任务提供了高质量的特征基础。

2)实体识别通道。该通道是一个基于跨度预测的识别头(head)。它接收编码器输出的每个token的隐藏状态向量,并将其输入到两个独立的线性分类层中,分别用于预测该token作为实体“起始位置”(start)和“结束位置”(end)的概率。通过Softmax函数归一化后,解码器根据概率最高的start-end对来确定实体边界。这种方式相比传统的序列标注,对实体边界的识别更为灵活和精准。

3)关系分类通道。该通道是一个关系分类头。在实体识别通道确定候选实体对后,关系分类通道提取这两个实体对应文本片段(span)的向量表示。随后,将这两个实体的向量表示进行拼接或通过其他方式融合,输入到1个多头自注意力层,以捕捉实体间的深层语义依赖。最后,将注意力层的输出通过1个前馈神经网络(Feedforward Neural Network, FFN)和1个Softmax层,分类预测它们之间的关系类型。

通过将实体识别损失 L_{NER} 与关系分类损失 L_{RC} 进行联合优化,并引入由本体驱动的类型一致性损失 $L_{\text{type_consist}}$,整个模型的目标函数定义为

$$L_{\text{joint}} = \lambda \cdot L_{\text{NER}} + (1 - \lambda)L_{\text{RC}} + \beta \cdot L_{\text{type_consist}} \quad (4)$$

式中, λ, β 为超参数,用于平衡不同任务的重要性。这种联合训练方式使得实体识别和关系分类两个任务能够相互促进:精确的实体边界有助于关系分类,而关系约束也能反过来指导实体类型的识别,从而提升整体抽取性能。

2.2.3 课程学习优化策略

针对目标领域低资源特性,本文采用课程学习策略提升领域抽取效率。该策略通过与大模型的预训练逻辑、提示学习框架深度耦合,解决大模型在小样本条件下易过拟合、领域语义理解不精准及知识抽取泛化能力弱等核心问题,从而最大化发挥大模型的驱动效能。

课程学习借鉴人类认知规律,通过分阶段调控任务难度引导模型逐步累积知识,可显著提升复杂任务的收敛速度与泛化能力。其核心在于模拟人类认知过程,通过有序规划学习任务难度实现高效知识习得。在领域知识抽取场景中,本文

将课程学习实施流程分为以下4个阶段:1)遵循从简到难的递进路径,基于本体层级与文本复杂度将任务划分为基础实体、简单关系、复合实体、复杂关系和跨句关联5级难度;2)强调累积学习机制,确保低难度阶段习得的知识可迁移至高难度任务,依托参数共享实现知识复用;3)实施任务分解策略,将长文本跨句关联抽取拆解为句内实体识别、句内关系抽取、句间实体共指和跨句关系推理4个子任务,分步攻克后整合为完整能力;4)引入适时挑战机制,通过动态评估模型性能设定难度跃迁阈值,规避任务过简导致的资源冗余或过难引发的学习停滞。

为规范课程学习成果的输出形式,设计了结构化JSON格式,明确标注实体类型、文本位置、关系类型和置信度,并标记样本所属课程阶段以便后续难度适配。

3 实验与结果分析

3.1 实验数据及评估指标

针对目标领域数据稀疏且保密性强的特点,本文构建了一个高质量的领域专用数据集。数据来源主要包括两部分:1)通过CNKI、arXiv等学术数据库检索“网络攻击”“卫星通信链路”“加密协议”等关键词获取的专业文献;2)通过Google等搜索引擎采集的公开防务报告与技术白皮书。

原始语料共包含148份长文文档。数据构建流程遵循严格规范。首先,对文档进行PDF转文本、去除乱码与停用词等预处理;其次,由3名具有5年以上经验的领域专家进行标注;最后,生成包含12000条高质量标注样本的基准测试集,涵盖“攻击方”“受害设施”“通信协议”等7类核心实体及“利用”“属于”等12类关系。为模拟真实的低资源场景,实验严格划分数据集,训练集仅随机抽取3%的标注数据(约360条),其余作为验证集与测试集。

为了保证实验的公平性,本文复现的所有对比方法均统一在上述构建的同一目标领域数据集上进行了微调与测试。与对比方法在原研究中使用的通用领域数据集相比,本文构建的数据集具有以下显著差异。1)语义专业性更强。通用数据集多源于新闻或百科,语义直白;而本数据集包含大量如“多跳代理”“侧信道攻击”等高度专业且晦涩的术语,对模型的领域知识理解能力要求更高。2)数据分布更稀疏。传统数据集通常具备万级以

上的均衡标注样本,而本实验设定为低资源场景(仅360条训练数据),且存在严重的长尾分布问题。3)隐含关系更多。目标领域的情报实体往往跨句分布,且关系隐含在复杂的上下文逻辑中,这使得在通用数据集上表现优异的浅层序列标注模型在本数据集上难以捕捉深层语义关联。

3.2 对比方法设置

为全面评估本文方法在目标领域知识抽取任务中的有效性,选取了4种具有代表性的方法作为基准模型进行对比实验。1)Pipeline。采用经典的串行抽取范式,先通过BiLSTM-CRF模型进行实体识别,再利用基于注意力的CNN模型进行关系分类,是传统深度学习时代的典型基准。2)BERT-base^[24]。基于双向Transformer编码器的预训练语言模型,通过微调(fine-tuning)方式应用于下游序列标注与分类任务,代表了预训练时代的通用强基线。3)T5-11B^[25]。基于Encoder-Decoder架构的文本生成式预训练模型,将信息抽取任务转化为文本生成任务,用于评估生成式模型在零样本或少样本下的表现。4)LLM-Alignment。直接使用未经特定提示工程优化的通用大模型(如Qwen2.5-72B基础版)进行指令微调,用于验证本体约束与提示策略相对于直接使用大模型的增益。

3.3 对比实验结果

在领域数据集上,将所提方法与传统Pipeline、BERT-base、T5-11B等模型进行对比,结果如表1所示。其中:实体精确匹配率(Exact Match, EM),用于衡量模型预测实体与真实标注实体完全一致的比例; $F1$ 值为精确率(precision)与召回率(recall)的调和平均,@3表示取模型置信度最高的3个预测结果进行评估;类型一致性(Type Consistency, TC),用于衡量抽取结果与领域本体类型约束的符合程度,即模型预测实体类型与本体定义语义规范的一致性比例;每秒查询数(Queries Per Second, QPS),用于衡量知识抽取系统的实时处理效率,即系统每秒可完成的知识查询或抽取请求数量。从表1可以看出,与T5-11B方法相比,本文方法在实体精确匹配率和关系 $F1@3$ 值上分别高出9.9和12.1个百分点。这主要得益于本体约束和结构化提示机制,它将领域知识有效注入到大模型中,减少了模型的语义漂移和事实性幻觉,使其抽取结果更贴合目标领域的专业规范。并且其类型一致性值高达91.5%,充分验证本体约束作为大模型驱动方法核心适配组件的有效性。

表 1 不同方法在领域数据集上的性能对比				
方法	实体精确 匹配率/%	关系 F1 @3 值/%	类型 一致性值/%	每秒 查询数
Pipeline	68.4	61.1	67.8	14
BERT-base	72.3	65.4	71.2	22
T5-11B	75.8	70.3	71.2	18
LLM-Alignment	78.2	75.6	79.3	28
本文方法	85.7	82.4	91.5	34

消融实验结果如表 2 所示。移除结构化提示模板(-prompt)后,模型性能显著下降,实体精确匹配率下降 12.6 个百分点,表明精确的提示引导对于激活大模型领域知识至关重要。取消联合训练架构(-joint),采用流水线模式后,关系 F1@3 值下降 9.8 个百分点,证实了联合优化能够有效避免误差累积。关闭类型一致性约束(-TC)后,类型一致性值下降 35.2 个百分点,直接证明了本体约束在保证抽取结果语义正确性方面的核心作用。这些结果验证了框架中各组件的必要性和协同效应。

表 2 消融实验结果			
模型	实体精确 匹配率/%	关系 F1 @3 值/%	类型 一致性值/%
完整模型	85.7	82.4	91.5
-prompt	73.1	70.2	68.3
-joint	78.9	72.6	89.1
-TC	83.2	80.5	56.3

在领域数据集上对比无课程学习、逆序课程和渐进课程 3 种课程学习策略,结果如表 3 所示。本文渐进课程的实体精确匹配率与关系 F1@3 值分别较无课程学习提升 9.4 个百分点与 10.9 个百分点,训练收敛轮数降低 12 轮,过拟合率降低 9.5 个百分点。这表明渐进课程策略通过有序引导模型认知过程,有效缓解了目标领域的样本稀疏性问题,同时结构化输出为知识图谱构建提供统一接口,为系统搭建奠定数据基础。

表 3 不同课程策略的性能对比				
策略	实体精确 匹配率/%	关系 F1 @3 值/%	训练 收敛轮数	过拟合 率/%
无课程学习	76.3	71.5	32	18.7
逆序课程	72.1	68.3	38	23.5
渐进课程	85.7	82.4	20	9.2

4 结束语

提出一种融合本体与提示学习的目标领域知识抽取方法,旨在解决目标领域在低资源、高动态

环境下知识获取的难题。通过专家知识引导的本体动态构建、基于提示学习的混合抽取及规则引擎校验,本文方法实现了大模型通用推理能力与领域专业知识的深度耦合。实验证明,该方法显著提升了知识抽取的准确率、关系一致性与领域适应性,为目标领域的自动化知识构建提供了一种有效的新范式。

未来研究将聚焦 3 个方面:1)探索跨模态数据融合,将图像、视频等多模态信息纳入知识图谱构建流程;2)研发增量学习机制,实现领域知识动态更新;3)研究基于知识图谱的下游应用,如攻击溯源、威胁预警等,以期在目标领域中发挥更重要的决策支持作用。

参考文献:

- [1] 郑佳明,陈家宾,胡杰鑫,等.基于大模型和知识图谱的标准领域融合应用方法研究[J].中国标准化,2023(23):39-46.
- [2] 汤萌萌,罗鹏,方晨,等.网络安全领域的事件抽取研究综述[J].信息工程大学学报,2025,26(2):224-230.
- [3] Liu Y, Yao Y S, Ton J F, et al. Trustworthy LLMs: a survey and guideline for evaluating large language models' alignment [PP/OL]. V2. arXiv (2024-03-21) [2025-10-30]. <https://doi.org/10.48550/arXiv.2308.05374>.
- [4] Zhao W X, Zhou K, Li J Y, et al. A survey of large language models [PP/OL]. V16. arXiv (2025-03-11) [2025-10-30]. <https://doi.org/10.48550/arXiv.2303.18223>.
- [5] 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.
- [6] 朱良兵,刘发德.大语言模型与知识图谱的比较和融合研究[J].情报探索,2024(12):8-16.
- [7] 黄勃,吴申奥,王文广,等.图模互补:知识图谱与大模型融合综述[J].武汉大学学报(理学版),2024,70(4):397-412.
- [8] Han H Y, Wang Y, Shomer H, et al. Retrieval-augmented generation with graphs (GraphRAG) [PP/OL]. V2. arXiv (2025-01-08) [2025-10-30] <https://doi.org/10.48550/arXiv.2501.00309>.
- [9] 叶光辉,彭泽,陈国梁,等.学术文献中的知识单元抽取及其分布特征识别研究[J].情报理论与实践,2023,46(4):90-98.
- [10] 朱闻达.基于深度学习的知识抽取算法研究[D].北京:北京交通大学,2023:1-74.
- [11] Tang X Y, Huang Y, Xia M, et al. A multi-task BERT-BiLSTM-AM-CRF strategy for Chinese named entity recognition [J]. Neural Processing Letters, 2023, 55 (2) : 1209-1229.
- [12] Wei X, Cui X Y, Cheng N, et al. ChatIE: zero-shot infor-

- mation extraction via chatting with ChatGPT [PP/OL]. (2023-02-20)[2025-10-30]. <https://doi.org/10.48550/arXiv.2302.10205>.
- [13] Polak M P, Morgan D. Extracting accurate materials data from research papers with conversational language models and prompt engineering [J]. *Nature Communications*, 2024, 15: 1569.
- [14] Hedderich M A, Lange L, Adel H, et al. A survey on recent approaches for natural language processing in low-resource scenarios [C]//*Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: human language technologies*. Stroudsburg, USA: ACL, 2021. DOI: 10.18653/v1/2021.naacl-main.201.
- [15] 于浏洋,郭志刚,陈刚,等.面向知识图谱构建的知识抽取技术综述[J]. *信息工程大学学报*, 2020, 21(2): 227-235.
- [16] Fritzler A, Logacheva V, Kretov M. Few-shot classification in named entity recognition task [C]//*Proceedings of the 34th ACM/SIGAPP symposium on applied computing*. New York, USA: ACM, 2019:993-1000.
- [17] Hou Y T, Che W X, Lai Y K, et al. Few-shot slot tagging with collapsed dependency transfer and label-enhanced task-adaptive projection network [C]//*Proceedings of the 58th annual meeting of the association for computational linguistics*. Stroudsburg, USA: ACL, 2020:1381-1393.
- [18] Li J, Shang S, Shao L. MetaNER: named entity recognition with meta-learning [C]//*Proceedings of the web conference 2020*. New York, USA: ACM, 2020. DOI: 10.1145/3366423.3380127.
- [19] Bapna A, Tür G, Hakkani-Tür D, et al. Towards zero-shot frame semantic parsing for domain scaling [C]//*Proceedings of the interspeech 2017*. Baixas, France: ISCA, 2017:2476-2480.
- [20] Liu Z H, Winata G I, Fung P. Zero-resource cross-domain named entity recognition [C]//*Proceedings of the 5th workshop on representation learning for NLP*. Stroudsburg, USA: ACL, 2020. DOI: 10.18653/v1/2020.repl4nlp-1.1.
- [21] Cui L Y, Wu Y, Liu J, et al. Template-based named entity recognition using BART [C]//*Findings of the association for computational linguistics: ACL-IJCNLP 2021*. Stroudsburg, USA: ACL, 2021. DOI: 10.18653/v1/2021.findings-acl.161.
- [22] Ma R T, Zhou X, Gui T, et al. Template-free prompt tuning for few-shot NER [PP/OL]. V3. arXiv (2022-11-23)[2025-10-30]. <https://doi.org/10.48550/arXiv.2109.13532>.
- [23] Li D F, Hu B T, Chen Q C. Prompt-based text entailment for low-resource named entity recognition [C]//*Proceedings of the 29th international conference on computational linguistics*. Stroudsburg, USA: ACL, 2022:1896-1903.
- [24] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//*Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics: human language technologies*. Stroudsburg, USA: ACL, 2019:4171-4186.
- [25] Roberts A, Raffel C, Shazeer N. How much knowledge can you pack into the parameters of a language model? [C]//*Proceedings of the 2020 conference on empirical methods in natural language processing*. Stroudsburg, USA: ACL, 2020. DOI: 10.18653/v1/2020.emnlp-main.437.

(编辑:李志豪)

(上接第472页)

- [29] 刘磊,王卓浩,李芃羽.量子操作系统(QuOS)及软件栈的设计[J]. *电子科技大学学报*, 2025, 54(2): 210-220.
- [30] 潘建伟.量子信息科技的发展现状与展望[J]. *物理学报*, 2024, 73(1): 7-14.
- [31] 王东升.通用量子计算模型:一个资源理论的视角[J]. *物理学报*, 2024, 73(22): 16-37.
- [32] Bertels K, Sarkar A, Hubregtsen T, et al. Quantum computer architecture: towards full-stack quantum accelerators [C]//*Proceedings of the 23rd conference on design, automation and test in Europe*. New York, USA: ACM, 2020:139-144.
- [33] 范桁.量子计算纠错取得突破性进展[J]. *物理学报*, 2023, 72(7): 18-20.
- [34] 周文豪,王耀,翁文康,等.集成光量子计算的研究进展[J]. *物理学报*, 2022, 71(24): 22-35.
- [35] 龙桂鲁,范桁,郑超,等.量子计算新进展:硬件、算法和软件专题编者按[J]. *物理学报*, 2022, 71(7): 1.
- [36] Yang Y C, Shen Z T, Zhu X, et al. FPGA-based electronic system for the control and readout of superconducting quantum processors [J]. *Review of Scientific Instruments*, 2022, 93(7): 074701.
- [37] Engler D R. The exokernel operating system architecture [D]. Cambridge, USA: Massachusetts Institute of Technology, 1998: 1-120.
- [38] Yuan Y F, Huang J H, Sun Y, et al. Lambda: RDMA-driven acceleration framework for memory-intensive μ -scale datacenter applications [C]//*Proceedings of the 2023 IEEE international symposium on high-performance computer architecture*. Piscataway, USA: IEEE, 2023: 499-515.
- [39] Qi S X, Monis L, Zeng Z T, et al. SPRIGHT: extracting the server from serverless computing! high-performance eBPF-based event-driven, shared-memory processing [C]//*Proceedings of the ACM SIGCOMM 2022 conference*. New York, USA: ACM, 2022: 780-794.

(编辑:李志豪)