

基于知识图谱的个性化习题推荐算法研究

曾明星 杨玲玲* 李忠娟
广东工业大学自动化学院 广东 广州 510006

摘要：随着智能设备的普及，在线学习逐渐成为线下学习的有益补充。针对《电力电子技术》课程内容多，难度大的问题，设计基于知识图谱的个性化习题推荐系统。系统采用分层架构，结合 Neo4j 存储知识图谱数据，集成 DeepSeek 大模型支持知识抽取、智能问答及习题推荐功能。系统对课程内容进行结构化表达和可视化呈现，通过个性化习题推荐算法优化学习体验，提升学习效率。

关键词：知识图谱；电力电子技术；个性化推荐；DeepSeek 大模型

引言：

电力电子技术是采用电力电子器件对电能进行变换和控制的技术，广泛应用于工业制造、新能源发电、轨道交通及智能电网等领域。《电力电子技术》作为自动化专业的核心课程，涉及复杂的电路拓扑、波形分析以及参数计算，学生普遍反映学习难度大。

在学习过程中，由于教材、MOOC 平台与习题库之间缺乏联系，学生很难将知识点与习题、课程内容相关联。系统的推荐多依赖简单规则（如按章节顺序），较少考虑学生知识掌握程度、薄弱环节以及知识点间的相互关系，导致学生难以根据本人的学习情况开展个性化学习。习题反馈机制不足，学生遇到问题时，难以及时获取精准解答。随着人工智能的快速发展，ChatGPT、DeepSeek 等聊天机器人迅速普及，人们只需要在聊天框中输入问题，就能得到个性化的回答。这不仅解决了传统习题推荐系统交互欠缺、难以实现个性化的问题，而且能够借助大数据的分析，更好地为学生提供习题的推荐及相关知识点学习。

关于个性化推荐模型的实现方法，学者们进行了广泛研究：聚合用户特征，使用注意力机制学习知识图谱高阶潜在关系，设计融合知识图谱与注意力机制的项目推荐算法^[1]；基于学习风格量表设计个性化推荐系统并应用于教学中^[2]；利用协同过滤算法和内容过滤算法，根据学生的偏好进行推荐^[3]；利用 K-means 算法分析学生对不同数学资源的喜好，将偏好分析结果输入到深度神经网络中实现个性化推荐^[4]。

针对学生在学习过程中遇到的问题，本文提出基于知识图谱的个性化习题推荐系统，深度融合知识图谱、推荐算法与人工智能大模型，为《电力电子技术》课程提供智能化教学支持。

一、知识图谱构建与存储

（一）文档转换

为了构建知识图谱和习题推荐系统，需要将教学文档转换为易于处理的文本格式，这是数据处理的第一步。

针对 PDF 格式的文档，使用 PDFMiner 库进行解析，将提取的文本保存为 Markdown 形式。PDFMiner 是一个开源库，用于从 PDF 文档中提取文本和元数据，能够解析 PDF 的结构化内容并转化。

针对 Word 格式的文档，使用 python-docx 进行读取。python-docx 是一个 Python 库，功能是提取 Word 文档中的段落文本，并将文本保存为 TXT 格式。

（二）知识抽取

首先配置好实体类型及关系，然后按照章节和知

识点层面，对教学文档进行分块处理^[5]。按照章节分块的实现方式是：通过 `loop_files('response', 'json')` 遍历 response 目录下的所有 .json 文件，如果令每个文件对应一个章节，那么可以通过属于关系关联到对应的章节节点。按照知识点层面分块的实现方式是：由 JSON 数据中每条记录的 subject 和 object 作为实体节点，实体通过 relation 字段定义的关系连接，形成三元组（如 A-[关系]->B），最后每个三元组可视为一个最小的知识单元，用于构建图谱中的边。

接着调用 DeepSeek 大语言模型对教学文档进行知识抽取，输入数据为预处理的 JSON 文件，每个文件对应一个章节。JSON 文件中每条数据包含 subject、relation、object 标签和属性，利用大语言模型将每个实体映射为图数据库中的节点，关系映射为边，通过

系统配置。

(二) 习题推荐系统设计

融合知识图谱查询和大语言模型推荐, 可以使得推荐结果更加准确。用户在个人信息页面输入自己正在学习的内容, 系统会先基于知识图谱对用户输入内容进行关联查询, 获取知识点之间的关系网络以及知识上下文, 导入到 DeepSeek 模型中, 通过设置好的提示词引导大语言模型输出。

关联的流程如下: 将知识点序列化为 JSON 字符串, 再拼接到系统角色提示中, 告知大模型“请基于给定知识点列表推荐 10 个应学的知识点”。接着把用户的学习情况作为用户角色消息传递给 ollama_api 去调用 LLM, 对模型返回的内容用 json.loads 进行解析。如果返回内容里带有 Markdown 风格的 JSON 区块, 就使用正则表达式提取 JSON 再解析, 否则返回一个空列表。

(三) API 接口

接入 DeepSeek 中最重要的就是其 API 接口, 如果设置不当, 会导致系统部分功能出现错误。在系统代码中, @app.before_request 的 before() 函数首先获取当前请求路径, 检查该路径是否属于静态资源或预设的白名单, 是则直接放行; 否则从请求头中读取 Token 令牌, 若令牌不存在则返回错误提示用户登录, 若令牌存在则解析 user_id 保存到全局变量。

知识检索模块是习题推荐和智能问答准确与否的关键模块。根据实体名称和查询深度, 首先在图数据库中查找与之模糊匹配的所有路径, 如果没有给出实体, 则随机返回任意 100 条路径。然后遍历查询结果, 把所有节点去重后抽取出每个节点的属性, 组装到 nodes 列表里; 同时把每条关系提取为一个字典, 存入 lines 列表。最终返回一个包含所有节点和连线信息的字典 { "nodes": [...], "lines": [...] }, 作为前端绘制知识图谱的 JSON 数据。

结 论 :

本文针对《电力电子技术》课程学习中存在的习题资源分散、推荐方式单一、反馈机制缺失等问题, 设计实现基于知识图谱的个性化习题推荐系统。通过多源数据整合与结构化处理, 完成了课程知识的语义化表达, 实现了知识网络的可视化存储; 融合知识图谱的结构化推理与大语言模型的动态语义分析能力, 通过提取用户薄弱知识点生成自然语言描述作为 DeepSeek 的提示词, 由大语言模型推理结合深度知识追踪 DKT 模型实时更新用户学习状态, 有效匹配更适合的习题, 实现个性化习题推荐功能。

参考文献 :

[1] 邢峻也, 邢星, 贾志淳, 等. 融合知识图谱与注意力机制的项目推荐算法 [J]. 计算机工程与应用, 2024,60(10):173-179.

[2] 薛雅倩, 松云. 人工智能课程知识图谱构建及个性化推荐探索 [J]. 计算机教育, 2024,(06):151-155.

[3] 陈强, 贺丹, 邓美玲. 基于个性化推荐的高校教学模式探索——以 Python 程序设计通识课为例 [J]. 电脑知识与技术, 2024,20(12):131-133.

[4] 方贝, 何文博. 应用知识图谱的高等数学教学资源个性化推荐研究 [J]. 信息与电脑 (理论版), 2023,35(22): 220-222.

[5] 王志刚, 吴士泓, 李孟全, 等. 基于知识图谱的项目文档智能管理与应用系统 [J]. 现代计算机, 2023,29(03):111-120.

[6] LIU Y, HUA W, XIN K, et al. TEA: Timeaware Entity Alignment in Knowledge Graphs[C]. Proceedings of the ACM Web Conference 2023. New York, NY, USA: ACM,2023:2591-2599.

基金项目: 教育部高等学校自动化类专业教学指导委员会教育教学改革研究课题 (2024047); 广东工业大学本科教学工程项目 (广工大教字 [2024]75 号)。

作者简介 :

曾明星 (2002.10-), 女, 汉族, 贵州黔西人, 工学学士, 研究方向: 知识工程;

通讯作者: 杨玲玲 (1978.03-), 女, 汉族, 河南平顶山人, 博士, 副教授, 研究方向: 非线性控制。