

基于数据挖掘技术的混合式教学实践研究

——以《计算机程序设计》课程为例

李俊杰

(陆军军医大学基础医学院计算机教研室, 重庆 400030)

摘要: 随着教学平台中各类教育类数据的海量产生, 推动了大数据技术与教育教学的深度融合, 如何深入分析和挖掘对教学效果有益的行为和反馈数据, 以期改善教学质量、变革教学模式提供数据决策支持, 成为了在线教育者研究热点。本文通过利用 PYTHON123 网络教学平台开展混合式教学, 将某高校学生学习行为全过程数据进行科学处理, 并将有效数据进行提取分析, 对混合式教学中的教师授课、学生自主学习、学习反馈与师生交互各项数据实现量化, 同时构建了聚类分析模型进行相关性分析, 以期能够对精准支持课程教学、优化学习任务、提高学习质量提供支撑作用。

关键词: 大数据技术; 混合式教学; 聚类算法

移动互联网和大数据时代的到来, 产生的“互联网+”模式对教育领域影响深远, 对传统的教育模式和学习分析提出了挑战, 对高等教育的教学理念和学习方式产生了深刻变革。这些教学平台产生了学生学习行为的各种动态数据。本文选取某校 2019–2021 级本科学员开展《计算机程序设计》为研究对象, 利用“PYTHON123”教学平台, 对在线学习平台的学生学习行为分析着手, 跟踪课程学习过程, 并利用机器学习中的算法对学生的行为进行分析, 从中挖掘学习行为与学习结果之间的映射关系, 从而使得能够更客观更科学地了解影响学生成绩的因素。基于此, 教师还可以根据每个学生的学习行为来总结分析学生学习特点, 采取针对性教学指导, 根据不同的情况给予不同的建议, 有效改善学生学习结果。

一、理论基础

(一) 混合式教学

混合教学模式核心内容指通过使多种学习资源、学习地理时空、课程组织形式和教学方法工具进行深度融合, 在教学过程注重培养学生的自主性, 以学生为主体, 加强互动交流、运用各种教学资源全面支持学生的知识建构和技能提高的一种灵活自由的教学模式。混合教学模式的方式主要有: MOOC、微课、翻转课堂等。混合式教学模式的实施需要高效的在线工具, 具有教学研讨、教学测试、角色互换、游戏教学、个别辅导、虚拟实验室等多种功能在发挥传统课堂学习特点基础上, 结合在线的网络学习优势, 开展创新性教学, 不断激发学生兴趣, 提高课程教学质量。

(二) 教学数据挖掘技术

教学数据挖掘技术以学习行为大数据为基础, 研究学生在线学习行为特征、学习投入、学习效果之间影响关联, 为教师和教学管理人员提供重要参考。在技术层面, 通常利用数据与模型, 进行测量、收集、分析和报告关于学习者及其学习情景的数据, 以期了解和优化学习和学习发生的情境。国外有 KhanAcademy、Moodle 等多个在线教育平台采用“学习仪表盘”等学习支持工具, 对学生学习过程跟踪, 进行游戏化学习评价、可视化数据展示、学习结果分析等功能, 为在线教育的学习者、教师、研究者、教育管理者等提供多层次的学习支持。

(三) K-Means 聚类算法

聚类分析又称群分析, 它是研究(样品或指标)分类问题的一种统计分析方法, 同时也是数据挖掘的一个重要算法。K-Means 是一种比较经典的基于距离的分层聚类算法, 它是数据点到原型的某种距离作为优化的目标函数, 在最小化误差函数的基础上不断迭代运算, 来调整规则, 在教学数据挖掘分析中可以应用于数

据的聚类分析。这样就能够进一步了解学生在线学习行为路径与学习效果, 确保教师教学设计做到“因材施教”。

二、数据挖掘应用于《计算机程序设计》课程的实证研究

(一) 混合式教学平台实施过程

实施阶段: 一是课程建设阶段, 包括确定教学内容、设置教学目标、组建教师授课队伍、选用优秀教程、设计混合式教学平台框架和相应教学资源; 二是在课程发展阶段, 更新了教学资源, 开展了课程体系改革(包括线上自习、课堂教学、随堂实践、线上实训等要素进行混合)、重点利用 Python123 OJ (Online Judge) 在线测评平台开展 SPOC 教学。三是在课程完善阶段: 教学内容向第二课堂拓展、选拔优秀学员参加创新竞赛、完善课程建设质量评估机制、进行平台教学内容更新。

课程教学内容模块: 教学内容均采用模块化设计, 有习题(OJ 自主勘误), 有作业(有讲解视频)。课堂讨论内容提前发布, 结合程序设计实践性强的特点, 结合 OBE 教育教学理念进行教学设计。

(二) 混合式教学数据资源

课程通过 Python123 开展 SPOC 教学。平台开放大量精品 Python 教学资源以及 OJ (Online Judge) 在线测评平台, 便于学员进行自主学习。MOOC 公开课(Python 语言程序设计)在学生登录后下方有链接入口, 教师团队自建课程导航栏内有导学案例和学习掌握目标。其中课件资源数 62 项(视频 32 个、PPT 22 个、思维导图 8 个), 外链 65 个; 问答 48 个; 课程内容共 76 (必做)+211 (拓展) 道习题, 作业 9 次, 一次期中摸查、一次模拟测试、一次考试、16 个课件, 总发帖量 85 个, 发帖人数 60 人, 学生代码 35213 行。

我的题库		
所有习题	未分类习题	我的校内习题
python内题库	python理论题	python程序填空
python程序改错题	python程序判断题	python程序简答题
程序改错题		
校内习题库		
Python 基础题库 (123)		
English (四学生)	Turtle图形绘制(人工评测)	公开课程实例(非自动)
函数和代码块用	基本数据类型	基本语法元素
文件与数据格式化	程序控制结构	程序设计基本方法
程序设计方法类	组合数据类型	

图 1 《计算机程序设计》教学资源题库

通过数据预处理, 提取相关性最大的常用数据, 包括登录课

程次数、发帖次数、答疑次数、学习笔记、课程 PPT 浏览次数、提问答疑次数、时长、作业、成绩等。

(三) 聚类算法的实现

1. 基于 K-means 聚类算法

流程为：输入集：待聚类分析的数据集合 Z，聚类的数目 s；
输出结果：s 个初始的聚类中心。

步骤：①按照维度对 Z 数据集合重新分布，每一个维度的数据按照从小到大的方式排列，形成新的集合 Z1；②对数据集 Z1 进行分割得到 s 个区间；③对每个区间的中心值在区间内位置或序号进行计算；④取出中间序号所对应的数据，确定其为初始化聚类中心；⑤利用每一个分组中的所有数据计算初始中心：

$$q = \left(\frac{n \times j}{s} + \frac{n \times (j-1)}{s} \right) / 2$$

其中，q 表示中心值序号；

n 为集合总个数；s 为类簇的总个数；j 为其中一类的序号。

2. 数据的处理

选取“视频学习时长”“章节作业完成率”“模拟测试得分”“讨论答疑次数”“登录总时长”进行处理，建立数据集，如表 1。

表 1 部分学习行为数据样本集

序号	学号	视频学习时长 (min)	章节作业完成率 (%)	登录总时长 (min)	讨论答疑次数	模拟测试得分
1	20190010020	640	75	2747	60	优
2	20190010021	580	67	2415	45	良
3	20190010022	620	58	3215	76	优
4	20190010023	570	46	1596	65	中
5	20190010024	620	80	1398	32	中

通过数据跟踪发现，不同的行为路径对结果有不同相关性，每一步在学习行为路径下的持续与否具有随机性，可通过聚类算法完成相应的分析处理过程，具体表示：学习视频课件(1)，在线做程序题(2)，查阅资料(3)，作业完成情况(4)，讨论答疑(5)，未采取行动(0)。通过 1-5 数字进行路径分析。同时进一步进行数据清洗，精简数据分析过程。

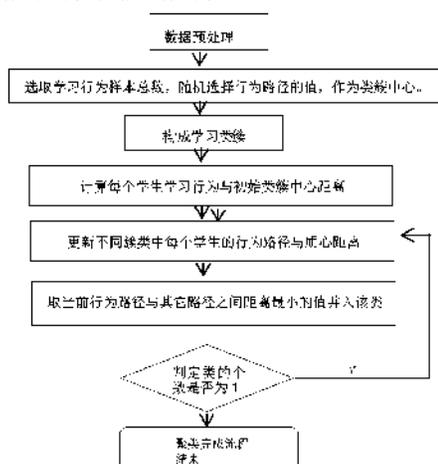


图 2 学习行为聚类流程

接着，通过 spssstatistics 工具完成聚类训练过程，采用迭代与分类方法，设置聚类数为 3，聚类最大迭代次数为 15 次，经多次类比和分析，得到初始聚类中心和迭代历史记录，最终得到最优的学习步骤，说明在线学习的首要目的在于浏览讲解视频和查找资料；而对于知识巩固和扩展，完成作业和论坛答疑发帖是最主要也是最有效的学习路径。并继续验证发现，大多数等级成绩较好的同学，采取了浏览视频课件—完成作业—论坛答疑讨论—交

流互动—拓展阅读的学习路径。

3. 聚类结果相关分析

结合统计表与可视化可以看出，对于观看时长、访问数、作业完成率和期末成绩四个特征，呈现正向关系，最终成绩越高；而对于作业得分和测试两个特征，结合整体来看，对最终成绩影响不大。



图 3 聚类结果可视化

三、分析及结论

(一) 影响因素比较

从影响比较大的因素依次排列为作业完成率、观看时长、平台访问数、期末考试前测试等，反过来也提醒学生，要取得较好成绩，应多花些时间去完成老师布置的任务；及时观看老师在平台上发布的视频，深入学习理解知识点；同时应多次进入平台去巩固自己的知识，最后对期末前综合测试应给予一定的重视，通过练习巩固和复习知识。同时也提醒教师，在课程成绩评定时，要建立全过程跟踪评价体系，包括在线表现(互助答疑)和平时作业完成情况(章节练习、在线测试等)。平时可利用聚类分析，观测学生的难易点和掌握程度，对预警同学进行了针对性一对一或一对多辅导。

(二) 需完善的策略机制

一是继续研发平台优势功能，对学习者的个性化跟踪与反馈机制进行研究，优化完善对学生的考核评价体系，建立完善的评价指标体系；二是进一步研究促进学生参与计算机程序学习的策略机制，设计出一系列提升教学效果的教学活动(课前线上活动、课堂线下活动、课后在线活动等设计)。

综上所述，本文利用数据挖掘的方法对混合式教学进行了学习分析研究，将程序设计教学渗透到日常网络和智能终端的使用中，实现学习空间、学习资源、学习方式的深度融合。创新之处在于，利用聚类算法，对学习过程中产生的数据进行挖掘分析，进行科学评判，预测学习者学习成绩并提供个性化指导干预对策，有利于对课程进行精准教学支持，优化学习任务，创新教学方法，提升学生学习质量。

参考文献：

[1] 吴永和, 陈丹, 马晓玲等, 学习分析: 教育信息化的新浪潮[J]. 远程教育杂志, 2013, 31(4): 11-19.
 [2] ARORA S, GOEL M, SABITHA A S, et al. Learner groups in massive open online courses[J]. American Journal of Distance Education, 2017, 31(2), 80-97
 [3] 周庆, 牟超, 杨丹. 教育数据挖掘研究进展综述[J]. 软件学报, 2015, 26(11): 3026-3042.

基金支持项目：陆军军医大学 2019 年教育研究课题(2019B02)：数据挖掘技术在《计算机程序设计》网络教学平台中的应用研究；陆军军医大学基础医学院 2019 年教育研究课题(2019JCYXB12)：基于大数据技术的混合式教学实践研究

作者简介：李俊杰(1980-)，男，硕士，副教授，计算机科学与技术。