

基于知识图谱的大数据采集与处理课程 教学资源整合与优化研究

周 勇 司利红 陆万里

重庆工程学院, 中国·重庆 400056

【摘 要】针对大数据采集与处理课程的评价粗放、知识割裂、资源分散问题,本研究基于DeepSeek大语言模型构建了包含118个原子知识点的三层课程知识图谱,并在超星学习通和头哥实践教学平台完成了285份教学资源的精准映射与动态更新。教学实验表明,知识图谱有效提升了教学效果,教师定位薄弱知识点耗时显著减少,学生知识点掌握度明显提高,跨模块综合应用能力增强,资源利用效率提升。该研究为数据类课程精准教学提供了可复制的范式。

【关键词】知识图谱; 大数据采集与处理; 教学资源整合

【基金资助】重庆工程学院教育教学改革研究项目“大数据采集与处理课程知识图谱建设与应用”(项目编号: JY2023103)

在“教育新基建”与“数字中国”战略的推动下,高等教育正从传统的“资源供给型”向“精准服务型”转型。2021年教育部发布的推进教育新型基础设施建设指导意见明确提出,需构建国家统一的学科知识图谱,以系统化梳理知识脉络、精准标识教学资源,从而实现因材施教与个性化学习。2023年怀进鹏部长进一步强调,教育知识图谱应成为数字化教学的核心基础设施,支撑教学模式创新与学习效果优化。

大数据采集与处理作为数据科学与大数据技术专业的核心课程,其教学内容覆盖互联网数据采集、数据预处理,具有知识点密集、实践性强的特点。然而,当前课程教学仍面临三大挑战:(1)评价粗放:学习评价以“章节”为单位,难以追踪学生对具体知识点的掌握情况;(2)知识割裂:模块化教学导致学生难以建立跨章节、跨课程的知识关联;(3)资源分散:教学资源(如课件、实验案例、MOOC视频)分散于不同平台,整合效率低下。知识图谱技术通过结构化表征知识节点及其关联关系,可为上述问题提供系统性解决方案。本研究以大数据采集与处理课程为例,探索基于DeepSeek大语言模型的知识图谱构建方法,并研究其在教学资源整合与精准教学中的应用效果,为同类课程提供可复用的模式。

1 课程现状

大数据采集与处理课程依托校企合作平台,采用真实业务场景案例教学,已形成较完善的教学文件体系。然而,通过教师访谈与学生问卷调查(样本量N=146),发现以下突出问题:

1.1 过程性评价颗粒度不足

现有评价以章节为最小单元,如数据采集模块得分70分,无法定位具体薄弱知识点,如反爬虫策略;教师需手动汇总成绩,平均耗时0.9小时/次(调研数据),且难以积累历次学习数据,导致个性化反馈缺失。

1.2 知识体系碎片化

课程模块如静态网页采集和动态网页采集间缺乏显性关联,75%的学生反馈“学完即忘,无法串联应用”。与关联课程数据库原理及应用、Python程序设计的知识点割裂。

1.3 教学资源整合效率低

资源分散于超星平台(课件、教学大纲、视频)、头哥平台(实验案例)、教师本地(实验代码),备课时间中30%用于资源搜索与格式调整(教师日志统计)。

2 知识图谱构建

从实现技术角度看,课程知识图谱的主流构建范式可归纳为四种:一是基于规则的人工构建,由授课教师先手工设计本体框架,再逐条填充实体与关系,虽然精确但效率较低;二是传统机器学习方法,借助CRF、SVM等统计模型在小规模语料上自动抽取并辅以人工校验,需要大量标注数据;三是深度学习与大模型技术,依托BERT、GPT或DeepSeek大语言模型等预训练模型实现端到端的实体-关系抽取;四是混合流水线方法,融合规则、传统机器学习和深度学习技术,实现难度较高。综合考量课程图谱的建设难度、准确度以及构建效率等因素,本课程采用DeepSeek大语言模型和人工审核的混合构建方法。

图谱构建前需要明确划分层级问题,知识图谱的层级划分并无统一标准,但根据教育技术领域的研究和实践经验,通常遵循3-5级分层原则。结合本课程的实际需求和应

用特点，最终确定采用“课程模块、知识单元、原子知识点”的三级结构设计。具体标准经过课程组二次论证：课程模块参照ACM/IEEE计算机课程体系规范；知识单元保留企业工程师建议的12个典型工作场景；原子知识点确保每个可对应1-2个课时的教学目标。这种设计既避免了过度拆分带来的问题，如违反认知规律增加学生记忆负担、使能力培养从解决问题偏离为语法记忆，又能有效控制维护成本，避免因技术迭代而频繁更新细节点。

本课程依托超星学习通平台的知识图谱功能开展建设，该平台支持通过Excel格式进行知识图谱数据导入。具体输入格式如表1所示：

表1 超星学习通知识图谱模版

节点名称	节点名称	节点名称	前置节点	后置节点	关联节点	标签

课程图谱构建以教学大纲为输入源，充分挖掘大纲中的模块名称、教学内容、教学目标以及课程重难点等结构化信息作为知识节点的基础素材。具体构建过程采用提示词工程方法与DeepSeek大语言模型进行交互，通过输入结构化提示词（如“基于教学大纲生成课程知识图谱，分为课程模块、知识单元、原子知识点三级结构，知识点120个左右，输出格式为CSV，输出列包括课程模块、知识单元、原子知识点、前置知识点、关联知识点和标签，标签标注为重点

或难点，如果前置知识点、关联知识点和标签涉及多个以分号隔开”）并经过多轮调试优化，最终输出结果经人工审核后导入超星学习通，形成符合课程要求的完整知识图谱。生成的课程图谱如图1所示：

在节点密度设计上，采用每学分20-40个节点的标准，这一设计基于Sweiller认知负荷理论的工作记忆容量研究，结合布鲁姆分类法的认知层级分布特点，并参考了ABET工程认证的学时、知识转化标准。实践表明，该设计能确保每个节点保持2-3个最优关联关系，在实现82%-90%课程内容覆盖率的同时，有效维持合理的认知负荷水平。

在完成课程知识图谱118个节点构建后，基于布鲁姆分类法的认知维度理论，建立了系统的知识点和能力映射体系。通过教学团队协同工作，为每个原子知识点标注了六级认知目标：记忆（如列举常见反爬虫策略）、理解（如解释动态渲染网页的爬取原理）、应用（如使用Scrapy框架实现增量爬虫）、分析（如诊断AJAX数据加载失败原因）、评价（如评估不同IP代理方案的性价比）和创造（如设计分布式爬虫的负载均衡方案）。这种基于网络爬虫技术特点的认知目标分级，既覆盖了基础理论知识的掌握，又强调了工程实践能力的培养，特别是将行业中的实际问题融入高阶认知目标的设定中，使能力培养与职业需求紧密衔接。

3 课程资源整合

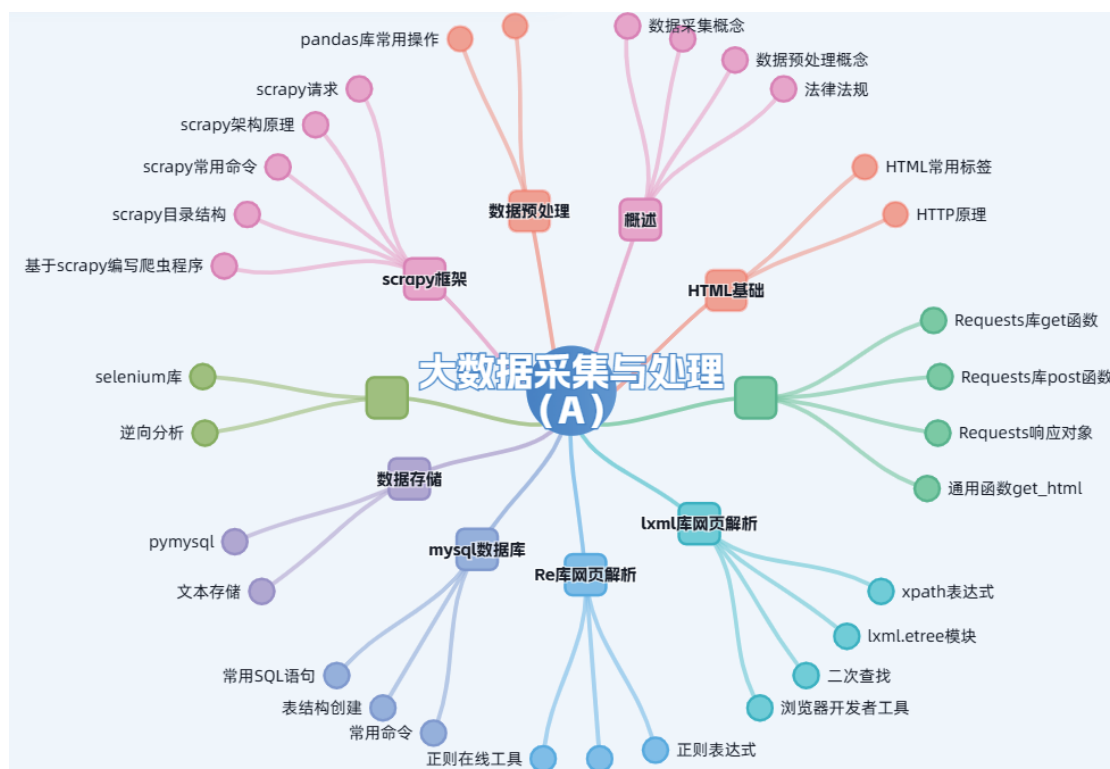


图1 课程图谱

为充分利用超星学习通平台的现有功能,课程组经反复研讨,形成一套可落地的教学资源整合方案。确立了标准化的资源组织标准,按照“课程模块、知识单元、资源类型”构建层级化文件夹结构,并配套统一命名规范。

在课程视频整合阶段,团队依托知识图谱,精准检索中国大学MOOC、B站优质资源;在严守版权的前提下,按需裁剪并完整保留原作者标识与出处。剪辑后的视频片段按知识点挂载至图谱节点,构建结构清晰、检索高效的多媒体资源库。

为支撑线上自学、前测诊断与课后拓展,围绕核心知识点重构结构化智能题库。题库以课程图谱为骨架,匹配“基础记忆、进阶应用、高阶综合”三类题型,涵盖选择、填空、案例、自动评测编程等多元形式;通过标签化与难度系数实现系统归类。前测模块精准定位知识盲区,课后拓展题则链接真实案例强化迁移,最终形成“自适应评测、精准巩固”的闭环学习支持体系。

在校企资源整合方面,依托头哥实践教学平台,开发18个贴合真实业务场景的企业典型案例,并与课程图谱能力精准映射。每个案例均配备实验指导书、一键部署的实验环境及可自动评分的评测代码,支持学生即时自测、即时诊断。每学期召开线上资源对接会,动态同步企业案例与教学需求,确保内容常新。

在平台功能应用方面,深度挖掘超星学习通内置工具潜力:通过“知识点关联”功能,将118个原子知识点与285个教学资源建立精确映射;利用标签系统为资源标注“理论/实践”属性与“基础/进阶”难度等级;平台提供的资源使用统计功能,则帮助教师实时掌握各类资源的实际使用效果,持续优化教学策略。

4 教学实践与效果分析

4.1 教学实验

2024-2025学年秋季学期,课程以2个实验班为实施对象,学生数93人,教学实施16周,并平行设置89人的对照班沿用传统章节式教学。教学流程依托超星学习通课程知识图谱与资源标签功能,构建“图谱导航、资源推送、精准评价、即时反馈”闭环:

学生登录后,在章节目录中展示本次课的知识图谱;教师提前为每个节点配套基础或进阶标签的微课、案例或实验包,平台根据学生前测得分与学习时长,通过智能分组规则自动匹配并推送对应资源;每次课内嵌3-5道微测验,正确率低于70%时触发任务再学习;学生完成头哥

平台的校企案例实验后,点击提交可进行自动评测。学期内借助腾讯会议与头哥平台共组织2次线上校企资源对接会,将3个业务案例的实验包与指导书同步更新至课程资源库,确保案例内容与产业需求同频。

4.2 效果分析

教学实施后,课程痛点得到缓解。其一,评价粒度细化:教师通过仪表盘一键导出薄弱点分布,定位耗时缩减至不足4分钟;实验班118个原子知识点的平均掌握度由前测52%升至后测85%,同期对照班仅提升约18个百分点。其二,知识关联显性化:期末跨模块综合题实验班平均得分79分,较对照班高17分,问卷显示超八成学生能清晰描述模块间依赖关系。其三,资源整合高效化:备课中资源搜索与格式调整时间占比由30%降至8%,学生对资源查找便捷性满意度达92%,分散低效问题明显改善。

5 结论与展望

本研究通过构建课程知识图谱并整合教学资源,显著提升了教学效果。评价粒度细化至知识节点,教师定位薄弱知识点耗时大幅缩短,学生知识点掌握度显著提升;知识关联显性化,学生跨模块综合应用能力增强;资源整合高效化,教师备课效率提高,学生资源查找便捷性提升。基于知识图谱的教学资源整合模式在提升教学精度、打通知识壁垒、优化资源利用方面效果显著,为数据类课程精准教学提供了可复制、可推广的范式。

未来课题将在二个方向持续深化:一是纵向延伸,将图谱节点与高阶项目制任务对接,引入校企联合命题的创新项目,进一步提升学生解决复杂工程问题的能力;二是技术迭代,结合DeepSeek大模型的持续升级,引入自动问答与代码诊断插件,降低教师维护成本,推动知识图谱向“自适应、自演化”方向发展。

参考文献:

- [1] 王佐旭. 知识图谱和大语言模型辅助新工科课程教学资源建设方法[J]. 高等工程教育研究, 2025, (01): 40-46+110.
- [2] 王洪江, 侯荣旭, 任娜, 等. 基于知识图谱的计算机类专业课程创新设计与实践——以“计算机多媒体技术”课程为例[J]. 沈阳工程学院学报(社会科学版), 2024, 20(04): 94-99.
- [3] 刘强, 徐小龙, 赵龙龙, 等. 基于知识图谱的大数据类课程PBL教学实践[J]. 计算机教育, 2023, (12): 325-330.