

李子亮, 李兴春. 《科学计算与 MATLAB 语言》课程知识图谱的构建[J]. 智能计算机与应用, 2025, 15(2): 116-123. DOI: 10.20169/j.issn.2095-2163.24082004

《科学计算与 MATLAB 语言》课程知识图谱的构建

李子亮, 李兴春

(五邑大学 电子与信息工程学院, 广东 江门 529020)

摘要: 在建设教育强国和推进教育数字化的背景下, 知识图谱在教育领域展现出巨大的潜力。针对《科学计算与 MATLAB 语言》课程内容复杂、知识点分散、学生学习难度大等问题, 研究分析该课程的知识结构和学习需求, 提出构建课程知识图谱, 以系统化、可视化的方法改进教学效果。采用 Protégé 本体编辑器构建知识图谱框架, 并利用 Neo4j 图数据库进行可视化展示。通过 Bi-LSTM+CRF 模型提取关键知识点, 利用关系图卷积网络(R-GCN)识别知识点之间的关联, 并结合基于规则匹配的方法进行属性抽取, 丰富知识图谱内容。基于该图谱, 使用前后端分离方法开发整合教学资源与个性化学习等功能的教学平台。结果表明, 基于课程专题一构建的知识图谱共包含 242 个节点, 320 条边, 覆盖 95% 以上核心知识点, 显著提升了学生的学习效率和教学效果, 为高等教育数字化转型提供了参考, 也为其他课程提供可借鉴的经验。

关键词: 自然语言处理; 知识图谱; MATLAB 语言; 关系图神经网络; 数字化教学

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2025)02-0116-08

Construction of the knowledge graph for the course "Scientific Computing and MATLAB Language"

LI Ziliang, LI Xingchun

(School of Electronic and Information Engineering, Wuyi University, Jiangmen 529020, Guangdong, China)

Abstract: In the context of building a strong education system and advancing digital education, knowledge graphs have shown significant potential in the educational field. The course "Scientific Computing and MATLAB Language" presents challenges due to its complex and dispersed content. This study analyzes the course's structure and learning needs, proposing the construction of a course knowledge graph to improve teaching effectiveness through a systematic and visualized approach. Using the Protégé ontology editor and Neo4j graph database, a knowledge graph was built and visualized. Key points were extracted using a Bi-LSTM + CRF model, and their relationships were identified through a Relational Graph Convolutional Network (R-GCN). Rule-based methods were employed to further enrich the graph. A platform was developed based on this graph, integrating teaching resources and personalized learning features. The knowledge graph for the first module contains 242 nodes and 320 edges, covering over 95% of core knowledge points, significantly enhancing both learning efficiency and teaching outcomes. This study provides a practical reference for the digital transformation of higher education and offers a model for other courses, driving innovation in the educational field.

Key words: natural language processing; knowledge graph; MATLAB language; relational graph convolutional network; digital teaching

0 引言

2022年1月, 国务院印发《“十四五”数字经济发展规划》强调推进教育新型基础设施建设^[1], 推动“互联网+教育”持续健康发展。同年10月, 党的二十大报告首次将教育、科技、人才进行“三位一体”统筹安排、一体部署, 并将“推进教育数字化”写入报告,

明确了教育数字化未来发展的行动纲领^[2-4]。

知识图谱的概念由谷歌于2012年正式提出, 是一种结构化的语义知识库, 以图结构表示知识点及其关系。Han等^[5]提出了一种基于知识增强文档级实体和关系抽取的知识图谱构建框架, 引入领域特定知识和创新的知识增强图推理网络(KEGI), 并成功应用于钢铁生产线故障报告, 实现了对生产报告

基金项目: 广东省本科高校教学质量与教学改革工程建设项目(GDJX2023013); 五邑大学教学质量工程与教学改革工程项目(JX2023012)。

作者简介: 李子亮(2001—), 男, 硕士研究生, 主要研究方向: 知识图谱及其应用。

通信作者: 李兴春(1976—), 男, 博士, 副教授, 主要研究方向: 数据处理及智能建模。Email: 249469543@qq.com。

收稿日期: 2024-08-20

哈尔滨工业大学主办 ◆ 专题设计与应用

的深入挖掘,为生产管理提供全面的解决方案。Enayat 等^[6]利用 Neo4j 构建疾病知识图谱,以提高医疗信息检索效率,并通过整合相关医学概念之间的语义关系和知识推理能力,加速临床发现并帮助开创新疾病的治疗方式。Ezekia 等^[7]提出了一种利用图结构和类似人类思维的推理方法,通过知识图谱识别和应对网络攻击,显著改进了网络安全系统的抗攻击能力,为网络安全管理奠定了自动化检测的基础。Wan 等^[8]以吴门画派为例,提出了构建多模态知识图谱的框架,利用印章和主题提取方法,构建了吴门多模态知识图谱(WuMKG),实现了跨模态检索、知识问答和可视化应用。冷敏敏等^[9]采用自顶向下的方法,整合 1 012 例失智个案的个性化照护方案,以规范的“实体-关系-实体”三元组形式构建了失智照护知识图谱,为智能化照护提供了基础和保障。丁国富等^[10]提出基于知识图谱的产教融合课程体系持续改进模式,通过精确采集产业界对高等工程教育的能力需求信息,并与课程知识点进行匹配映射,实现了产教融合的双向能力评估和课程体系的持续优化,并在西南交通大学的专业课程中应用。Yang 等^[11]通过处理超过 80 000 篇论文摘要,利用 BERT 模型进行命名实体识别和关系抽取,成功提取了超过 200 000 个高质量的实体和关系,建立了一个建筑环境领域的知识图谱,促进了城市规划和管理中的高级分析和决策。可见,知识图谱在各个领域广泛应用,并在高等教育数字化转型中发挥着重要作用,提升教育治理的科学化和智能化水平,推动教育创新发展^[12]。

当前,科学计算已成为科学研究和技术创新的重要手段,而 MATLAB 作为高效、灵活的科学计算语言,是科学和工程计算的首选工具之一。然而,《科学计算与 MATLAB 语言》课程在教学中遇到学

生需理解并应用大量复杂的数学和算法概念等挑战。由于该课程缺乏系统的知识图谱,导致知识点呈现零散,学生学习难度大,教学效果难以保障。因此,本文以课程第一章节为例,介绍知识图谱的构建及其应用,旨在通过图谱化组织知识,解决当前无图谱状态下学生难以快速理解知识点的问题,并填补这一领域现有图谱的空白,从而提升课程的教学质量和学生的学习效率。

1 知识图谱构建方法

目前,构建知识图谱主要分为自顶向下和自底向上两种方法。本研究结合七步法^[13],借助 Protégé 开发工具进行课程本体构建,采用自顶向下的方法,在建立模式层后,不断挖掘新知识以更新数据层。首先将课程分为课程目标、课程内容和考核标准 3 个本体,再将课程内容分为十个专题,每个专题对应一章内容,再细分为节、知识集和知识点,如图 1 所示。在顶层框架的指导下,逐步填充具体实例。

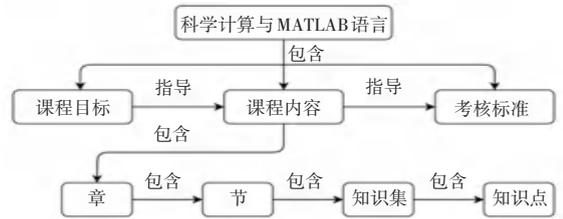


图 1 本体构建

Fig. 1 Ontology construction

构建过程中,通过人工校验和专家评审等方式进行修正和优化,确保每个节点的准确性和完备性,再对抽取的数据进行知识融合。最后,使用图数据库进行系统化建模,形成一个动态调整和优化的知识系统。在实际使用过程中,通过应用反馈不断更新和完善知识图谱,确保其在实际教学和应用中的有效性和前沿性。构建过程如图 2 所示。

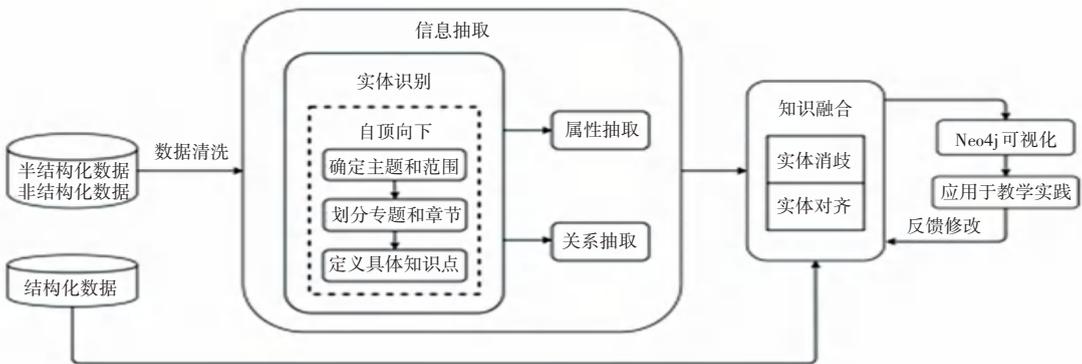


图 2 知识图谱的构建过程

Fig. 2 Construction process of knowledge graph

2 知识图谱的构建

构建《科学计算与 MATLAB 语言》课程知识图谱可以分为获取数据、实体抽取、关系抽取、属性抽取、知识融合和可视化展示 6 个步骤。

2.1 数据来源

本研究依托多种数据来源,核心资源为《轻松学 MATLAB 2021 从入门到实战》课程教材^[14]。该教材详细介绍了 MATLAB 语言的基础知识、编程技巧和科学计算方法,为知识图谱的构建提供了基础数据。此外,MathWorks 官方网站提供的 MATLAB 使用手册、功能说明、案例分析、用户指南和技术支持文档也是重要的数据来源,确保知识图谱涵盖最新的功能,保持知识的前沿性。最后,在线教育平台以及高校的公开课程提供了多样化的教学资源,为知识图谱的动态更新提供了数据支持。通过不断吸纳最新的教学资料和科研成果,知识图谱能够持续优化和完善。

2.2 实体抽取

实体识别是知识图谱构建过程中最核心的步

骤,其目的是从大量文本数据中自动识别并提取与课程相关且有意义的实体。

本研究使用 Bi-LSTM+CRF 模型与人工检验相结合的方式命名实体识别。Bi-LSTM 相较于传统 LSTM 添加了一个逆向处理层,在模型中融合了从右到左的信息^[15-16],Bi-LSTM+CRF 模型框架如图 3 所示。根据课程需求,确定“基本语法、数值计算、数据处理、编程、工具箱”5 个实体类型。首先对文本预处理,分词和去停用词,采用 BIO 标注策略进行标注;然后使用 Bert 模型初始化词向量,并将生成的词向量输入到 Bi-LSTM 层进行序列建模。前向 LSTM 处理输入序列,从左到右生成前向隐状态序列,后向 LSTM 处理输入序列,从右到左生成后向隐状态序列,并将前向和后向隐状态序列结合起来,形成每个时间步的完整上下文表示,得到带有上下文信息的隐藏层表达,并计算输出每个标签的概率得分。最后,CRF 层通过有效地约束预测标签之间的依赖关系,对标签序列进行建模,获得全局最优序列,确保预测结果的一致性。

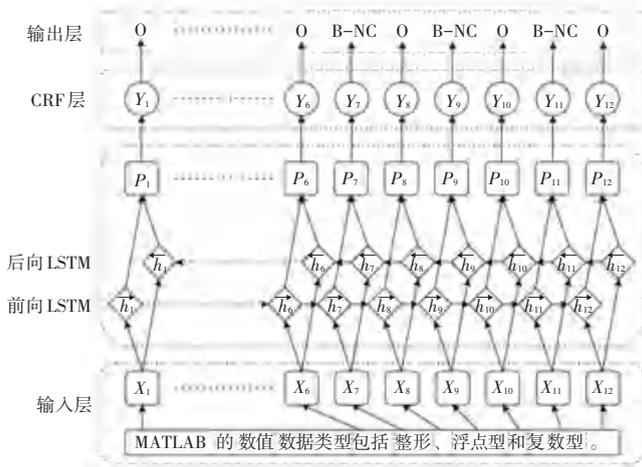


图 3 Bi-LSTM+CRF 模型框架

Fig. 3 Bi-LSTM+CRF model framework

2.3 关系抽取

在识别出实体之后,确定这些实体之间的关系。本研究采用关系图卷积网络(R-GCN)进行关系抽

取^[17-21]。根据课程实际情况,将实体间的关系定义为 7 种类型,关系类型描述和举例见表 1。

表 1 知识图谱节点关系

Table 1 Knowledge graph node relationship

关系名称	英文表示	关系说明	三元组举例
包含	Contain	知识点是另一个知识点组成部分	<字符串处理,包含,字符串查找>
前导	Precedence	在学习某个知识点之前必须掌握的基础知识或技能	<矩阵运算,前导,算术运算>
后继	FollowUp	在掌握某个知识点之后,可以进一步学习的相关知识或技能	<字符串处理,后继,文件输入输出>
同义	Synonymy	不同的术语或概念在某些上下文中表示相同的意义	<for 循环,同义,while 循环>
参考	Reference	知识点引用或依赖于另一个知识点作为背景或支持	<plot 函数,参考,数据类型>
平行	Parallel	知识点间无主次顺序关系	<strcmp 函数,平行, strcmpi 函数>
方式方法	Method	特定任务或功能的实现方法	<数据可视化,方式方法,scatter 函数>

首先,根据第一阶段提取得到的实体作为节点,参考本体模型中不同层级之间的关系建立为初始边,构建 R-GCN 的基本图结构。接下来,将 Bi-LSTM 层输出的相关节点的特征向量和根据不同关系生成的邻接矩阵输入 R-GCN 层。通过公式(1)进行多层卷积操作,双层循环遍历(不同关系类型选用不同的权重矩阵)所有关系,不断聚合邻居节点和自身的信息,随后进行归一化处理,最后经过 ReLU 激活函数输出中心节点的特征向量,从而更

新节点的特征表示。

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} \mathbf{W}_r^{(l)} h_j^{(l)} + \mathbf{W}_0^{(l)} h_i^{(l)} \right) \quad (1)$$

式中: $h_i^{(l)}$ 是节点 i 在第 l 层的表示, N_i^r 是通过关系 r 与节点 i 相连的节点集合, $\mathbf{W}_r^{(l)}$ 和 $\mathbf{W}_0^{(l)}$ 是第 l 层的可学习权重矩阵, $c_{i,r}$ 是一个归一化常数, 设置为节点 i 在关系 r 下的邻居节点数, σ 是非线性激活函数, 本研究采用 ReLU 函数。节点的一次卷积如图 4 所示。

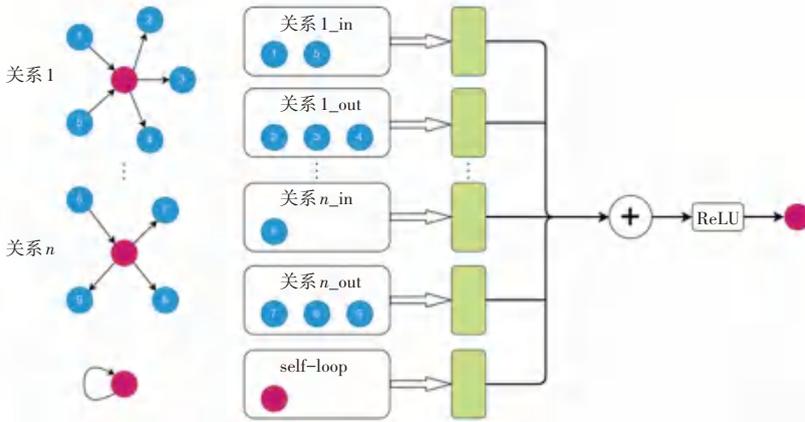


图 4 R-GCN 模型单节点更新

Fig. 4 Single node update of R-GCN model

经过多层 R-GCN 更新后,最终得到每个节点的表示向量。将这些节点表示输入 DistMult 模型解码器中,通过下式评分函数计算三元组得分,以预测新的链接和关系。

$$f(s, r, o) = h_s^T \mathbf{R}_r h_o \quad (2)$$

式中: h_s 、 h_o 是待预测节点对经过 R-GCN 更新后的实体表示, \mathbf{R}_r 是关系 r 的对角矩阵。

2.4 属性抽取

属性抽取是知识图谱构建中的关键步骤,本研究中将实体的数据属性分为 5 类:知识名称、知识内容、学习目标、学习难度和相关资源。

知识内容属性的抽取采用基于规则匹配的方法。对于“章、节”等标题,根据课程教材中的常见结构和格式,定义正则表达式和关键词匹配规则;对于知识点,分析 MathWorks 网站上的 MATLAB 使用手册,利用正则表达式和 XPath 编写具体的模板匹配规则,并用 Python 实现网页数据的抽取逻辑;学习目标属性的抽取也采用上述方法,从结构明确且格式固定的课程大纲和教学计划中提取;对于学习难度和相关资源属性,则根据教学反馈逐步完善。这种基于规则的方法既能保证准确性,又能提高抽取效率和自动化程度。

2.5 知识融合

知识融合是将从不同数据源中提取的信息整合为一个连贯且一致的知识图谱的过程,实体消歧和实体对齐是知识图谱融合的主要工作。

实体消歧旨在区分具有相同名称但实际意义不同的实体。例如,“函数”一词可能有多个含义,其可以指代 MATLAB 中的内置函数(如 sin、cos 等),也可以指代自定义函数。通过实体消歧,可以将内置函数和自定义函数区分开,并为其分别创建清晰的定义和示例。

实体对齐是将来自不同数据源但指向同一对象的实体进行匹配和统一。例如,“MATLAB”和“MatLab”这两个实体实际上指的是同一个对象。在实体对齐过程中,对于识别出的重复实体,保留一个主要实体节点,删除其余重复节点。

2.6 数据可视化

本研究利用 Neo4j 进行可视化展示。通过 Cypher 查询语言,将相关数据以三元组形式导入 Neo4j 数据库,存储为节点和边,从而直观展示知识点及其关系。基于 Neo4j 图数据库的局部显示效果如图 5 所示。在教学过程中,Neo4j 的可视化功能帮助教师更好地规划课程结构,了解学生的学习进

度和薄弱环节,从而进行针对性的教学调整。学生也可以通过知识图谱直观理解知识点关系,规划学习路径,提高学习效率。Neo4j 在知识管理、个性化教学和智能化教育中提供了有力支持。

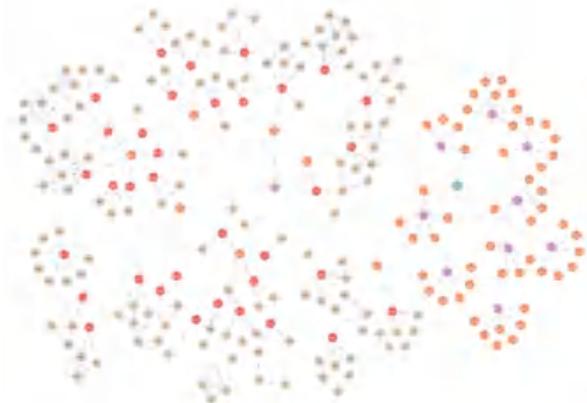


图5 知识图谱可视化展示(局部)

Fig. 5 Visualization display of knowledge graph

3 知识图谱应用

为方便师生使用知识图谱,本研究设计开发了基于《科学计算与 MATLAB 语言》课程知识图谱的教学平台。平台采用前后端分离的开发模式,Web 前端交互界面使用 Vue.js 框架,结合 Element UI 组件库,通过 Neovis.js 将图数据库中的知识图谱系统化、结构化地展示在平台界面上。知识图谱在教学平台中的局部展示如图 6 所示。后端采用 Spring Boot 框架,通过 RESTful API 实现前后端连接,确保数据传输的高效和安全,同时采用 MyBatis 进行数据库的相关操作。平台包含两种数据库:存储用户数据的 MySQL 关系型数据库和存储知识图谱数据的 Neo4j 图数据库。通过将知识图谱融入教学平台,能够显著提升课程教学效果和学生的学习体验,推动教学模式的创新。



图6 知识图谱在教学平台的展示(局部)

Fig. 6 Display of Knowledge Graph on Teaching Platform (Partial)

3.1 个性化学习路径

知识图谱将课程内容结构化为知识点网络,系统能够实时追踪和记录学生的学习进度和掌握情况。利用这些数据,系统能够精确地定位学生在知识图谱中的位置,根据掌握情况并评估与其他知识点的关联性,

精准推荐下一步的学习内容,动态生成个性化的学习路径(见图 7),确保学生在牢固掌握基础知识的基础上,稳步进阶到更高层次的知识点。这种个性化的学习方式为学生提供了精准和定制化的学习指导,充分发挥其学习潜能,进而达到更好的学习效果^[22-24]。

课程资源 扩展资源 学习路径



图 7 个性化学习路径

Fig. 7 Personalized learning path

3.2 教学资源整合

教学资源的整合可以根据学生当前学习的知识点和需求,推荐最适合的教材和学习资料,使学生能够快速找到和有效利用相关资源^[25-27]。此外,教学资源的整合还能促进多样化的学习体验和教学方法。学生可以根据个人学习偏好选择不同形式的资源,如高校开放课、期刊、图书或相关视频音频,从而提升学习的趣味性和深度(如图 8、图 9 所示)。教师可以在此基础上根据学生的学习情况实时调整和优化教学资源的使用,以达到更好的教学效果。

度以及对教学资源的利用情况,还可以及时发现教学中的问题和学生的困惑点,为课程的持续改进提供数据支持,有针对性地调整和优化教学策略和资源配置,使教学设计更加科学高效,提升教学质量和学生的学习效果。



图 8 相关资源 (期刊)

Fig. 8 Related resources (journals)



图 9 相关资源 (图书)

Fig. 9 Related resources (books)

3.3 教学质量评估

知识图谱可以帮助系统化地记录和分析课程设计、教学内容覆盖以及学习目标的达成情况,进而评估课程的结构合理性和知识点的完整性,确保课程结构合理、层次分明、教学内容覆盖到位,避免遗漏重要知识点。此外,通过追踪学生的学习进度、参与

研究中收集并分析了学生引入知识图谱前后的习题完成情况。结果表明,知识图谱的引入显著提高了学生在大多数知识点上的平均得分率,尤其是在复杂概念(如矩阵运算、算法设计等)上,得分提升尤为明显,平均增幅达到了 5% 至 10%。仅有极少数基础知识点的得分率变化较小。部分知识点的平均得分率情况如图 10 所示。

除了对单个知识点的分析,还计算了学生的整体知识掌握度,即所有知识点得分的加权平均值。统计结果显示,使用个性化学习路径和资料推荐功能的学生,较未使用这些功能的学生整体掌握度提高了约 4%。此外,个性化学习路径还有效帮助学生减少了平均学习时间,学习效率提升约 15%。

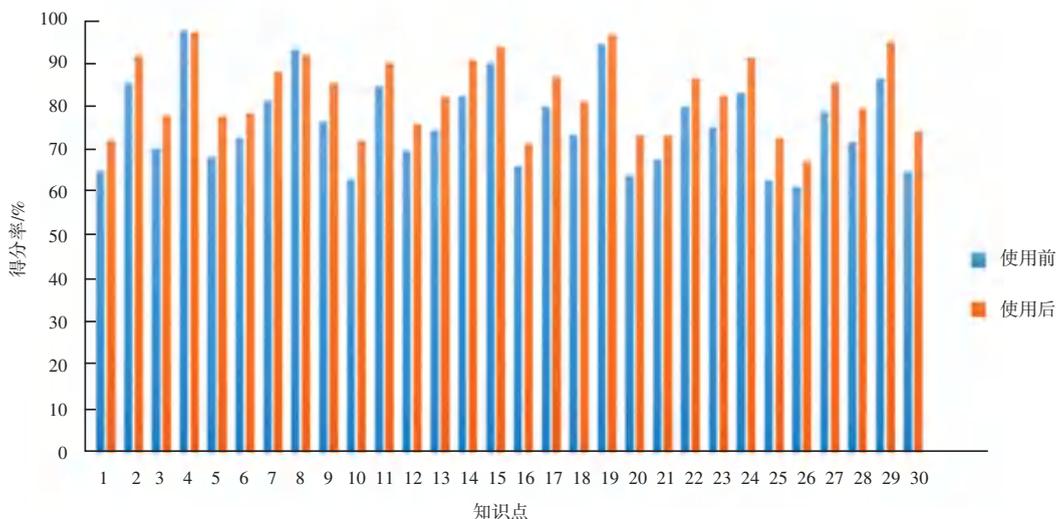


图 10 部分得分率

Fig. 10 Partial score rate

4 结束语

构建《科学计算与 MATLAB 语言》课程知识图谱不仅是对传统教育模式的革新,更是面向数字时代教育的前沿探索和实践。本文详细介绍了课程知识图谱的构建方式,并采用深度学习相关方法进行实体和关系的抽取,利用自动化技术提升构建效率。同时,还搭建了基于知识图谱的课程教学平台,介绍了平台的设计与使用,实现了个性化学习路径推荐和教学资源的动态整合等功能。不仅为学生提供了更具针对性的学习指导,也使教师能够更好地规划课程结构,了解学生的学习进度,并进行适当的教学调整,充分展示了知识图谱在教育数字化转型中的重要作用和巨大潜力。

尽管研究取得了一定成果,但当前平台在用户体验方面还有改进空间,并且随着课程内容不断丰富,知识图谱的数据还需要持续更新和维护。未来,将继续完善这一知识图谱,探索和扩展其应用范围,并融入更多先进的方法和实践案例,以应对不断变化的教育教学需求。通过持续改进和创新,知识图谱将进一步促进教育质量的提升,推动教学模式的革新,为数字时代的教育发展提供强有力的支持。

参考文献

[1] 中华人民共和国国务院. 国务院关于印发“十四五”数字经济发展规划的通知[J]. 中华人民共和国国务院公报, 2022(3): 5-18.

[2] 祝智庭,张博,戴岭. 数智赋能智慧教育的变与不变之道[J].

中国教育信息化,2024,30(3):3-14.

- [3] 冯婷婷,刘德建,黄璐璐,等. 数字教育:应用、共享、创新:2024世界数字教育大会综述[J]. 中国电化教育,2024(3):20-36.
- [4] 李惠乾,钟柏昌. 教育知识图谱:研究进展与未来发展:基于2013—2023年中文核心期刊载文的分析[J]. 计算机工程, 2024,50(7):1-12.
- [5] HAN Z, WANG J. Knowledge enhanced graph inference network based entity-relation extraction and knowledge graph construction for industrial domain[J]. Frontiers of Engineering Management, 2024,11(1):143-158.
- [6] ENAYAT R, SOMAYEH K. Building a disease knowledge graph [J]. Studies in Health Technology and Informatics, 2023,302:701-705.
- [7] EZEKIA G, JINSHUO L, AHMED A. Knowledge graph reasoning for cyber attack detection [J]. IET Communications, 2024,18(4):297-308.
- [8] WAN J, ZHANG H, ZOU J, et al. WuMKG: A Chinese painting and calligraphy multimodal knowledge graph [J]. Heritage Science, 2024,12(1):1-18.
- [9] 冷敏敏,孙月,鲁卫华,等. 失智照护领域知识图谱的构建[J]. 中华护理杂志,2024,59(4):432-438.
- [10] 丁国富,王淑营,马术文,等. 基于知识图谱的产教融合课程体系建设模式探索[J]. 高等工程教育研究,2024(2):79-83.
- [11] YANG X, ZHONG H, WANG Z, et al. BEKG: A built environment knowledge graph[J]. Building Research & Information, 2023,52(1):19-37.
- [12] 刘晓玲,王炜. 基于知识图谱的课程教学改革[J]. 中国冶金教育,2024(1):7-11.
- [13] 周育伟,杨朝红,王宏宇. 军事领域本体构建[J]. 计算机时代, 2022(9):96-99.
- [14] 刘卫国. 轻松学 MATLAB 2021 从入门到实战:案例·视频·彩色版[M]. 北京:中国水利水电出版社,2021:2-16.
- [15] ZHANG Y, XIAO G. Named entity recognition datasets: A classification framework [J]. International Journal of Computational Intelligence Systems, 2024,17(1):1-17.
- [16] 龚帅奎,蒋路茸,范骥凯. 基于 Bi-LSTM-Attention 的癫痫发作

- 检测方法[J]. 智能计算机与应用, 2024, 14(2): 69-75.
- [17] 王炼红, 林飞鹏, 李潇瑶, 等. 融入课程知识图谱的 KMAKT 预测[J]. 计算机工程, 2024, 50(7): 23-31.
- [18] 张欢, 李卫疆. 基于类型注意力和 GCN 的远程监督关系抽取[J]. 计算机工程与科学, 2024, 46(2): 316-324.
- [19] 刘铭, 何利力, 郑军红. 融合多源异构数据的图卷积神经网络混合推荐模型[J]. 智能计算机与应用, 2024, 14(2): 1-8.
- [20] 孙文洁, 李宗民, 孙浩森. 基于神经网络的多智能体强化学习值函数分解方法[J]. 计算机工程, 2024, 50(5): 62-70.
- [21] 姚砺, 魏钰菁, 万燕. 基于密集图卷积和注意力的攀爬行为识别技术[J]. 智能计算机与应用, 2024, 14(6): 50-55.
- [22] 刘凤娟, 赵蔚, 姜强, 等. 基于知识图谱的个性化学习模型与支持机制研究[J]. 中国电化教育, 2022(5): 75-81.
- [23] 赵万祥, 李滔, 刘强, 等. 以活动为导向的有机化学知识图谱构建与实践[J]. 化学教育(中英文), 2024, 45(4): 113-120.
- [24] 赵宇博, 张丽萍, 闫盛, 等. 个性化学习中学科知识图谱构建与应用综述[J]. 计算机工程与应用, 2023, 59(10): 1-21.
- [25] 吴昊, 徐行健, 孟繁军. 课程资源的融合知识图谱多任务特征推荐算法[J]. 计算机工程与应用, 2021, 57(21): 132-139.
- [26] YANG S, CAI X. Bilateral knowledge graph enhanced online course recommendation [J]. Information Systems, 2022, 107: 102000.
- [27] GUAN Honglei. An online education course recommendation method based on knowledge graphs and reinforcement learning [J]. Journal of Circuits, Systems & Computers, 2023, 32(6): 1-20.