

大数据背景下《数据分析实践》课程教学研究

程 静, 闫 嘉, 陈 睿

西南大学人工智能学院, 重庆
Email: 42546062@qq.com

收稿日期: 2021年4月20日; 录用日期: 2021年6月2日; 发布日期: 2021年6月9日

摘 要

“大数据”背景下数据科学专业核心课程的实验教学环节, 在培养学生具备较强实践能力和自主学习能力, 具备一定的实际问题求解能力以及良好团队协作精神等方面, 发挥着举足轻重的作用。本文就西南大学的数据科学与大数据专业下设的核心课程《数据分析实践》进行研究, 依次探讨了实验内容体系建设、实验教学方法改革、实验平台构建以及课程评价手段改革等方面, 最后采用问卷的方式调查了学生的学习效果。结果显示采用项目驱动、多种教学模式相结合的实践课程的教学方式, 能够较好满足数据科学专业培养目标, 有利于学生更好地适应未来工作的挑战, 培养出具备终生学习、主动学习和协作学习精神的高素质人才。

关键词

数据科学与大数据, 数据分析, 实验课程教学改革, 项目驱动的教学

Studies on Teaching Reform of *Data Analysis Practice* Course under the Background of Big Data

Jing Cheng, Jia Yan, Rui Chen

School of Artificial Intelligence, Southwest University, Chongqing
Email: 42546062@qq.com

Received: Apr. 20th, 2021; accepted: Jun. 2nd, 2021; published: Jun. 9th, 2021

Abstract

In the context of “big data”, the experiment courses of the data science major will play a very im-

portant role in cultivating students to have strong practical ability, independent learning ability, practical problem-solving ability, and good teamwork spirits. This paper studies the core course—*Data Analysis Practice* in the major of Data Science and Big Data at Southwest University, and the construction of experimental content system and experimental platform, the reform of experimental teaching methods and course evaluation methods have been discussed there. Finally, a questionnaire is used to investigate this effect of teaching reform in the experimental course. The results of this survey show that the project-driven method combined with multiple teaching strategies, can better meet the training objectives of data science major, help students better adapt to the challenges of future work, and cultivate high-quality talents with the spirit of lifelong learning, active learning and collaborative learning.

Keywords

Data Science and Big Data, Data Analysis, Teaching Reform of Experimental Courses, Project-Driven Teaching

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着信息技术和互联网技术的飞速发展，各行各业积累了海量异质数据。如何从数据中发现知识，已经成为各个行业的共性问题，并由此催生了“大数据”、“数据科学”等学科，并成为新兴技术革命的核心驱动力之一。我国已经从国家战略层面对大数据相关产业进行了布局。与我国高速发展的大数据应用形成鲜明对比的是专业人才的极度匮乏。2017年1月，国家工信部编制了《大数据产业发展规划》，大力支持高校探索建立大数据领域专业型人才和跨界复合型人才培养机制[1]。之后全国各大知名高校，如北京大学、中国人民大学、复旦大学、电子科技大学等先后申请增设数据科学与大数据本科专业。截止2018年获批新增的数据科学与大数据专业的高校达到283所，这其中包括西南大学人工智能学院的“数据科学与大数据”专业。虽然大数据相关学科呈现蓬勃发展态势，但在专业建设方面还存在以下问题：第一，大数据科学在我国还处于起步阶段，各高校对于该新增专业还存在专业定位模糊、课程体系不完善、考核方式不够清晰、实验室建设不到位等问题[2][3][4][5]。第二，国内高校对于该新增专业鲜有提出具有本校特色的办学思路。大数据的专业知识能够在多个领域、多行业中得到广泛使用，但是各高校很少能够依据本校已有积淀较深的学科与数据科学专业进行的融合，没有形成具有学校文化的数据学科。第三，由于是新设置专业，因此在课程建设、教学资源建设、实验室建设等方面还存在大量工作，几乎从零开始。

本文聚焦西大人工智能学院下设数据科学与大数据专业的核心骨干课程——《数据分析实践》，以西南大学计算机科学与技术、统计学两个一级学科博士点为依托，探讨拥有“西大特色”的课程体系设置、实验环节设计、实验平台建设、教学方法改革等方面的问题，以培养具有良好的综合素质、系统的理论知识、较强的实践和创新能力的新型卓越工程人才为目标，最终为整个专业的建设和发展提供探索和改革之路。

2. 《数据分析实践》课程定位与现状

由于数据科学蓬勃的发展态势，以及西南大学在智能科学、统计学、心理学等传统学科上的优势，

学校于 2019 年创建人工智能学院，新增“智能科学与技术”以及“数据科学与大数据”两个专业，以支持学校在这两个方向上进行高水平的教学和科研。数据科学与大数据专业所制定的培养目标是以“理论厚实、能力本位、市场需求”为导向，采取“校企政协同、产教研融合”的人才培养模式。即，数据科学专业重视学生的实践环节，强调学生的实践活动能够与社会需求相结合，强调学生能够面向各行业的海量数据进行分析，数据科学专业所培养的学生要能够服务社会实践，满足国家对大数据人才的急迫需求，如图 1 所示。因此，实践课程的开设、实验环节设计、实验平台的开发等在本专业在建设和发展过程中将起到非常关键的作用，具有特别重要的地位。

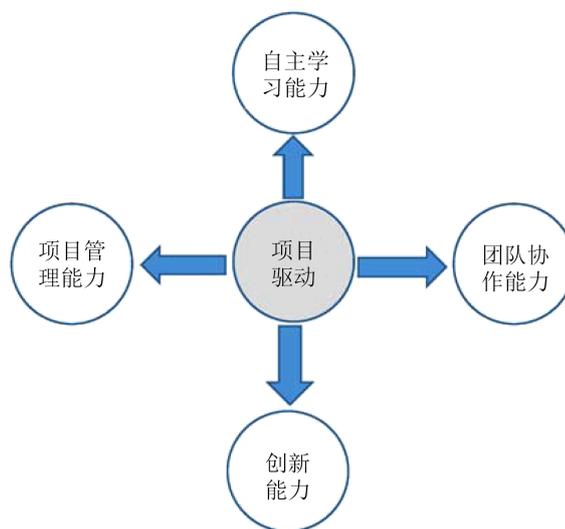


Figure 1. Training goal
图 1. 人才培养目标

《数据分析实践》课程是一门以 Python 语言为基础的数据分析实践课程，该课程学时 54，全部在实验室讲授，课程的重点在于将数据分析的基本原理穿插在多个案例的讲解中，通过项目驱动的方式，引导学生完成指定任务，从而掌握数据分析的基本方法和基本理论。因此，课程中实验案例的选择、实验环节的设计、教学资源的丰富补充、具有创新精神和主动精神的实验教学方法的推动，以及课程评价体系的革新，都将对学生最终成就这门课程起到非常关键的作用。

承担该课程的教学团队，已经从事了数据分析相关课程的教学工作多年，积累了较为丰富的教学经验。本文作者在 2018 年于美国伊利诺伊大学香槟分校(UIUC)信息学院进行访问交流，主要选修了《统计学》、《Python 程序设计》和《数据可视化》等课程，并在外导 Linda Smith 的带领下，对图书情报系统的数据分析的相关课题进行了一定的实践。回国后与奥克兰大学统计系的 Russell Miller 教授合作，承担了全英文课程《高级回归分析》的理论讲授和实践任务，在 2019 年下半年独立承担了《数据可视化》课程。不仅如此，团队教师全部教师均来自人工智能学院，该学院是由多学科的教师队伍组成，无论在理论或者在工科实践方面均有卓越的成果，这也为本课程的建设提供了多学科交叉的土壤。

3. 《数据分析实践》课程教学改革的具体措施

3.1. 教学目标

《数据分析实践》是训练学生数据分析基本素质和实践能力的核心课程之一，该课程的教学目标为：将数据科学的基本概念和基本原理融会至一个又一个案例(项目)教学中，以项目驱动的方式，让学生通过

Python 语言实现项目, 由此掌握数据科学领域的核心概念、关键算法, 从而达到训练学生面向大数据领域问题的分析和求解能力, 锻炼学生们的自主学习、协作学习的精神, 训练专业素养和职业精神[6]。

3.2. 实验内容体系建设

根据以上教学目标, 该门课程的关键在于以实践项目推动理论学习, 因此基于实践和项目驱动方式的实验内容体系建设是课程得以实施的关键。为了将数据分析的基本方法和基本原理拆解至不同的项目中, 我们将实验内容进行模块化划分, 如图 2 所示, 并构建实验教学体系, 如表 1 所示。

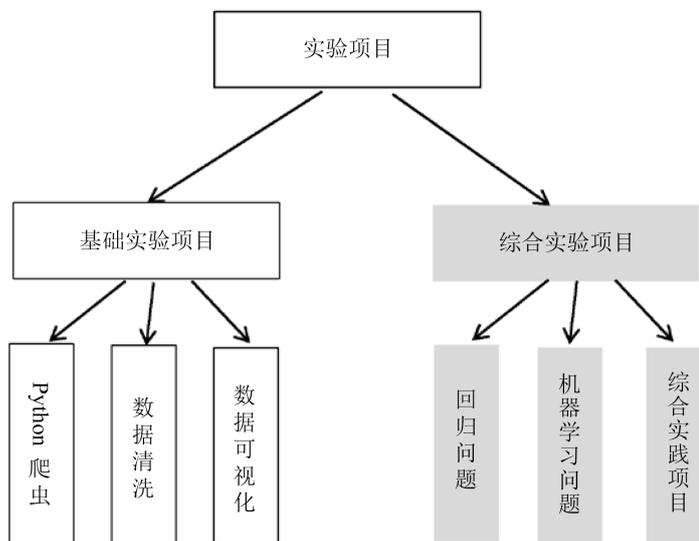


Figure 2. The system structure of experiment contents

图 2. 实验内容体系结构

Table 1. The system of experiment contents

表 1. 实验内容体系

序号	实验内容	教学要求	案例/项目	学时
1	爬虫	1) 掌握数据获取的常用方法 2) 掌握数据存储的基本格式 3) 掌握 HTTP 协议的基本概念 4) 掌握简单 Python 爬虫的方法	1) COVID-19 数据的爬取 2) 新闻数据的爬取 3) 微博数据的爬取	6
2	数据清洗	1) 掌握数据清洗的基本概念 2) 掌握 pandas 包的使用技巧	1) COVID-19 数据的清洗 2) 酒品消耗数据初步探索	6
3	数据可视化	1) 掌握常用的数据可视化手段 2) 掌握 matplotlib 包的使用 3) 掌握 seaborn 包的使用	1) COVID-19 数据的可视化 2) 亚马逊雨林数据的可视化	12
4	回归分析	1) 掌握线性回归的基本概念 2) 建立回归模型 3) 回归诊断分析	1) 使用回归构建房价预测系统 2) 使用回归模型预测鲍鱼年龄	9
5	机器学习算法	1) 掌握机器学习的基本概念 2) 掌握 sklearn 包的使用	1) 基于机器学习的宫颈癌诊断 2) 基于机器学习的心血管疾病诊断 3) 基于 sklearn 的文本分类	9
6	综合实践项目	1) 搜集资料 2) 项目实现 3) 完成一份报告和 PPT 展示(以小组为单位, 2 人/组)	1) 基于卷积神经网络的人脸识别 2) 人群接触网络中的 SIR 疫情模拟 3) 基于分类算法的肝病诊断	12

3.3. 实验课程教学方法改革

传统的实验教学模式，往往上教师在讲台上演示代码，学生按部就班复现教师所给代码。这种方式会必然造成学生照抄照搬知识后，仍然无法将这些案例内化成为自己的能力。学生只是被动模仿，缺乏主动思考、主动搜索答案、自主学习和创造性解决问题的能力。为此，在《数据分析实践》课程中我们将积极探索多元化的实验教学方法[7] [8] [9]。具体措施包括：

1) “面授” + “线上”的混合教学模式

教师除了在课堂上对基本知识点和教学案例进行讲解之外，还需要给学生布置大量的课后阅读材料，如微课、MOOC、案例、代码、报告和博客等，任课教师也可以根据情况制作一些教学短视频作为课后补充材料，这些资源可以利用QQ、微信、钉钉、超星云平台等向学生进行发布。线上资源的学习难度分为“必选”和“自选”两个层次，“必选”要求每一位学生必须自学掌握，“自选”部分学生可以根据实际情况进行选学。

2) 项目驱动式教学模式

本课程是以项目作为基本的教学单元，通过案例的讲解和演示，将数据分析的核心概念、主要思想和核心算法教授给学生。这种授课方法好处在于，一方面能够加深学生对数据分析基本原理的理解，加强学生的实际动手能力，另一方面让学生了解项目开发的基本流程(从问题背景调研、数据采集、问题分解、逐步求解，以及项目总结)，初步了解项目中的进度规划、人员协作、任务分解等内容，训练学生们的团队协作、自主学习的精神。

3) 协作学习和探究学习的教学模式

在课堂的教学活动中，教师也要调整教学的方式，不能再采用“填鸭”的灌输模式，而应该由以“教师为中心”的讲授和演示模式，转变为以“学生为中心”的启发式教学模式。教师在课堂活动中尽量以引导者而非灌输者的身份出现，鼓励学生提出开放式问题，鼓励学生自行搜集问题的求解方案，鼓励学生自行实现问题解决，在学生遇到无法解决问题的时候再进行辅导，在师生之间实现平等、协作的氛围，共同完成实验任务和课程项目。

以上三种教学方式的根本目标在于，训练学生的实际动手能力，培养学生自主学习和终身学习的精神，鼓励他们进行创造性思维，培养他们的团队意识，为后续的实习实训课程奠定基础，为实现本专业的培养目标——“理论厚实、能力本位、市场需求”奠定基础。

3.4. 课程实验平台搭建

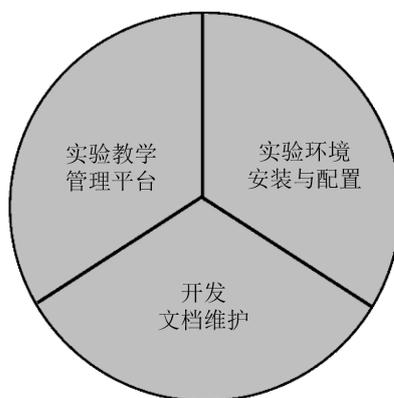


Figure 3. Three tasks of experimental platform construction

图 3. 实验平台构建的三个工作

课程实验平台的建设需要三个部分的工作，如图 3 所示。

首先需要引导学生安装和配置项目运行环境。本课程所使用的教学语言是 Python，所有项目均需运行在 Python 环境下。实验室需要安装两个开发平台，一是 Python IDLE，二是 Jupyter Notebook。其中 Python IDLE 是 Python 安装包自带的一个轻量级集成开发环境，采用编译方式运行，对系统资源消耗极少，开发环境非常简洁，缺点是对编程人员的支持较少，只适合中小规模的程序开发。Jupyter Notebook 是基于网页的交互式开发环境，可以一边运行代码，一边展示结果，一边进行文档撰写，已经成为数据分析领域最为主流的开发环境之一，因此本课程项目主要在 Jupyter Notebook 环境下运行。

实验教学管理平台的主要任务包括：1) 发布各种课程资源，如上课演示的 PPT、代码、案例讲解、课后资源的链接等；2) 为学生提交实验报告和作业提供接口；3) 展示优秀的作业和报告；4) 提供课程的交流讨论区域。这部分工作拟采用西南大学的超星云平台实现，部分功能也可以通过 QQ 群或微信群实现。

实验文档包括实验环境的安装和维护文档、学生的作业和实验报告，学生们进行项目展示的 PPT，任课教师所创建的各种文档和视频等，这部分资料需要加以保留存底，以备后来者继续参考借鉴。

3.5. 课程评价手段改革

科学的课程评价体系是实现课程目标的重要保障，是实现教学全过程有效监督的重要手段，能够帮助学生们在课程的学习过程不断体验进步与成功，获得学习的内驱力，并使教师获取教学的反馈信息，从而及时调整教学策略和教学方法。教学改革实践表明，传统的只重视结果的评价机制已经不能适应“新工科”的培养目标和教学理念。为此，在本门课程中我们将从评价方法、评价目标、评价内容和评价指标四个方面对传统实验课程的评价体系进行改革[10][11][12]。

Table 2. The curriculum evaluation system of data analysis practice

表 2. 《数据分析实践》课程评价体系

评价方法	评价目标	评价内容	评价指标	占比
形成性评价	考察学生的学习态度、基本知识点的掌握情况，帮助学生建立学习内驱力，为完成期末复杂的综合实践项目奠定基础	单元作业 课堂提问 课堂考察 考勤	1. 学习态度 2. 实践能力 3. 书写格式	20%
总结性评价	考察学生分析问题和解决问题的综合能力，训练学生对项目机制的理解和感受	综合项目验收： 1. 项目开题报告 2. 项目进度 3. 人员分工报告 3. 项目展示 4. 项目结题报告	1. 态度 2. 资料文献的准备 3. 实践能力 4. 书写格式 4. 项目完成度 5. 项目结题陈述	30%
	综合考察本课程基本知识和基本原理的掌握情况	期末考试	笔试成绩	50%

在评价方法上，我们将采取形成性评价和总结性评价相结合的机制。形成性评价伴随在整个课程的学习过程中，注重评价学生在学习过程中表现出的学习态度和um学习能力，能够帮助学生建立学习的内驱力，同时也为教师的教学活动提供及时的反馈。因此，形成性评价的考查内容是学生所提交的作业、考勤以及课堂的发言情况，这部分成绩在总成绩的占比为 20%。总结性评价分为两个部分，一个是期末学生需要完成一个有一定难度的综合性项目，另一个是需要参加期末的闭卷考试。综合性项目着重考察学生的实践能力、问题分析求解能力以及对本课程基本原理的融会贯通能力，同时将项目运行机制贯穿在实验报告的撰写过程中，让学生了解项目的人员分工、进度安排、项目立项、项目展示、项目文档撰写

等内容, 这个部分占比为 30%。总结性评价的另一个内容是期末考试, 这个部分占比 50%, 主要考察学生对整个课程中基本知识和基本原理的掌握情况, 以闭卷形式进行考察(表 2)。

4. 教学改革实施效果分析

我们对在 2020 年度选修了《数据分析实践》课程的 115 位学生展开了问卷调查[13], 问卷调查结果如图 4 所示。图 4(a)显示了课程结束后分别取得“优”、“良”、“中”、“及格”和“不及格”成绩的学生比例。结果显示采用改革后的教学方法和课程体系让 40% 的学生能够较好掌握课程内容和数据分析的基本方法, 但仍然有 7.8% 的学生不能够很好掌握该门课程, 这与课程难度以及学生个体本身的能力和态度有关。图 4(b)显示了教学改革后, 学生对“知识点掌握”、“提升动手能力”、“增加项目经验”、“增加学习趣味性”四个维度进行评价的情况。结果显示, 成绩较好的学生在这四个维度上均有较好体验, 比如对于成绩优秀或者良好的学生而言, 他们均认为教改后的课程能够更加有效地促进知识点的理解、锻炼动手能力和实践能力, 增加学习的趣味性和提升项目管理经验。但是对于不及格的学生, 在这四个维度的体验都较差, 这充分说明他们不能很好掌握知识点, 学习态度不够端正, 较难融入到改革后的课程中。

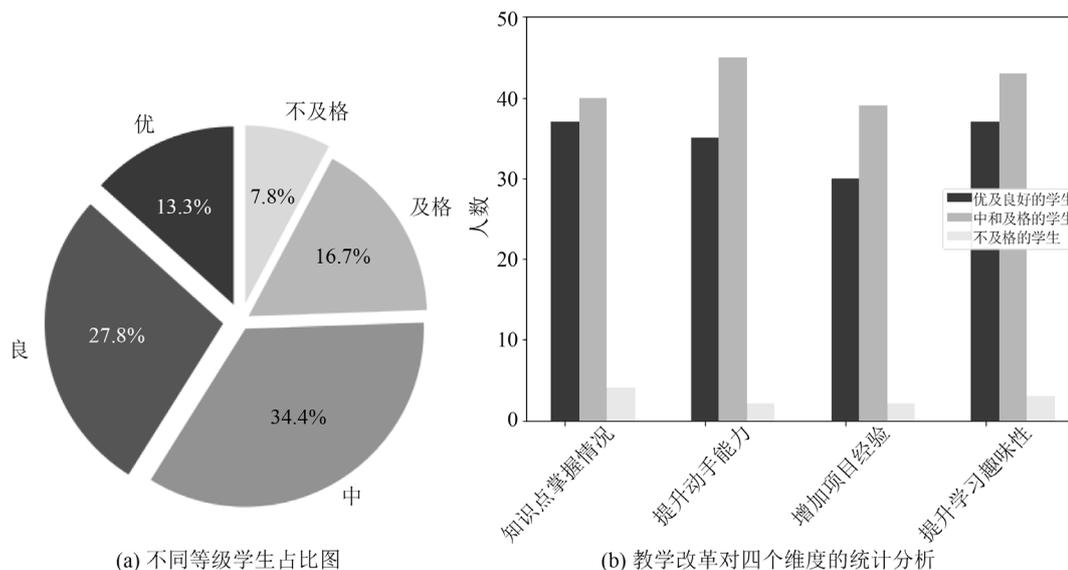


Figure 4. The statistics chart of the implementation of teaching reform

图 4. 教学改革实施效果统计图

5. 结束语

在数据科学飞速发展的今天, 各行各业对数据分析人才的要求越来越高, 数据科学的毕业生们不应该仅仅具有扎实理论基础, 还应该具有较强的实践动手能力、问题解决能力和创新精神, 他们应该能够满足市场对专业人才的需求。在这样的背景下, 本文探讨了数据科学专业下设的核心课程——《数据分析实践》在实验课程改革方面的问题, 包括实验课程内容设置、多元实验教学方式的变革、实验平台搭建和考核机制创新等。上述种种教学改革的核心在于, 以项目驱动的方式, 将数据分析的基本原理融入到各个实践环节, 帮助学生建立学习的内驱力, 形成终身学习、协作学习和探究式学习的综合素质, 培养他们对数据科学专业的热爱。同时, 本门课程在实验课程中所积累的各种资源, 如各种案例、数据、素材等, 也将为西南大学数据科学专业后续相关课程的建设奠定基础。

基金项目

西南大学实验技术项目(SYJ2020038); 西南大学校级教育教学改革项目(2020JY053)。

参考文献

- [1] 吕小光, 姜乐, 成青松. 数据科学与大数据技术专业人才培养模式探析[J]. 淮海工学院学报(社会科学版), 2018, 16(9): 132-135.
- [2] 马海群, 蒲攀. 大数据视阈下我国数据人才培养的思考[J]. 数字图书馆论坛, 2016(1): 2-9.
- [3] 曹高辉, 胡紫祎, 郭家乐, 等. 美国数据科学硕士专业培养要求与课程设置研究[J]. 数字图书馆论坛, 2018(5): 40-47.
- [4] 许嘉, 吕品. 哈佛大学数据科学课程教学初探[J]. 教育探索, 2015(15): 109-110.
- [5] 付希善, 曲国庆. 美国数据科学专业教学大纲调查[J]. 数字图书馆论坛, 2020(5): 43-48.
- [6] 周开发, 曾玉珍. 新工科的核心能力与教学模式探索[J]. 重庆高教研究, 2017, 5(3): 22-35.
- [7] 蔡剑红, 霍亮, 朱凌. 新工科理念下的实验课程教学改革与实践探索[J]. 测绘通报, 2019(2): 147-152.
- [8] 谢涛, 薄钧戈, 房琛琛. 新工科背景下 Web 编程技术课程的实验教学改革研究[J]. 计算机教育, 2019(1): 63-67.
- [9] 刘波, 毛先柏, 刘曼玲. 新工科电工电子创新实验教学改革探索[J]. 电气电子教学学报, 2020, 42(2): 138-142.
- [10] 农健. 新工科背景下计算机类专业实训教学质量评价体系研究[J]. 教育现代化, 2019, 6(81): 97-98.
- [11] 唐华芳, 曹蕾, 王梦豪. 虚拟仿真实训平台评价机制及指标体系的构建研究[J]. 中国教育信息化, 2020(2): 88-92.
- [12] 郭娟, 陈莹. 大数据专业实验教学与创新驱动型人才培养[J]. 现代信息科技, 2019, 3(21): 174-176.
- [13] 李春梅, 何洪, 李元, 程南璞. “新工科”背景下材料类专业“虚实互补”实验教学体系深化研究[J]. 西南师范大学学报(自然科学版), 2020, 45(4): 143-148.