学科知识图谱构建中的大模型关系抽取：信息感知与后验证策略

Relation Extraction for Subject Knowledge Graph Construction using Large Language Models: Information-Aware and Post-Verification Strategies

李华政[1]，张准[2*]
华南师范大学
lilhuaz@m.scnu.edu.c
zhzhun@scnu.edu.cn

【摘要】 在人工智能与个性化学习深度融合的趋势下，自动化抽取知识点逻辑关系以构建学科知识图谱，是实现个性化学习路径动态生成与智能学伴研发的基础设施。针对大语言模型在关系抽取任务中存在的上下文利用率低及“幻觉”现象，本文提出一种信息感知的检索增强与后验证策略。通过集成测试样本的实体与语义信息，增强检索器获取高质量示例的能力；并设计后验证机制，利用模型的连续对话能力对预测结果进行一致性校验。在 SciERC 等包含大量学术知识的数据集上进行实验，结果表明，该方法显著提升了复杂学术关系的抽取精度。本研究为自动化教育资源组织与智能教学系统的知识库构建提供了有效且稳健的技术方案。
【关键词】 关系抽取；大语言模型；学科知识图谱；信息感知；后验证

Abstract: *Automatically extracting knowledge relations from massive resources is key to constructing subject knowledge graphs in intelligent education. Addressing the issues of low context utilization and "hallucination" in Large Language Models (LLMs) for Relation Extraction (RE), we propose an information-aware retrieval and post-verification strategy. By integrating entity and semantic details, we improve the quality of demonstration retrieval. A post-verification mechanism is further designed to ensure prediction consistency via LLMs' dialogue capabilities. Experiments on academic datasets like SciERC demonstrate significant accuracy improvements in complex relation extraction. This study offers a robust technical solution for automated educational resource organization and knowledge base construction in intelligent tutoring systems.*
Keywords: Relation Extraction; Large Language Models; Subject Knowledge Graph; Information-Awareness; Post-Verification

1. 前言

在数字化转型的浪潮下，智慧教育已进入以人工智能为核心的 2.0 时代。随着在线开放课程（MOOCs）和数字化教材的普及，海量的非结构化教学资源呈爆炸式增长。然而，这些资源往往以孤立、零散的形式存在，学生在面对复杂的学科知识体系时容易产生“认知负荷过载”。学科知识图谱作为一种能够结构化呈现知识点及其逻辑脉络的工具，已成为实现个性化学习路径推荐、智能辅导系统（ITS）以及自动化题目生成的关键基础设施。然而，传统学科知识图谱的构建高度依赖教育专家的手工标注，不仅成本高昂且难以实时更新。因此，研发精准、可靠的自动化关系抽取（Relation Extraction, RE）技术，已成为构建“智能教育地图”的迫切需求 (Chen et al., 2023)。

目前，利用大语言模型（Large Language Models, LLMs）进行关系抽取已成为人工智能教育应用领域的研究热点。然而，在教育场景下，学科知识具有极高的专业性与严谨性。通用大模型在执行抽取任务时仍面临两大挑战：首先，由于缺乏特定学科的上下文指引，模型容易误解专业术语间的深层语义；其次，模型在处理复杂逻辑推理时易产生“幻觉”现象（Hallucination），输出错误的知识关系，这严重影响了自动构建知识库的可信度。尽管通用领域的大语言模型在零样本（Zero-shot）学习中展现了惊人的理解力，但将其直接应用于教育场景仍面临严峻挑战 (Ouyang et al., 2022)。教育文本（如数学定义、生物代谢过程、历史因果逻辑）具有高度的逻辑严密性。研究表明，通用模型由于缺乏学科本体知识的约束，在处理学科专业词汇时常表现出语义偏移。例如，在处理学科术语间的“组成”或“演化” 关系时，模型可能仅捕捉到词汇共现的统计概率，而非真正的逻辑关联。这种幻觉现象一旦 进入教学系统，可能会传播错误的学科知识，误导学习者，这与教育的严肃性本质相违背。

针对上述挑战，本文提出一种面向学科知识图谱构建的信息感知与后验证关系抽取方 法。本文的贡献主要体现在：1) 提出一种信息感知检索器，通过集成测试样本的实体与语 义信息，显著提升了语境学习（In-Context Learning）中示例抽取的质量；2) 设计了针对 RE 任务的后验证策略，利用模型的连续对话能力对预测结果进行一致性校验，有效降低了 逻辑错误的发生率；3) 在 SciERC 等包含大量科学知识的数据集上进行了验证，结果表明 该方法能为自动化教育资源组织提供稳健的技术支撑；4) 该框架通过大模型自动抽取与逻辑自我校验的人机协同机制，实现了高质量的知识建构表征。

2. 相关工作

2.1 语境学习与学科关系抽取

语境学习（In-Context Learning, ICL）的核心逻辑在于利用少量的演示示例（Demonstration）激活模型的特定下游任务能力(Brown et al., 2020)。在教育领域，研究者已尝试利用 ICL 从科学教材中提取实体关系。例如，Wang 等人（2022）尝试通过手动设计的模板引导模型识别知识点。然而，固定模板的局限性在于无法适应学科文本的多样性。学科语言（Subject-specific language）通常包含复杂的句式结构和长程依赖，简单的示例往往不足以覆盖所有可能的语义变体。因此，如何从动态变化的教学资源库中自动筛选出最具“代表性”和“类比价值”的示例，成为 ICL 在教育应用中的核心瓶颈 (Min et al., 2022)。

2.2 示例检索技术

演示示例的选择直接影响 ICL 的最终效果。现有的检索方案主要基于向量空间模型。例如，基于 SIMCSE 的对比学习方法通过计算句向量余弦相似度来匹配示例(Gao et al., 2021)。但在处理“光合作用”与“叶绿体”这类具有特定学科属性的关系时，单纯的句子级嵌入往往由于忽略了实体本身的语义特征而导致匹配失效 (Reimers & Gurevych, 2019)。Li 等人（2023）的研究指出，如果检索器无法感知实体层面的信息，检索出的示例可能在句式上相似但在逻辑结构上完全脱节。本文提出的信息感知检索策略正是为了填补这一空白，通过多维度特征提取，确保检索到的示例不仅在文法上相近，在学科逻辑上也具有强关联性。

2.3 后验证策略

针对大模型的生成不确定性，研究界提出了多种校验机制。Chain-of-Thought（CoT）虽能提升推理能力，但在长文本处理中常出现“一步错、步步错”的漂移现象 (Wang et al., 2022)。在教育测评中，对准确率的要求近乎苛刻。Self-Refine 等迭代策略虽有改善，但往往耗费过高的计算资源。本文提出的后验证策略（Post-Verification）借鉴了教育心理学中的“反向论证”

思维。通过构建正向预测与逆向校验的闭环对话，本方法能更轻量化地实现在线逻辑审查(Weng et al., 2022; Dhuliawala et al., 2023)。这种设计不仅是技术上的改进，更是对教育系统中“严谨知识产出”这一准则的技术映射。

3. 研究方法

3.1 任务定义

关系抽取(RE)的目标是从非结构化文本中识别实体间的语义联系。令 S 为输入的学科句子，$\boldsymbol{E}_{head}$ 和 $\boldsymbol{E}_{tail}$ 为待识别实体。本文的目标是在预定义的学科关系集 R 中，预测最符合语义的标签 $\boldsymbol{r} \in \boldsymbol{R}$。

3.2 总体架构

本文提出的面向学科知识图谱构建的关系抽取框架如图 1 所示。该框架主要由两个核心模块组成：

1) 信息感知演示检索器，负责从学科训练集中提取与测试样本最相关的示例文本；

2) 后验证模块，通过构建正序与逆序的 $Prompt$ 提示词，并利用 LLM 的一致性校验能力对初步抽取结果进行修正。

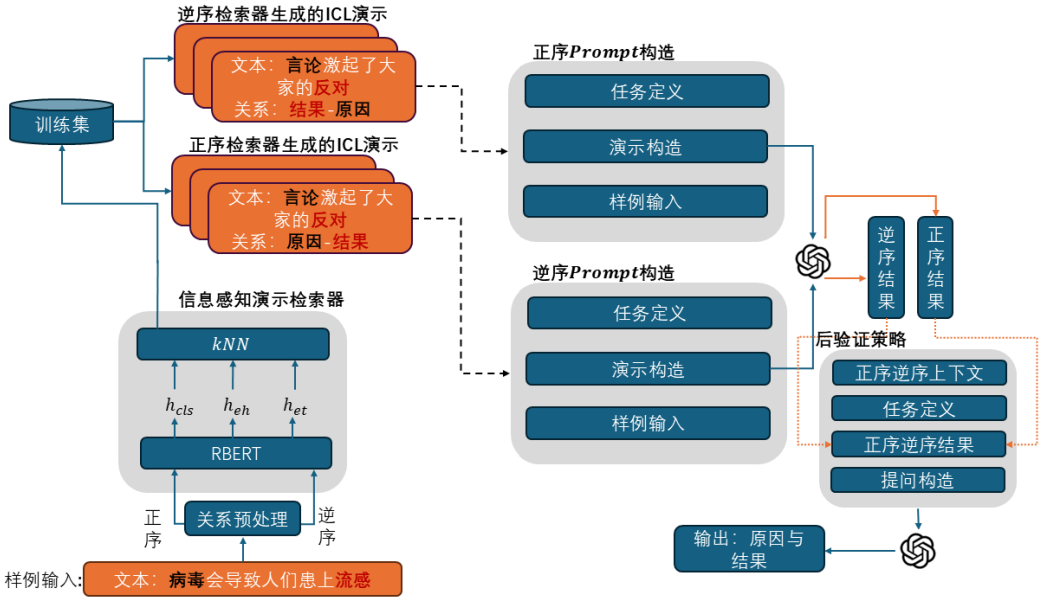

图 1. 信息感知与后验证关系抽取框架总体架构图

3.3 信息感知演示检索器

为使模型理解特定的学科背景，本文利用 RBERT 模型提取隐在此处键入公式。藏层信息(Wu & He, 2019)。通过连接[CLS] 向量、头实体（h_{eh}）与尾实体（h_{et}）向量，获得包含学科语义特征的特征向量 h_i。通过公式 (1) 将隐藏层信息进行拼接：

$$\boldsymbol{h_i} = [\boldsymbol{h_{cls}} \oplus \boldsymbol{h_{en}} \oplus \boldsymbol{h_{et}}] \tag{1}$$

其中，$\oplus$表示向量拼接操作。公式 (1) 实现了句子级语义与实体级特征的深度融合，使检索器能更精准地识别学科背景下的知识关联。随后，利用 k-最近邻（kNN）算法在训练空间中检索相似示例。这种方法不仅考虑了句子的整体语义，更强化了对教育文本中关键实体及其交互关系的感知能力。

3.4 提示词构造与后验证策略

在提示词（Prompt）构造阶段，本文分别构建正序（$Forward$）与逆序（$Reversed$）任务定义。如图 1 所示，该过程分为预测与验证两个阶段。首先，模型利用检索到的 k 个相似示例构建语境学习提示词。一个典型的正序提示词结构如下所示（采用区块引用格式）：

> 任务定义:你是一名资深的学科专家。请根据以下示例，判断给定句子中两个知识点间的逻辑关系。
> 示例演示:给定句子S_i，实体 E_{ih}与E_{it}的关系是r_i。
> 待测输入:给定句子S，预测E_{head}与E_{tail}之间的关系。

模型首先分别生成正序和逆序的初步抽取结果y_f与 y_r。若两者结果不一致，则触发后验证策略。该策略将两个方向的上下文及结果再次输入 LLM，引导模型进行逻辑自洽性校验。这种机制通过“二次确认”，有效降低了 LLM 在处理复杂学术逻辑时的幻觉风险，确保了学科知识图谱构建的严谨性。

4. 实验结果与分析

4.1 研究设计与数据来源说明

本研究定位于智能教学系统投入课堂应用前的“底层知识建构预研阶段”。研究对象聚焦于“数字化教学文本的语义表征”，而非直接采集终端学生的交互数据。为模拟真实且严谨的高等教育教学情境，本研究选取了富含复杂科学实体与关系的 SciERC 数据集作为核心样本来源(Luan et al., 2018)，以此代表真实的理工科数字化文献与教材。此外，辅以 SemEval 与 ACE05 数据集作为基准对照(Hendrickx et al., 2019)。整个实验实施过程模拟了“智能学伴自动阅读并构建课程逻辑地图”的后台数据处理环节，采用 Micro-F1 值对知识抽取的准确性进行规范化与量化的评估。该指标通过聚合所有类别的贡献来衡量模型的整体性能。

4.2 实验实施与分析逻辑

本文基于 GPT-3.5-Turbo 与 GPT-4o-mini 接口进行实验。对比算法包括传统的全监督微调模型（如 RBERT, PURE）以及先进的大模型关系抽取方法 (如 GPTRE-FT)(Wan et al., 2023)。实验统一采用 25-shot 设置。

为提升研究过程的透明性并紧扣教育技术应用需求，本文的方法应用与实验分析逻辑分为三个递进步骤：1) 基线性能评估：通过与上述全监督模型及大模型框架进行横向对照，确立学科知识抽取的性能上限与基准；2) 核心模块消融分析：采用控制变量法，分别剥离“信息感知检索”与“后验证”模块，量化解析各组件在抑制 AI 逻辑幻觉中的独立贡献；3) 教学情境白盒剖析：结合具体的医学常识错判案例，透明化展示模型从“产生认知偏差”到“自我逻辑纠错”的完整推演路径。此外，所有实验均在相同算力环境下运行 5 次并取平均值，以消除生成随机性干扰，确保结果的稳定性和可复现性。

4.3 实验结果分析

4.3.1 总体性能表现分析

实验结果如表 1 所示。在 SemEval-2010 Task 8 数据集上，本文方法（GPT3.5-FT+后验证）取得了 93.18% 的最高 F1 值，显著优于传统微调模型。这表明在通用语义环境下，信息感知检索能为模型提供极具类比价值的上下文，使其对关系类别的判定更具确定性。

特别值得关注的是在 ACE05 数据集上的表现。由于该数据集文本结构复杂，基础模型 GPT3.5-Random 仅获得了 9.04% 的极低分，这反映了 LLM 在处理多重实体干扰时的脆弱性。然而，通过引入本文的信息感知检索策略，性能大幅跃升至 73.38%（提升了 64.34 个百分点），这充分证明了通过拼接实体向量（ h_{eh} 和 h_{et}）实现的特征增强，能有效引导模型在长难句中准确锁定核心语义点。

4.3.2 针对学科知识抽取的深入探讨

在模拟真实教学场景的 SciERC 科学文献数据集上，本文方法表现出了极强的鲁棒性（F1 值达 77.49%）。这一提升幅度（相比随机检索提升了约 60%）对教育知识图谱的构建具有重要意义。分析原因主要有两点：

1. 学科背景的一致性：信息感知检索器通过语义感知，从训练集中检索出具有相同学科逻辑的示例（如“组成”、“因果”关系），为模型提供了精准的推理模版，从而降低了跨学科术语带来的理解门槛。

2. 后验证机制的逻辑兜底作用：在严谨的学科知识建模中，幻觉导致的逻辑错误（如反转因果）是致命的。后验证策略通过“对话式纠错”，在 SciERC 数据集上为模型带来了稳定的性能增量。实验证明，该策略能有效捕捉模型在处理高难度学术关系时的逻辑矛盾，显著提升了学科知识自动抽取的严肃性与可信度。

表 1. 各数据集上的关系抽取性能对比（Micro-F1 值）。星号（*）表示在相同 k-shot 实验条件下引入后验证模块后的结果；下划线表示该结果超过了全监督微调模型；加粗结果代表性能优于此前最先进的 GPTRE-FT 方法。

方法	检索器	SemEval	ACE05	SciERC
		GPT 基线 (Best k-shot)		
GPT3.5_random		70.04(30)	9.04(25)	17.92(25)
GPT4_random		48.82(25)	81.22(25)	69.48(25)
		本文方法(Best k-shot)		
GPT3.5-SimCSE*	SimCSE	80.22(25)	63.00(25)	30.75(25)
+后验证	SimCSE	83.81(25)	68.00(25)	38.70(25)
GPT3.5_FT	RBERT	90.25(15)	64.31(15)	59.75(15)
GPT3.5-FT*	RBERT	**92.37(25)**	**73.38(25)**	**77.49(25)**
+后验证	RBERT	**93.18(25)**	**74.50(25)**	**73.60(25)**
GPT4-FT	RBERT	90.15(25)	**93.00(25)**	**86.00(25)**
		微调模型以及其他方法		
PURE		89.90	70.09	68.45
GPT-RE_FT	PURE	91.90(25)	68.70(25)	69.00(30)
RBERT		89.10		

4.4 消融实验与分析

为了进一步验证本文提出的信息感知检索器（Information-aware Retriever）与后验证策略（Post-verification Strategy）的各自贡献，本文在 SemEval 数据集上进行了消融实验，结果如图 2 所示。

4.4.1 实验设计与实现细节

为了精确量化各模块对性能的贡献，本文设计了三组对照实验：1) w/o inf：移除信息感知检索器，验证实体特征对语义锚定的作用；2) w/o PV：移除后验证模块，衡量逻辑校验对幻觉的抑制效果；3) w/o inf&PV：同时移除两者作为基准。

在实现细节方面，本文基于 Python 3.9 环境开发，调用 OpenAI 官方 API 进行模型交互。GPT-3.5-turbo 的温度系数（Temperature）统一设为 0.1 以确保结果的确定性，top_p 设为 1，最大生成长度限制为 256 词。检索器采用 RBERT 模型提取 768 维特征向量，kNN 检索基于欧氏距离计算。所有消融实验均在相同的硬件环境（Intel Xeon CPU, 64GB RAM）下运行，并对实验结果取 5 次运行的平均值，以消除生成随机性带来的干扰。

在实验超参数的选择上，本文进行了细致的逻辑权衡。首先，对比 15、25 和 30 三种 k-shot 配置，旨在探究不同密度的学科背景对模型性能的边际效应；实验发现 25-shot 在维持提示词（Prompt）长度适中的同时，能提供最稳定的语义参考。其次，将模型生成温度（Temperature）固定为 0.1，是为了在严谨的学科知识抽取任务中最大程度抑制 LLM 的“随

机创造性”，确保教育知识图谱构建过程中的逻辑一致性与结果可复现性。此外，检索算法采用欧氏距离而非余弦相似度，是因为前者在捕捉 RBERT 隐藏层特征空间中的绝对特征差异方面表现更为稳健。

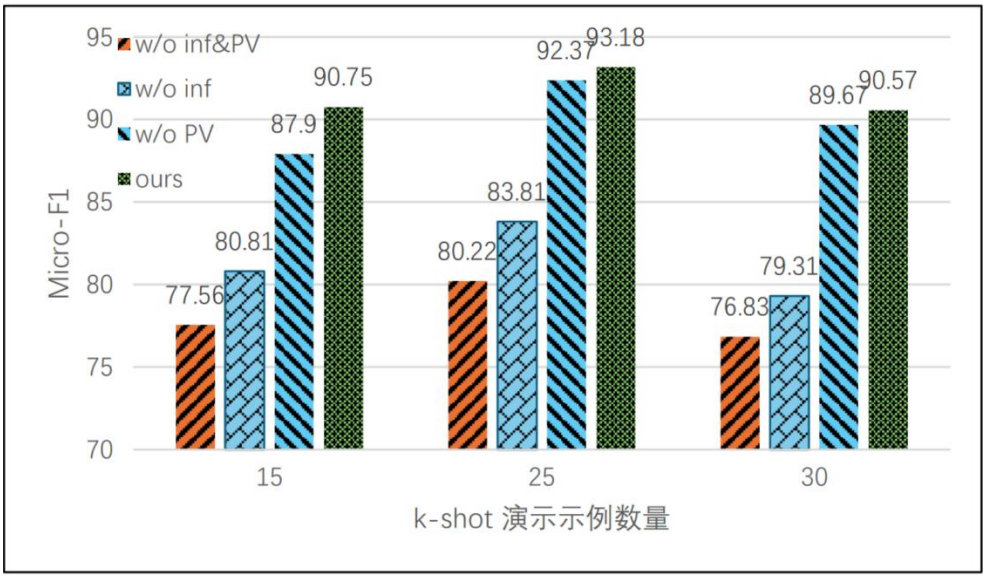

图 2. 不同模块对模型性能的影响对比（Micro-F1 值）。其中，ours 指代本文方法；w/o inf 指代去除信息感知检索器；w/o PV 指代去除后验证策略；w/o inf&PV 指代同时去除两个模块。

4.4.2 消融实验结果讨论

观察图 2 的实验数据，我们可以从以下三个维度进行深层次探讨：

1. 学科语义特征的引导作用：在移除信息感知检索（w/o inf）后，模型 F1 值平均下降了 10.19%。这从侧面证明了通用 LLM 在处理专业学科文本时，存在明显的“上下文错配”问题。本文提出的 h_j 特征向量通过显式地将实体属性引入检索空间，实际上为模型提供了一种“认知脚手架”，使其能够通过类比快速掌握学科关系的判定逻辑。

2. 逻辑校验对幻觉的阻断机制：移除后验证策略（w/o PV）导致的性能下跌（0.81%-2.85%）揭示了 LLM 在处理严谨知识时的“过度自信”倾向。尽管降幅在数值上小于检索器，但在教育场景下，这种校验机制的意义在于将逻辑矛盾率降至最低。后验证模块通过引入“反向推导”思维，强制模型进入系统 2（慢思考）模式，从而有效拦截了大量潜在的逻辑幻觉。

3. 模块间的协同增效效应：当同时去除两个模块后，性能出现了显著的断崖式下跌（约 12.66%）。这说明信息感知检索（提供高质量输入）与后验证策略（实施严格输出审校）之间存在强耦合关系。只有在确保输入背景精准的前提下，逻辑校验才能获得正确的参照基准，进而实现对学科知识图谱构建全流程的质量把控。

4.5 案例分析

为了直观展示本文方法在处理严谨学科知识时的纠错能力，本节选取了一个典型的医学常识关系抽取案例进行对比分析，具体过程如表 2 所示。

表 2. 学科知识关系抽取案例分析

步骤	详细内容	逻辑说明
输入文本	“病毒会导致人们患上流感。”	包含学科因果逻辑的句子
初步预测	关系:相关关系	**[错误]**通用模型倾向于给出泛化关联。
后验证机制	正向:原因-结果;逆向:流感导致病毒	**[冲突]**发现逻辑不自洽。
本文输出	**原因-结果**	**[修正]**成功识别出学科逻辑关系。

案例讨论：

如表 2 所示，通用大模型（基准模型）由于缺乏学科领域知识的约束，容易给出模糊的预测结果。在教育场景中，这种不确定性会直接导致知识图谱的逻辑断裂。

本文方法的优势在于：1) 检索模块通过信息感知能力，从知识库中精准匹配了具有相同“因果逻辑”的学科示例，为 LLM 提供了强有力的类比参考；2) 后验证模块利用 LLM 的自我博弈能力，通过逆向思维发现逻辑矛盾（即“患病”不能反向推导为“产生病毒”），从而迫使模型从语义深层进行修正。该案例证明了本文方法能显著提升学科知识发现的严谨性，确保了在构建学科知识图谱时，知识点之间的逻辑链条准确无误。

5. 讨论与应用前景

5.1 人机协同在知识建构中的作用机制

本文提出的关系抽取框架，实质上在智能教育底层构建了一种高效的“人机协同知识处理机制”。在该机制中，AI（大模型）作为认知代理，承担了海量教学资源的高负荷提取与初始表征工作。其具体作用路径体现为两个阶段：首先，“信息感知检索模块”模拟了学习者“激活先验知识”的认知过程，为 AI 提供精准的学科参考锚点；其次，“后验证策略”模拟了人类专家的“批判性反思”，通过正逆向逻辑博弈，强制 AI 过滤违背学科常识的错误（如误判“疾病产生病毒”）。这种协同机制不仅从源头上保障了系统底层知识库的绝对严谨，更极大提升了教学内容的组织效率，为最终向学习者输出高质量的结构化知识提供了核心支撑。

5.2 面向真实教学场景的实践启示

结合具体的教学实践，本研究的底层算法框架可直接转化为以下两项具操作性的教育应用：(1) 赋能智能教材的动态编排：教师或教研团队可将本框架作为智能辅助插件嵌入在线备课平台。面对长篇晦涩的 STEM 课文，系统能一键精准提取核心概念及其因果、层级关系，辅助教师快速生成可视化的知识图谱与结构化板书设计，显著降低备课的认知负荷。(2) 构建自适应智能学伴的“逻辑安全网”：在研发面向学生的交互式答疑机器人时，可将本文的“双向后验证策略”作为内容输出前的必经防火墙。在 AI 向学生推送解答前，强制进行一轮逻辑自洽校验，彻底阻断 AI 幻觉引发的知识误导，从而营造高可信、高容错的人机协同学习生态。

5.3 局限性与挑战

尽管本研究取得了显著提升，但仍存在一定的局限性。首先，本文方法目前主要针对文本资源，对于教材中的图像、公式等跨模态信息的处理能力仍有待加强。其次，后验证策略虽然显著提升了精度，但也引入了额外的 API 交互开销，如何进一步优化推理效率是未来的研究方向。需要指出的是，尽管本文的实证研究基于 GPT-3.5/4 开展，但本文提出的‘信息感知检索+后验证’框架具有模型无关性（Model-agnostic），可无缝迁移至新一代大模型中，持续为个性化学习系统的知识底座提供高可靠的技术保障。

6. 结论

本文提出一种集成信息感知检索与后验证策略的关系抽取框架，有效缓解了学科图谱构建中的幻觉与上下文利用问题。实验表明，该方法在无需大规模微调的前提下，通过优化语境学习的示例质量并强化输出逻辑约束，显著提升了关系抽取的准确率。

本研究证明了该框架在智慧教育系统中的应用潜力。未来，我们将探索将特定学科的领域本体（Ontology）引入后验证流程，并结合多模态知识发现技术，为大规模数字化教育资源的深度组织提供更全面、更智能的技术保障。

参考文献

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33,* 1877-1901.

Chen, Q., Li, H., & Zhang, Z. (2023). Fact-condition statements and super relation extraction for geothermic knowledge graphs construction. *Geoscience Frontiers, 14*(5), 101412. https://doi.org/10.1016/j.gsf.2023.101412

Dhuliawala, S., Thompson, A., Zhou, D., Schuurmans, D., & Guu, K. (2023). Chain-of-verification reduces hallucination in large language models. *arXiv preprint arXiv:2309.11495.*

Gao, T., Yao, X., & Chen, D. (2021). SimCSE: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821.*

Hendrickx, I., Lefever, E., & Hoste, V. (2019). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. *arXiv preprint arXiv:1911.10422.*

Li, G., et al. (2023). Unlocking instructive in-context learning with tabular prompting for relational triple extraction. *arXiv preprint arXiv:2402.13741.*

Luan, Y., Wadden, D., & Wang, L. (2018). SciERC: Scientific entities, relations, and coreference dataset. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.*

Min, S., Lyu, X., Ariyaeeania, A., & Zettlemoyer, L. (2022). Rethinking the role of demonstrations: What makes in-context learning work? *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.*

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35,* 27730-27744.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks.*Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing.*

Wan, Z., et al. (2023). GPT-RE: In-context learning for relation extraction using large language models. *arXiv preprint arXiv:2305.02105.*

Wang, B., Deng, X., & Sun, H. (2022). Iteratively prompt pre-trained language models for chain of thought. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing.*

Weng, Y., Zhu, W., Xia, F., Li, C., He, S., Liu, K., & Zhao, J. (2022). Large language models are better reasoners with self-verification. *arXiv preprint arXiv:2212.09561.*

Wu, S., & He, Y. (2019). Enriching pre-trained language model with entity information for relation classification. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management.*