

doi:10.16104/j.issn.1673-1891.2021.04.021

基于“两性一度”的数据挖掘 R 语言教学设计与实践

张丽平, 李松, 郝晓红, 李成严, 万静

(哈尔滨理工大学计算机科学与技术学院, 哈尔滨 黑龙江 151000)

摘要:针对数据挖掘 R 语言课程在教学内容和教学方法方面存在的不足,围绕“金课”建设总要求,基于“两性一度”的标准,结合线上线下混合式教学模式,在教学设计、教学实践和教学评价 3 个方面对数据挖掘 R 语言课程进行了优化。相比于传统教学模式,学生的学习积极性、主动性得到了较大的提升,较好地培养了学生发现问题、分析问题和解决问题的能力。

关键词:金课;两性一度;数据挖掘 R 语言;线上线下混合式教学模式

中图分类号:TP311.13-4;G642 **文献标志码:**A **文章编号:**1673-1891(2021)04-0114-05

Design and Practice of Data Mining with R Course Based on the Standard of Being Advanced, Innovative and Challenging

ZHANG Liping, LI Song, HAO Xiaohong, LI Chengyan, WAN Jing

(School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, Heilongjiang 151000, China)

Abstract: In view of the shortcomings of data mining R language course in teaching content and teaching methods, according to the general requirements of "golden course" construction, based on the standard of "being advanced, innovative and challenging", combined with the online and offline hybrid teaching mode, the data mining R language course is optimized in three aspects of teaching design, teaching practice and teaching evaluation. As compared with the traditional teaching mode, students' learning enthusiasm and willingness have been greatly improved, and students' ability to find, analyze and solve problems has been promoted with the new mode.

Keywords: golden class; standard of being advanced, innovative, and challenging; data mining with R; online and offline hybrid teaching mode

0 引言

2018 年 8 月份,教育部专门印发了《关于狠抓新时代全国高等学校本科教育工作会议精神落实的通知》,提出“各高校要全面梳理各门课程的教学内容,淘汰‘水课’、打造‘金课’,合理提升学业挑战度、增加课程难度、拓展课程深度,切实提高课程教学质量”。“金课”可以归结为“两性一度”:高阶性、创新性和挑战度^[1]。目前,各高校都在积极地进行“金课”建设,许多高校教师在“金课”建设方面进行了积极的探索和实践。

大数据是信息化发展的新阶段,新工科建设背景下,不少信息类专业开设了大数据相关课程。近

年,学者们对大数据课程体系和教学模式进行了一些重要研究^[2]。作为大数据课程的核心专业课,数据挖掘 R 语言是统计领域广泛使用的数据分析语言,主要用于统计分析、数据挖掘和知识发现与预测,可以对大数据进行数据价值分析和挖掘,同时采用图形的方式对数据进行可视化。传统的 R 语言教学内容主要集中在 R 语言基本操作、R 语言数据管理、R 语言编程规范、R 语言绘图和基于 R 语言的基础分析等方面。传统授课方法则主要是以教师课堂逐一讲解知识点,学生课下完成作业模式进行。传统教学内容理论化较强,较为枯燥;教学方法较为单一,学生学习的主动性和积极性不高。近年来,一些学者对 R 语言课程教学方法进行了研

收稿日期:2021-08-06

基金项目:黑龙江省高等教育教学改革研究项目(SJGY20200310);国家自然科学基金项目(61872105);黑龙江省自然科学基金研究项目(LH2020F047)。

作者简介:张丽平(1976—),女,辽宁铁岭人,副教授,硕士,研究方向:数据库理论及应用、大数据技术。

究。柯朝甫^[3]从R语言教学的必要性、教学目标、教学内容、教学手段和方法及考核形式进行了探讨。阳红英^[4]阐述了在R语言课程中进行案例教学的实施步骤,并分析了实施案例教学的成效。严文颖^[5]分析了R语言软件在生物信息专业教学中的优势,探索了适合生物信息专业特点的R语言课程的教学方法。但已有的方法距离金课建设的“两性一度”的标准要求尚相差较大,教学设计中没有充分体现高阶性、创新性和挑战度。教学方法和考核模式尚有很多不足,学生实践能力薄弱,发现问题、解决问题和分析问题的能力缺乏系统的培养。

为了弥补已有教学内容和教学方法的不足,在“金课”建设主旋律下,基于“两性一度”的标准,结合线上线下混合式教学模式,我们在教学设计、教学实践和教学评价3个方面对数据挖掘R语言课程进行了调整和优化,在教学实践中取得了良好的效果。

1 在教学设计中体现“两性一度”标准

课程教学是以学生发展为中心、由“教”与“学”双主体共同展开活动的过程,包含教师对课程的顶层设计、教与学活动过程的引导以及学生的自主学习等多个环节^[6]。

1.1 教学设计的高阶性

教学设计的高阶性,就是将知识、能力、素质有机融合,培养学生发现问题、分析问题和解决复杂问题的综合能力和高级思维。

在数据挖掘R语言课程的教学设计中,我们充分借鉴对分课堂^[7]的教学组织方式,采用“多项目并发,一项目到底”的形式,即基于课程教学目标和知识结构,选取涉及多行业多需求的项目进行分组教学,每个项目分发给2个小组,该项目将贯穿课程学习的整个过程,项目进展跟随课程教学进度进行,学生需要分阶段完成项目的各个功能模块,具体设计如图1所示。

教学实践表明,我们的教学设计具有如下优点:(1)强化基础知识的理解和应用;(2)培养学生发现问题和解决问题的创新能力;(3)一项目到底,可以培养学生运用所学知识进行问题分析、设计和实现的能力,与单知识点小项目相比,更有利于学生从整体上掌握复杂问题的分析设计方法;(4)小组间的方案对比可以提高学生分析、探讨、优化能力,使学生能够依据现有知识和认知能力对设计进行评价;(5)多项目的选取可以扩大学生的视野,让

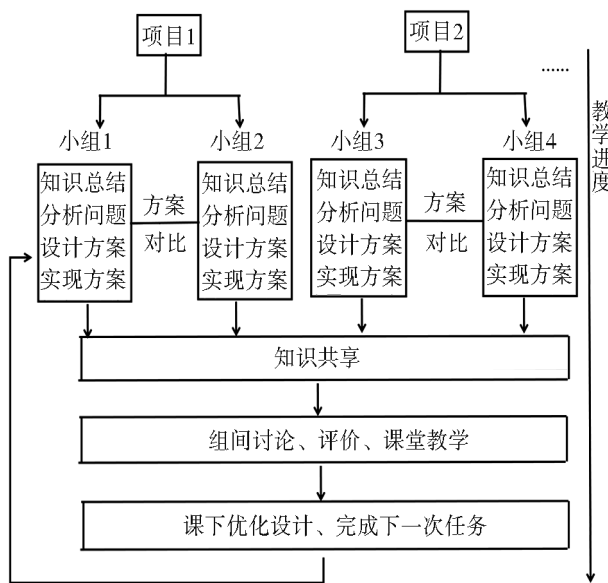


图1 数据挖掘R语言课程教学设计

学生在有限的学习时间内接触更多的项目实例,了解更多的行业需求、R语言应用场景及一些经典算法和常用的数据分析思路;(6)各组间的知识共享可以用来弥补学生对其他组项目知识的缺乏,组间讨论允许学生对其他组的项目进行质疑和探究,在讨论中提高对项目的理解和知识的渗透性,同时培养学生的团队协作能力。

1.2 教学设计的创新性

我们在数据挖掘R语言课程的教学设计中主要围绕以下3个方面保证了教学设计的创新性。

1.2.1 课程内容具有前沿性和时代性

数据挖掘R语言课程旨在培养具有如下能力的人才:能够从数据与实际需求中识别数据分析与挖掘目标,并选择分析方法;正确运用R语言完成对数据进行预处理、统计分析以及挖掘,并能对分析结果进行解释说明;针对不同领域的的数据,设计并实施挖掘蕴含于数据中的深层知识的能力。课程涵盖R语言知识、统计学知识、机器学习知识、数据挖掘知识、数据库知识、计算机算法和一些相关的行业知识。

课程在教材选择上力求经典和前沿。课程初期选择了葡萄牙作者 LuisTorgo 著的 *Data Mining with R: Learning with Case Studies* 中的部分案例作为实践项目,例如“预测股票市场收益”“侦测欺诈交易”等,这些项目虽然涵盖了无监督、半监督、有监督数据挖掘技术和基本的数据处理方法,但与学生的日常生活有一定距离,在实践过程中,学生存在对项目的理解不深入,需求分析不够准确,技术应用突破较难的问题。针对以上问题,我们将上面的

2 个实践项目替换为“餐饮企业综合分析”“天气情况分析”。表 1 以“天气情况分析”为例,说明了我们所设计的本课程实践的具体实施过程及覆盖的知识和技术。从表 1 可以看出,实践项目知识含量充足,可以逐步培养学生分析解决复杂问题、灵活运用知识和技术的能力。在不同项目的实现过程中,可以从数据量、分析目标和数据类型等方面引出挑战性任务。实践项目为学生提供了更多的知识应用场景,极大地提高了学生运用 R 语言和相关知识对特定行业进行分析和解决问题的能力。

表 1 “天气情况分析”实践项目

实践要求	步骤	要求	知识 & 技术
收集天气情况数据,对数据进行分析 and 处理,建立预测模型,对模型进行评估,使用模型对未来天气进行预测	数据获取	网络爬取/数据库获取	爬虫/数据库
	需求分析	根据行业特点和项目要求完成	软件工程
	数据预处理	(1) 缺失值与异常值分析 (2) 数据清洗 (3) 特征选取 (4) 数据转换	数据管理
	建模与应用	(1) 对数据进行统计分析并可视化显示结果 (2) 建立时间序列模型并进行预测,评估该模型 (3) 使用聚类算法对天气情况进行聚类,进行特征分析 (4) 对天气情况进行分类,建立分类模型(决策树/随机森林/SVM 并评价模型)	(1) 统计分析(描述性统计分析、推断性统计分析)、R 语言基础绘图(箱线图/折线图/饼图/核密度图等)、ggplot\lattice 高级绘图 (2) 生成时序对象、时序平滑化和分解、时序模型及预测(指数模型/ARIMA 模型) (3) 聚类算法:层次聚类/k-means 聚类等;聚类结果特征分析及可视化 (4) 建立分类模型(决策树/随机森林/SVM 等);评价模型

为了进一步提升学生对重要技术的掌握和动手实践能力,课程还设置了综合性实验,课程实验是对实践项目的有效补充,可以完善数据挖掘 R 语言的知识体系,对实践项目中涉及较少的知识点进行强化。

1.2.2 教学形式体现先进性和互动性

立足于“金课”建设目标,为了增强人才的培养效果,我们在数据挖掘 R 语言课程中采用线上和线下相结合的混合教学模式进行。在混合教学模式下,首先需要确定教学目标和建设教学资源,确定

教学目标即对本课程的培养目标进行分解细化,使其具有可操作性和可检验性。建设教学资源是根据教学目标收集、录制或者创作相应的教学资源。本课程的资源包括:课程视频、案例视频、知识题库、教学 PPT 等资料。知识题库、课程视频方便学生随时学习和查漏补缺;案例视频帮助学生在进行项目设计时进行借鉴和模仿。线下教学部分完成教学活动设计、效果评价设计和项目迭代设计 3 部分。教学活动设计包括:讨论时间分配、具有引导性的讨论主题设计、效果评价环节设计 3 个部分;效果评价设计包括:评价标准、“互评”点评 2 个部分;项目迭代设计是指确定项目分割为模块的方法和课程中进行迭代开发的策略。教师在课前完成确定教学目标、建设教学资源、教学活动设计、效果评价设计和项目迭代设计 5 部分工作,即可在线下教学中发挥课堂的教学主场作用。我们在课堂教学中借鉴了对分教学的思想:按照学生的情况,将学生按照 ABCC 或者 ABBC 进行分组,其中 A 为尖子生、B 为中等生、C 为后进生。课堂教学根据教学活动设计进行,将单一的教师主讲形式转变为教师+学生主讲、讨论+答辩、互评等多种形式,学生可以展示学习成果、考教重要知识点和寻求其他同学的帮助。教学形式多样、灵活,学生不仅是课堂教学的参与者,也可以是主讲人,较大程度提高了学生的学习积极性、主动性。

1.2.3 学习结果具有探究性和个性化

数据挖掘 R 语言的总体教学设计思想是“多项目并发,一项目到底”,项目的最终实现效果取决于学生对项目的需求分析、对方案的设计和对实现方法的选择,学生在不断的探索中设计完善方案和实现细节。在“天气情况分析”实践项目的实施过程中,开始时,我们将只包含“实践要求”和“步骤”2 行的表 1 发给两组同学,在项目实践过程中,学生填写表 1 中剩余内容,教师只提要求、做必要的指导,设定最低完成目标,不给出具体的实施方案,允许学生使用超出教师预设的“知识 & 技术”之外的相关知识和技术,鼓励学生从不同角度对项目做深入调研,作出更完整、更优秀的设计,因此,即使相同项目的两个小组,由于在各个阶段的设计差异,最终呈现的效果也会不同。这样的教学设计能够培养学生探究精神,发挥学生的个性特点。

1.3 教学设计具有挑战度

数据挖掘 R 语言课程选择的项目具有行业时代特点,并设有最低完成要求。学生在完成项目的过程中,需要掌握相关的行业知识和 R 语言实现方

法,涉及一些跨专业、跨学科的知识,此部分虽然高于教学内容,却是学生经过努力可以达到的高度,需要学生充分发挥学习主动性进行资料收集和自主学习。课堂内容的难度和挑战度会使学生认识到知识鸿沟,充分了解自己的差距,从而激发学生内在的学习动力和学习热情。

混合教学模式下,数据挖掘R语言课程的教学设计不再是简单地照本宣科,需要教师投入更多的时间去进行课前准备,包括了解前沿技术、行业需求、资源选择和教学设计等工作。教学过程中,对于学生提出的新设计和新技术,教师需要和学生共同学习,一起探讨。

2 数据挖掘R语言教学实践

2.1 线上教学资源建设

课程组开课即完成和优化了课程教学视频录制。在超星平台提供每一章课件、单元测试、教学视频、参考书籍等相关内容;提供教学案例。为保证线上线下教学的开展奠定坚实基础。

2.2 线上+线下混合教学活动

为了充分提升教学效果,在数据挖掘R语言课程教学中,我们采用线上和线下相结合的方式,科学、有序开展教学活动。上课前,教师在超星平台发布教学资源和疑点登记表,方便学生完成课前预习。项目小组发布项目相关知识,帮助其他组的学生在线下课堂上对项目进行理解。在课前学生需要完成这两部分内容的学习;上课中,借鉴对分课堂教学思想,将教学分为教师主讲、学生主讲、讨论、提问、点评等环节,根据课堂内容决定各环节时间分配;对于设计不达标的小组给出修改建议;教师根据本次课堂教学内容布置各小组下一功能模块要求;课后,学生运用所学知识完成该组项目的相关功能模块分析、设计和实现,对于跨学科、超范围知识点进行学习,总结相关知识点,发布到超星平台,供其他学生学习;教师对学生在设计 and 实现过程中出现的问题进行答疑和指导。具体教学实施过程如图2所示。

在教学过程中,教师要随时掌握学生的学习情况,包括:(1)课前预习情况:通过预习效果检验(课课练)和疑点登记表,掌握学生的普遍问题,对存在问题较多的个别同学进行单独辅导;(2)项目进展情况:通过学生分享的小组知识点总结和课堂讨论及项目讲解了解项目的进展情况;(3)项目改进情况:对于课堂上设计不满足最低要求的小组,跟进小组讨论、进行技术指导;(4)学生知识推进程度:

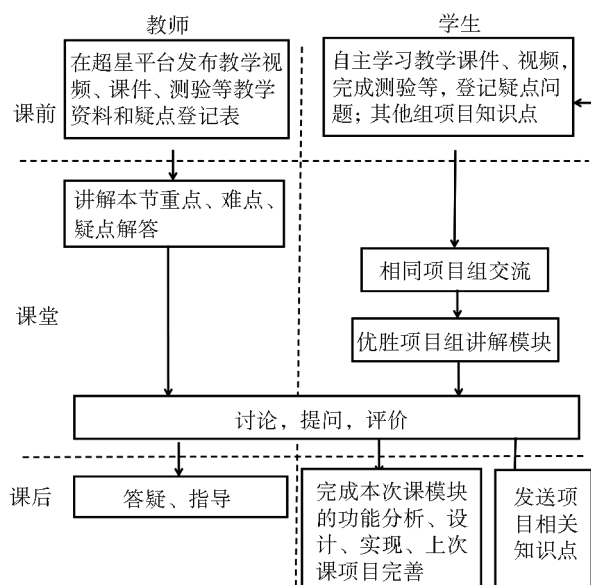


图2 数据挖掘R语言课程线上线下载混合教学过程

我们采用阶段测试、代码抽查、随机问答等形式跟踪检验。课程组会定时汇总教学情况,对于共性问题,适当调整线上和线下教学的时间分配或教学内容安排,对于学生个体存在的具体问题,由任课教师协助学生排除学习障碍,提高学习效果。

3 数据挖掘R语言教学评价

为了强化教学过程管理,我们采用O2O课程评价体系,对学生成绩的几个组成部分:课堂表现、综合实验、实践(项目)、阶段测试、期末成绩进行严格把控。课堂表现依据学生对R语言数据分析与挖掘技术的理解、掌握和应用程度评定成绩。综合实验依据学生完成实验题目的设计、实现、运行情况、实验报告的情况评定成绩。实践(项目)依据学生进行综合应用程序开发情况、项目实践报告情况评定成绩。阶段测试采用题库的形式对学生的基础知识进行检验。期末是开放式题目,采用一人一组,答辩报告形式,成绩构成包括:系统演示、答辩、报告3部分,成绩依据问题分析、技术掌握、文档内容及格式等方面评定成绩。要求能够准确掌握领域数据包含的信息,对领域问题准确分析,并得到分析目标;针对分析目标,准确选择R语言提供的数据统计分析方法,对数据进行分析,并得到准确性高的分析结果;能够利用R语言绘图对数据以及分析结果进行可视化;能够对分析结果进行较充分的解释。文档能够涵盖对问题的描述、对问题的分析、选择何种方法、如何应用方法解决问题,以及对结果的解释整个过程的完整描述。各项按一定权重计入总评成绩。

数据挖掘 R 语言课程从 4 人一组协作完成项目到期末的一人一组独立完成项目,这个过程既给了学生能力的成长时间,又对每个学生提出了独立的能力要求。近 3 a 的教学实践表明学生的总体能力有了显著提高,具体表现在:实践项目的技术和难度有所提高;项目中自学技术和专业知识增多;期末题目涉及的领域更加广泛、部分学生选择的题目具有明显的时效性和独特视角,例如:新冠疫情数据分析及预测等;项目分析更加准确、技术更加全面,分析和解决复杂问题的能力有较大程度的提升。

参考文献:

- [1] 吴岩.建设中国“金课”[J].中国大学教学,2018(12):4-9.
- [2] 杨维明,王时绘,余敦辉,等.新工科背景下 IT 类专业大数据课程体系建设与教学模式研究[J].计算机教育,2019(11):64-67
- [3] 柯朝甫.医学本科生 R 语言教学初探[J].教育现代化,2018,5(48):233-245.
- [4] 阳红英.案例教学在《数据挖掘与 R 语言》课程教学中实践与思考[J].中国多媒体与网络教学学报(上旬刊),2020(8):106-107
- [5] 严文颖.生物信息专业 R 语言课程教学的探索和研究[J].现代信息科技,2018,2(11):108-112
- [6] 宋专茂,刘荣华.课程教学“两性一度”的操作性分析[J].教育理论与实践,2021(12):48-51.
- [7] 张学新.对分课堂:大学课堂教学改革的新探索[J].复旦大学教育论坛,2014,12(5):5-10.
- [8] 高霏霏,孔令军,舒晓春.基于 MOOC 的“四段式”线上教学模式的应用研究[J].西昌学院学报(自然科学版),2021,35(2):120-123

4 结语

“金课”建设为数据挖掘 R 语言的课程建设提供了契机,“两性一度”的指导标准为课程改革提供了明确的方向,我们在教学设计、教学实践和教学评价 3 个方面对数据挖掘 R 语言课程进行了调整和优化,在教学实践中取得了良好的效果。在教学的过程中,我们将继续践行该标准,在实践中逐步完善课程体系和教学设计,引入“四段式”线上教学模式^[8],进一步整合和优化各种教学资源,致力于打造更加优质的课程。