

基于知识图谱增强大语言模型的课堂教学分析问答系统

A Knowledge Graph-Enhanced Large Language Model-Based Question-Answering System for Classroom Teaching Analysis

郑扬俊*, 李嘉, 杨瑞雨, 杨芷悦, 庄颖, 陈自力
华南师范大学

* 1944756625@qq.com

【摘要】 针对 CSMS 课堂教学分析报告指标复杂、难于解读的痛点, 本文提出融合知识图谱与大模型的问答系统 CSMS-Chat。系统首先将半结构化 CSMS 报告转化为领域知识图谱, 显式建模教学指标、分析事实与改进建议之间的结构关系; 在此基础上, 采用混合检索驱动的语义解析策略精准定位图谱子图, 并将其重构为具备推理线索的“知识路径”, 以弥合图谱结构与大模型文本输入之间的差异, 结合 DeepSeek-R1 实现深度推理。实验表明, 该系统在 BLEU 指标及专家主观评价中均显著优于传统 RAG 方法, 有效降低了教师解读报告的认知负荷。

【关键词】 课堂教学智慧评价; 大语言模型; 知识图谱问答; 系统设计与开发; 教学改进

Abstract: To address the difficulty of interpreting CSMS classroom analysis reports caused by complex indicators, this paper proposes CSMS-Chat, a question-answering system integrating knowledge graphs and large language models. Semi-structured CSMS reports are first transformed into a domain knowledge graph that explicitly models relationships among teaching indicators, analytical findings, and improvement recommendations. A hybrid-retrieval-driven semantic parsing strategy is then applied to accurately locate relevant subgraphs and reconstruct them into reasoning-aware knowledge paths, bridging the gap between graph structures and textual model inputs. Combined with the deep reasoning capability of DeepSeek-R1, the system enables interpretable analysis. Experimental results demonstrate that CSMS-Chat significantly outperforms conventional RAG methods in BLEU scores and expert evaluations, effectively reducing teachers' cognitive load.

Keywords: Intelligent Classroom Evaluation, Large Language Models, Knowledge Graph-based Question Answering, System Design and Development, Instructional Improvement

1.问题的提出

随着人工智能与大数据技术在教育领域的深入应用, 课堂教学评价正从主观经验主义向数据驱动的循证实践转型。课堂教学智慧评价系统 (Classroom Structure-based Multi-model-supported Scoring System, CSMS) 能够基于多模态课堂数据生成涵盖教学结构、思维激发和交流互动等多个量化指标的课堂大数据教学分析报告, 为教师专业发展和课堂质量提升提供新

基金项目: 教育部人文社会科学研究规划基金项目《人机协同学习机制研究——基于多智能体的角色建模与知识扩散仿真》(课题编号: 25YJA880065)

途径（中国社会科学网，2023）。然而，在实际应用中，CSMS 报告存在“理论性强、操作性弱”的问题（王冬青等人，2025），其包含的复杂指标体系（如知行模型、FIAS、S-T 分析等）和晦涩的理论术语，提高了教师理解和分析结果的门槛。

知识图谱以结构化方式组织实体及其关系，为复杂推理提供支撑。随着大语言模型的发展，利用其语义理解与少样本学习能力进行知识抽取已成为知识图谱构建的重要方向（张才科等人，2024），相关研究表明大模型在实体识别和关系抽取等任务中具有较强潜力（Wang et al, 2023）。在教育领域，面向课堂教学分析报告的知识图谱构建仍较为缺乏。与此同时，知识图谱问答结合大语言模型已成为研究热点，虽然现有方法在答案真实性方面有所提升，但在深层语义理解、多步推理和全局信息整合方面仍存在不足（萨日娜等人，2022）。

基于此，本研究设计并开发了一个融合知识图谱与大语言模型的课堂教学分析问答系统（CSMS-Chat），采用 DeepSeek-V3.2-Exp 模型结合提示词工程，从半结构化教育报告中自动构建领域知识图谱，并引入“知识路径”机制，以弥合结构化图谱与语言模型之间的数据模态差异，同时采用 DeepSeek-R1 作为内容生成模型，实现更精准、可解释的教学诊断问答。

2.CSMS-Chat 系统及关键技术实现

CSMS-Chat 系统旨在解决传统 RAG 在处理高逻辑密度教育报告时的局限性。系统架构如图 1 所示，主要包含三个核心模块：基于提示词工程的领域知识图谱构建、基于混合检索驱动的语义解析以及基于知识路径的推理增强生成。

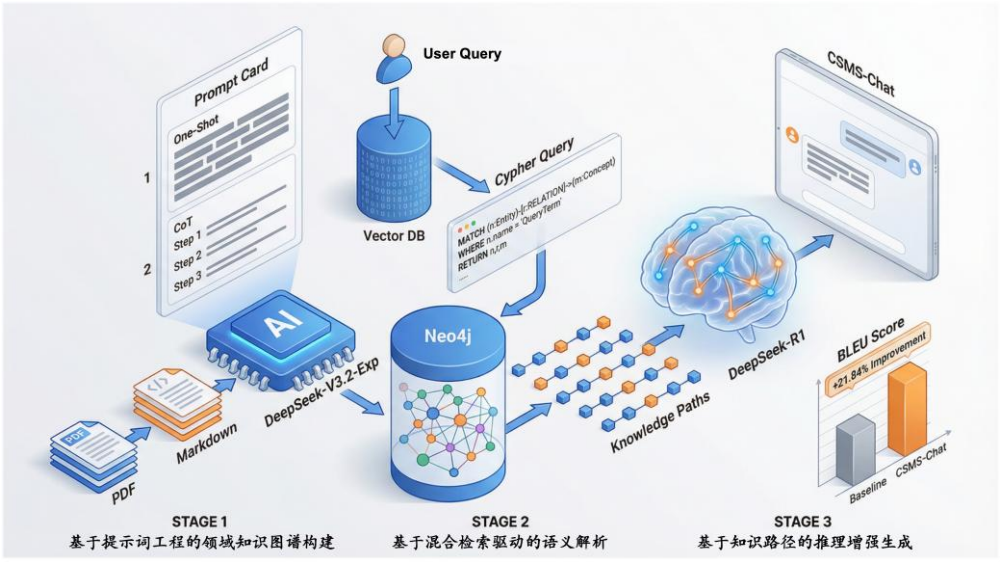


图 1 研究内容总体流程图

2.1. 基于提示词工程的领域知识图谱构建

CSMS 报告通常以包含大量非结构化文本和半结构化表格的 PDF 形式呈现，且指标之间存在复杂层级关系。为实现知识的结构化存储，本文构建了由数据预处理、模式层设计、知识抽取与图数据库存储组成的自动化流程（Sun et al, 2024），通过调用 DeepSeek-V3.2-Exp 模型，结合“任务指令—本体定义—输出约束—单一样本示例—思维链策略”的提示词模板，实现知识三元组的自动抽取，并将结果以 JSON 形式导入 Neo4j 图数据库完成存储。

2.2. 基于混合检索驱动的语义解析

针对教育分析场景中自然语言提问的多样性，CSMS-Chat 系统采用“语义解析”与“混合检索”相结合的策略，如图 2 所示。首先，系统构建了一个向量化的查询映射知识库，存储了常见自然语言问句与标准 Cypher 查询模板的键值对，即文本块（Chunks），当用户输入查询时

（例如：“本节课的教学效果怎样？”），系统基于向量模型（Embedding Model）对用户输入的自然语言问句进行向量化处理，通过与知识库匹配的方式，将自然语言问题转化为图数据库可执行的 Cypher 查询语句。其次，为确保检索的召回率与准确率，系统采用双阶段混合检索策略。初筛阶段利用向量检索快速定位与问题相关的文本块；重排序阶段利用检索排序模型（Rerank Model）对候选文本块进行精细打分，过滤掉语义相关但逻辑不匹配的噪声，最终执行最优的 Cypher 语句从 Neo4j 中检索出结构化的图谱子图。

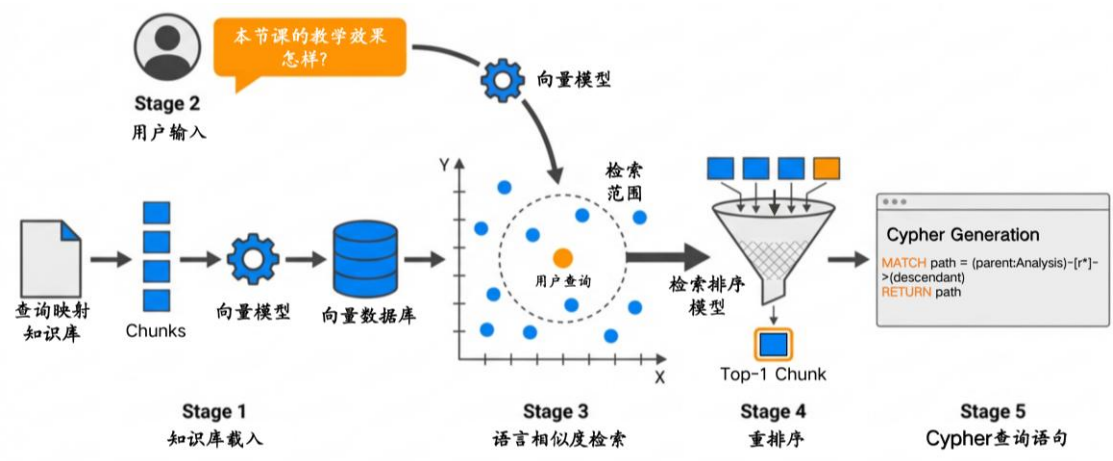


图 2 基于混合检索驱动的语义解析流程图

2.3. 基于知识路径的推理增强生成

检索得到的图谱子图虽然包含了与问题相关的实体和关系，但原始结果中往往夹杂节点编号、标签集合等元信息，不利于大语言模型直接理解。为此，本文将图谱子图进一步整理为“知识路径”，即按照问题求解所需的逻辑顺序，将相关实体、关系及其指向的事实连接成一条清晰的语义链，使原本分散在图结构中的信息转化为更适合语言模型处理的线性上下文。

在生成阶段，系统将用户问题与知识路径共同输入 DeepSeek-R1 模型，使模型能够沿着明确的推理线索关联不同教学指标，完成多步推理与综合分析，从而生成更具逻辑性、证据性和解释性的教学反馈，并降低回答中的“幻觉”现象。

3.实验与评估

3.1. 实验准备

为验证 CSMS-Chat 在课堂教学分析问答任务中的有效性，并考察知识图谱驱动的“知识路径”机制相较传统 RAG 的优势，本文选取覆盖小学、初中、高中三个学段、涉及数学、语文、历史、物理等学科的 6 份典型 CSMS 课堂教学分析报告作为测试语料。围绕各报告构建了共 115 个测试问题，涵盖基础查询、多跳查询和综合推理查询三类，以全面评估系统对不同复杂度问题的处理能力。

3.2. 实验方法及关键指标

本研究采用基准对比实验，将 CSMS-Chat 与传统 RAG 系统进行比较。基准系统以原始 CSMS 报告文本为知识来源，并使用与 CSMS-Chat 相同的生成模型 DeepSeek-R1、尽可能一致的提示词及相同模型参数，将上下文信息差异控制在是否引入知识图谱及“知识路径”机制这一关键因素上。最终，两类系统围绕 115 个问题分别生成回答。基于此，本研究设计了主客观两种测试方法：客观评价采用 BLEU 指标衡量生成答案与参考答案的匹配程度（Papineni et al, 2002）；主观评价邀请两位具有教育技术背景的专家开展单盲评分，从准确性、逻辑连贯性和教学改进实用性三个维度进行综合判断。

3.3. 实验结果分析

客观测试结果表明,在 BLEU 指标上,CSMS-Chat (0.76 ± 0.182) 显著优于基准系统 (0.62 ± 0.215),且差异达到极显著水平 ($t=5.116, p<0.001$),效应量 *Cohen's d* 为 0.675,表明系统性能提升具有中等到较高效应。

主观测试结果同样支持上述结论。专家评分显示,CSMS-Chat (8.64 ± 1.12) 显著高于基准系统 (6.69 ± 1.45),两者系统之间存在统计学意义上的显著差异 ($t=11.372, p<0.001, \text{Cohen's } d=1.503$)。专家进一步指出,传统 RAG 在处理跨层次推理问题时容易停留于零散片段,导致回答逻辑割裂。相比之下,CSMS-Chat 通过“知识路径”将分散指标加以关联,更有利于生成具有全局视角和逻辑严密性的教学诊断建议。

4. 结论

本文针对 CSMS 课堂教学分析报告专业性强、解读门槛高的问题,构建了融合知识图谱与大语言模型的智能问答系统 CSMS-Chat。该系统通过将半结构化教学评价报告转化为结构化领域知识,并引入“知识路径”机制,增强了系统对表格数值、层级关系和因果逻辑等关键信息的理解与推理能力。实验结果表明,相较于传统 RAG 方法,CSMS-Chat 能够生成更具逻辑性、全局性和证据支持的教学诊断建议,为教育数据驱动的教学反思与改进提供了新的技术路径。

尽管如此,CSMS-Chat 在图谱构建自动化程度和多模态数据融合方面仍有进一步优化空间。未来研究可继续探索端到端知识图谱构建方法,完善 RAG 检索与推理机制,并结合课堂视频、音频等多模态信息,构建更完善的课堂教学分析智能体。

参考文献

- 王冬青,陈自力,邵文豪,张粤芳,李赞坚 & 任光杰.(2025).从“负能”到“赋能”:基于 LLMs 的思维链提示设计与教研 AI 智能体构建——以课堂教学智能分析为例.中国电化教育,(03), 111-117+125.
- 中国社会科学网.(2023-06-15).广东省中小学课堂大数据报告暨数据开源发布会在穗举行..http://www.cssn.cn/skgz/bwyc/202306/t20230615_5645224.shtml
- 张才科,李小龙,郑胜,蔡家骏,叶小舟 & 罗静.(2024).基于大语言模型的知识图谱构建及应用研究.计算机科学与探索,18(10),2656-2667.
- 萨日娜,李艳玲 & 林民.(2022).知识图谱推理问答研究综述.计算机科学与探索,16(08),1727-1741.
- Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., ... & He, Y. (2025). Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. arXiv preprint arXiv:2501.12948.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).
- Sun, J., Zhang, Z., & He, X. (2024, August). Llm4edukg: Llm for automatic construction of educational knowledge graph. In 2024 International Conference on Networking and Network Applications (NaNA) (pp. 269-275). IEEE.

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., ... & Guo, C. (2025, April). Gpt-ner: Named entity recognition via large language models. In Findings of the association for computational linguistics: NAACL 2025 (pp. 4257-4275).