

新工科背景下《大数据原理及应用》课程项目式教学设计与实践研究

王晨莎 李宇 刘丽景

(西安培华学院 陕西西安 710125)

摘要: 本文从新工科育人要求出发,按照项目式教学法设计《大数据原理及应用》课程,重点阐述教学设计内容与实践过程,以期今后的同类型教学设计与实践提供实际参考意义。

关键词: 新工科;大数据;项目式教学;教学设计

1 引言

新时代,工科育人要坚持新理念、新模式、新课程,坚持守正创新,创新人才培养的理念、模式和措施^[1]。

《大数据原理及应用》作为一门传统工科转型升级的新课程,其教学理念和内容、教学方法和活动还处于探索阶段。课程以计算机类、电子信息类学生为对象,以大数据相关的的基础知识和核心技术为核心,力图引导学生掌握大数据相关技术的原理与应用。课程涉及 Linux 基础、数据库原理及应用、多语言编程设计、数据挖掘算法原理、网页前端设计等内容,知识范围广,教学内容多,既有理论,又需实践,传统的教学方法难以满足本课程的教学要求^[2]。

项目式教学作为一种工学结合、任务驱动、目标导向的新型教学模式^[3],遵循“以学生为中心”的理念,通过实施一个完整项目工作而实施教学活动。《大数据原理及应用》课程通过借鉴项目式教学法的设计与实践,重构课程教学内容、融入案例教学、激发学生兴趣,促进师生互动和合作,有助于课程目标的达成。

2 教学方案设计

在项目式教学中,将一个大项目进行拆解和整合,确保项目涵盖课程的知识 and 技能。

1.明确教学目标,选择项目并拆解

明确课程的教学目标,梳理学生需要掌握的知识点和技能点,构建认知型的知识体系。选择与课程主题相关、难度适中的项目,即某市天气分析及预测探索性分析项目为例,以教学目标为导向,将项目进行拆解和整合。项目分解的细致,对于知识点和技能点的教学管理和控制越有利,以便学生能够逐步完成。但过细的分解造成学生对于项目的认识不完整,学生往往因任务繁多而畏难情绪加重,因此,本项目综合考虑拆解为5个子项目。

(1) 搭建大数据处理平台

Hadoop 基于 Java 语言开发,首选运行在 Linux 系统之上。围绕 Hadoop 开发环境的搭建,安装 Linux 虚拟机并配置 Java 环

境。大数据集群结点间的频繁通信,需借助 SSH 免密码登录,将目标节点的公钥添加到本地节点的信任密钥列表中,获得授权。再进行 Hadoop 本地模式的安装,伪分布式配置修改 core-site.xml 文件和 hdfs-site.xml 文件。最后,通过词频统计任务验证伪分布式 Hadoop 集群可用。

(2) 采集数据并预处理

大数据技术的学习和研究需要借助大量数据作为实验材料,而数据被称为“软黄金”资产^[4],其资源价值巨大。为获取大量真实数据,采用计算机+网线,即可实现网络爬虫,是一种低成本快速有效的数据获取方式。利用当前热门的 Python 语言,编写爬虫程序,对目标网页的有效主题进行提取分析,爬取大量真实数据。获得某市近一年天气的所有相关数据,并进行预处理,保证后续数据分析的质量。

(3) 数据计算和数据存储

Spark 相较于 Hadoop 平台的 MapReduce,其计算效率和优势显著。Spark 在处理大规模数据时,需要从存储系统中读取数据,HDFS 作为 Hadoop 生态系统的默认存储系统,可以提供稳定、可靠的数据存储服务。在 YARN 上部署 Spark^[5],让两个框架共享一个集群和资源,提高资源利用率。因此,在子项目 1 伪分布式 Hadoop 集群搭建完毕的基础上部署 Spark 计算框架。

HBase 具有海量存储、列式存储、高并发、稀疏性的优点,为大数据存储提供强大的支持^[6]。子项目 2 获取的 5 万多条天气数据经过 Spark 计算,结果存储在 HBase 中。在安装 HBase 之前,确保已经正确安装和配置 Hadoop,确保版本兼容。

(4) 大数据分析环境和算法实现

Scikit-learn 作为 Python 的第三方库,能够提供丰富的机器学习算法和工具^[7]。其 API 简单易用,学生可以轻松地入手实践。以入门级的鸢尾花数据集为处理对象,实现分类、回归、聚类算法,让学生初尝机器学习的魅力,了解机器学习的基本流程和常用算法。同时,调用 HBase 数据库中的房天气数据建立模型。

(5) 数据可视化 Python

Seaborn 以 matplotlib 为基础进行二次开发,学生在已有编程基础上,更方便创建复杂图像和美观视觉效果^[9],使得受众更加清晰理解数据特征。将子项目 4 中建立的模型进行可视化图表展示。

2. 整合子项目并制定实施计划

拆分子项目之后,需要将子项目进行整合,形成一个有机的整体。整合的过程中,要确保各个子项目之间的逻辑关系清晰,形成一个连贯的学习路径,即图 1 所示。

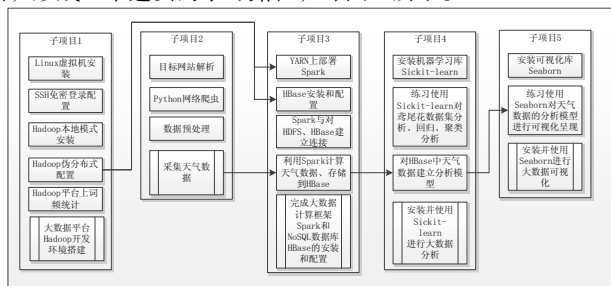


图 1 项目学习路径概览图

根据拆分和整合后的项目,制定详细的实施计划,包括每个子项目的具体任务、时间安排、所需的资源等。

3 教学实践过程与反馈

按照上述项目式教学法的教学设计方案进行教学实施过程。

首先,注重实践与理论相结合。大数据课程中理论知识是基础,仅仅停留在理论层面是远远不够的。通过项目学生可以将理论知识运用到实践中,更好地理解和掌握这些知识和技术。

其次,强调知识结构的系统性。明确项目目标和预期成果,分解为各个子项目、子任务,学生在逐步完成的过程中构建完整的知识体系。大项目的分解可以在不同里程碑上增强学生的自信心,小目标实现带来的成就感让学生感到充满激励。学习路线的明确能确保学生在这个过程中不迷失,沿着知识结构脉络不断前行,更有勇气面对未来的挑战。

再者,引导学生分组进行主动探索。一改往日传统的“教师讲,学生听”的模式,为课堂注入了极大的活力。并增强学生的团队协作能力,在探索过程中,同学们不断验证并修正实践方法和步骤,培养自主学习能力和解决问题能力。必要时给予学生适当的指导,帮助学生解决遇到的问题。

同时,作为教师,定期收集学生的反馈,对项目进行必要的调整,以确保项目能够有效地覆盖课程的知识 and 技能。关注学生在过程中的表现,引导学生思考解决问题的方法,评估他们在实践中的表现和技能掌握情况。了解学生对知识和技能掌握情况,对整个项目的实施过程进行反思和改进。

4 教学评估与总结

课程采用项目式教学以来,经过两个学年的实践和应用,取得了较为明显的成效。根据麦克斯教学质量管理平台相关统计数据得知,2022 年课程有 4 个线下课堂,学生评价应评人次 197 人,已评人次 172 人,参评率 87.31%,得分 88.87; 2023

年课程有 3 个线下课堂,学生评价应评人次 115 人,已评人次 61 人,参评率 53.04%,得分 90.79 分。

学生的反馈如图 2 所示,学生评教内容:“注重教学和案例分析”、“思路清晰、环节紧凑、重难点突出、设计合理”、“根据知识结构特点,重点突出、曾测分明,理论和实际相结合”、“体现知识结构,突出学生活动设计,课堂结构完整”等。说明项目式教学方式能够激发学生的学习兴趣 and 动力,提高学上的学业表现和综合能力。



图 2 麦克斯教学质量管理系统（截取部分）

5 结束语

项目式教学注重实际应用、主动学习、个性化学习、技能培养和解决问题的能力培养等,受到学生的欢迎。

项目式教学虽然具有诸多优点,但在实践中也存在一些不足之处。项目式教学实施难度大,需要学生在课外投入较多的时间和精力,由于学生的个体差异,可能导致部分学生无法适应项目式教学的方式,需要采取其他教学方法来弥补项目式教学的不足。项目式教学需要一定的教学设备和资源支持,需要具备较高的教学管理能力,以确保项目的顺利进行和学生的有效参与。

参考文献:

- [1]钟登华.新工科建设的内涵与行动[J].高等工程教育研究, 2017(3):6.
- [2]王岩,杨森,黄岚,等.大数据分析与应用课程体系构建[J].计算机教育, 2020(2):4.DOI:CNKI:SUN:JYJS.0.2020-02-008.
- [3]李雪威,王文俊,郅海霞,等.新工科项目式教学多元融合评价方法研究——以“智慧社会与大数据智能”课程为例[J].高等工程教育研究, 2023(6):27-33.
- [4]石秀峰.数字化时代下的数据安全治理[J].数字经济, 2022(3):4.
- [5]纪贵.Spark on Yarn 模式的电信大数据处理平台[J].计算机应用文摘, 2023, 39(13):52-54.
- [6]王帅.HBase 数据库评测关键技术的研究[D].哈尔滨工业大学 [2024-01-21].DOI:CNKI:CDMD:2.1015.982151.
- [7]Ksieniewicz P, Zyblewski P.Stream-learn – open-source Python library for difficult data stream batch analysis[J].Neurocomputing, 2022(Mar.14):478.
- [8]Weiss C J.Visualizing protein big data using Python and Jupyter notebooks[J].Biochemistry and molecular biology education, 2022.

基金: 西安市科技计划项目-高校院所科技人员服务企业项目 (22GXFW0098)