

基于大数据分析的在线课程教学质量评估

梅丹丹¹ 樊卫北^{2, 通信作者} 张玉杰² 王汝传²

(1. 南京邮电大学通达学院, 江苏 扬州 225127;

2. 南京邮电大学计算机学院, 江苏 南京 210023)

摘要: 随着云计算和大数据等技术的发展, 逐渐影响和改变到现有的课程教与学的模式。在线课程打破了传统教学在时间与地域上的局限性, 以慕课为代表的新型课程教学平台得以广泛使用。面对慕课平台中产生的大量课程学习信息, 为了解决传统的课程评价机制无法进行有效分析的问题, 借助 Spark 并行计算模型对在线课程信息进行研究分析, 设计实时数据与离线数据相结合批处理数据分析评估, 以便从多元化角度探讨学习成效与课程教学质量评估, 提升在线课程中教师授课与学生之间的参与度。

关键词: 大数据分析; 在线课程; 慕课; 教学评估

随着信息技术与多媒体的快速发展, 传统的教学模式逐步向线上教学过渡, 如中国大学慕课(MOOC)、腾讯课堂、雨课堂等。这些新兴教学形式的出现, 极大地丰富了传统教学模式的教学手段, 虽然克服了传统教学的局限性, 但是带来了教学质量评估的复杂性及多元性。网络教学平台会产生大量教学信息数据, 包括实时的与离线的, 因此, 有必要借助大数据分析工具来对线上教学数据进行处理分析, 从而为线上教学质量评估提供一种新颖的评估方法。

国务院《促进大数据发展行动纲要》提出“探索发挥大数据对变革教育方式、促进教育公平、提升教育质量的支撑作用”。大数据在教育中的应用主要涉及在线决策、学习分析和数据挖掘三个要素。其主要功能是应用和研究预测分析、行为分析和学术分析。大数据的含义是指学生在学习过程中产生的大量数据(数据来源包括两个方面: 显性行为 and 隐性行为)。其中, 隐性行为包括论坛发帖、课外活动和在线社交活动, 不直接作为教育活动进行评价。包括考试成绩、作业完成情况和课堂表现。大数据模型和显示的数据可以为学校和教师的教学提供参考, 及时准确地评估学生的学业状况, 发现学生潜在的问题。

教学评价对于教学质量的改进起到关键作用, 传统的教学评价还是以学生评课为主。对于信息化程度高的今日, 传统的教学评价有以下不足: 学生评课通常在学期期末, 对于教师授课的改进存在滞后性; 学生只能按照系统设定内容进行打分, 评课形式较单一; 对于 MOOC、雨课堂等线上教学形式, 传统手段不能有效地分析出学生线上学习的参与度及达成度。

基于此, 本文基于 Hadoop 环境, 构建分布式文件系统 HDFS+Spark MLlib 在线课程评估系统: 一方面, 结合大数据分析技术, 设计面向授课质量监测评估的大数据平台, 提升教学评价与改进的实效性; 另一方面, 通过对在线课程的数据深入分析与挖掘, 可以提升学生参与在线课程的参与度与达成度, 可以从多元角度直观地展示出学习成效, 便于教师随时掌握和了解学生的学习过程及掌握程度。

一、国内外研究现状

目前, 在国内外, 大数据分析技术已经应用到多种领域, 如智能交通、医疗数据、政务系统等, 并且大数据分析在教育领域也发挥了关键的作用。面向大数据的学情诊断、个性化学习分析智能决策支持, 在提升教育品质的同时, 对促进教育公平、提升教育质量、优化教育管理都起到十分重要的支撑。

比如美国采用学生学习调查, 主要包括新生基本信息调查、学习过程调查、综合性校友评价调查。目的是评估学生在经历大学学习生活后, 在能力、认知、情感等方面的发展变化, 以及了解校友的个人及职业发展情况, 作为大学教育绩效评定的重要参考及改进学生培养工作的依据。澳洲对学生学习质量的评估包括学习经历、生源质量与成果, 学生学习调查为院校改进学生教育流程, 提高教育质量提供了一个科学的分析手段, 从而起到保障作用。在英国, 大部分学者认为高校的绩效应该体现在能在多大程度上促进学生的学习与发展上, 将学生学习调查作为一种间接测量的定量研究方式。他们比其他直接的测量方式可以更高效、全面地了解学生的状况, 尤其是学生的态度、行为这些难以用传统测验直接测量学生状况。事实上, 无论是采用哪种形式的课程教学质量评估, 其目的都是为了促进学生素质教育的健康发展。

国内学者对课程教学质量监测与评估体系也进行了比较深入的研究。文献提出了证据理论与 BP 神经网络模型相结合的教学质量评估方法, 实现了教学质量的综合评估, 结果表明可有效评估互联网在线教学质量。蒋家琼等分别介绍了现有课程教学质量监测评估主体(高等教育质量保障署、社会组织及相关专业团体等)的组织架构、保障中的作用和标准。从政府职能转变、外部组织间联系、质量监测评估与社会各界的沟通交流等, 提出构建课程教学质量监测评估体系的可操作性建议。我国学者对国外发达国家的高等教育质量监控与评估体系进行了大量的研究, 以期建立适合我国的高等教育质量监控与评估体系。程海涛等通过构建知识图谱, 研究了在线慕课课程的知识成效分析, 更加直观地展示了学生的学习过程。廖卓凡等分析计算机专业在线课程教学的特

点,提出了双循环互促的计算机在线教学评估方法,可以充分提升在线教学的优势。

总的来说,我国学者在课程教学质量监测评估的研究中取得了一些成绩,为课程教学质量监测评估研究的进一步深入研究奠定了基础,具有一定的影响力。但上述大多数研究还停留于探索的初期,特别是,基于大数据的在线课程的多元化评估还较少,鉴于此,本文对基于云计算 Hadoop 环境的在线课程评估进行了研究。

二、基于大数据分析的教学质量评估

(一) 网络课程教学评估模型

将课程教学监测评估特征与大数据处理平台的有机结合,形成了一套较完整的大数据采集、数据处理、结果使用的多功能系统,系统整体架构图 1 如下:

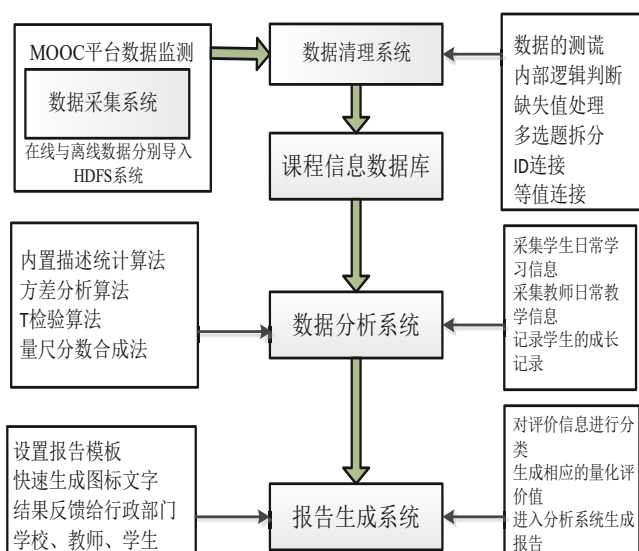


图 1 教学质量评估系统架构图

借助图像识别算法,自动化采集纸质数据。采用在线问卷调查的形式,可以有序地实现教师信息采集。其次采集教学资源情况、教师日常教学信息、学生学业成绩及校内外活动及学生与课程评价信息。最后基于教师的日常教学信息,生成相应的量化评价报告。

将采集的数据先存储到本地数据库,再统一导入到分布式文件系统中,实现数据统一管理共享。本文采用自定义分区类型,按照学校、教师、学生三种角色之间的关联性存储到相应的节点。将录入监测数据形成实时的数据流,把已关联的数据流切成数据块。作为 Spark 集群的 RDD(弹性分布式集)操作,在分析的过程中主要采用内置描述算法、方差分析算法、T 检验算法等,挖掘出有用知识。

课程教学质量评估体系依靠大数据网络系统,形成了一套数据采集、数据存储、数据处理、结果反馈的完整平台,只有将二者有机结合,才能构建教评一体化,协同推进的完整体系。

综合上述课程教学质量监测评估模型、课程教学质量监测评估理论,以及面向监测数据的大数据平台设计等研究内容的具体设计和实施方案,基于南京邮电大学本科教学审核评估的应用展示,并在此基础上,对上述所提技术方案进行测试和验证。

尽管数据的来源主要集中在学校范围,但学校各部门及学生信息并非一成不变,例如教师的岗位信息以及职称信息,还有学生的选课信息等,具备了动态性、异构性特点。以南京邮电大学现有教师数据为例,部分数据量较小的项,可由学院用户手工填写,如学生对南邮的评价信息,本文采用纸质问卷的形式,通过设定关于课程评估问题,发放给学生,填写完成后交给学校,或通过毕业生工作部门反馈的形式,让毕业生填写学校评价信息,以此获取学校知名度与社会认可度的相关数据。

监测数据的处理过程主要由 Spark 平台实现,实现一个多表处理的过程。对于采集的数据首先进行分类,划定以学校、教师和学生为三大角色,其次进行聚类分析,例如采集到与学校相关的数据就划分到学校属性中,聚类分析完成后就把相关联的数据划分成数据流,交给 Spark Streaming 处理,进行多节点分布式处理,最后汇聚到数据分析系统。

以现有的南京邮电大学本科生现有数据为例,2019-2020 学年学生的考试成绩、线上出勤率以及单元测试等真实数据,示范监测数据的处理过程。本文采用聚类算法,对学生考试成绩、线上出勤率以及单元测试信息进行聚类分析,创建三者之间的关联性,如图 2 所示:

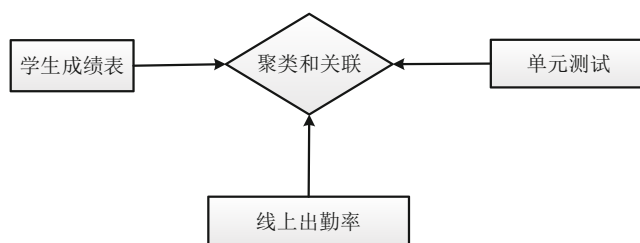


图 2 多表之间的关联

将创建好关联性的三张表格数据交由 Spark Streaming 处理,如图 2 所示,原始数据被划分成数据流,交给多节点处理,实现一个实现处理、实时分析的过程,各个节点处理完成后将结果汇聚到一张完整的表格信息中,输入到报告生成系统中。接下来给出本学年学院高级职称教师的上课评价标准,高级职称上课率 = (上课教授数 / 教授总数),副高级职称教师上课率 = (上课副教授数 / 副教授总数)。根据上面的得分标准判断:大于 15 分可以判定为课堂教学优秀;12-15 分之间,判定为课堂教学尚可;9-12 分之间,判定为课堂教学一般;0-9 分之间,判定为课堂教学较差。由此得到教学评估报告。

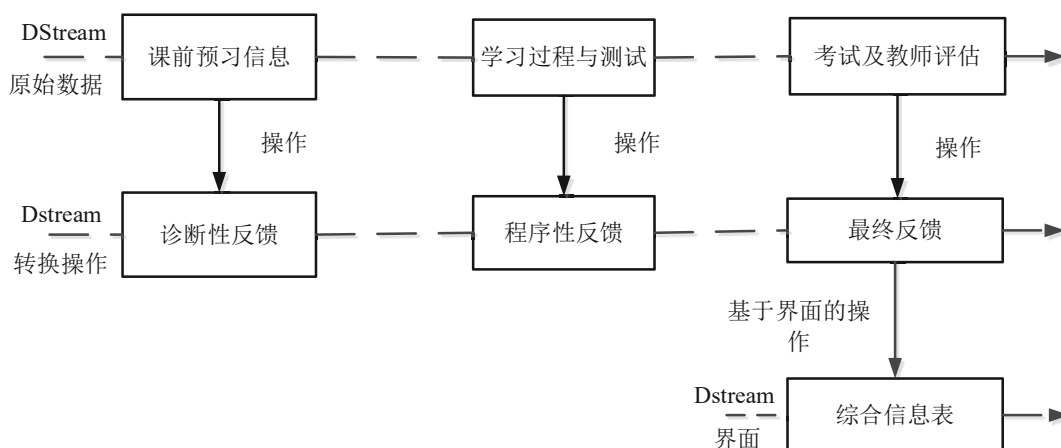


图3 相关联表格处理过程

综合信息表格包括“三种反馈”和“一种可视化工具”，“三种反馈”即诊断性反馈、程序性反馈和最终反馈。如图3所示，主要对学习和预测试提供及时反馈；过程反馈在学习过程中，主要提供对过程测试、章节测试、问答卡、交流学习活动的及时反馈，以及对学习过程的及时预警。在课程结束时进行，终止反馈主要用于同行评估、教师评估和期末考试，提供及时反馈。在反馈给学习者的同时，这三种反馈信息也会传递给教师。根据这些信息，教师在必要时对学习者的学习活动进行及时干预。

（二）网络课程质量评估系统

笔者基于五台高性能服务器，其中一台作为主节点，四台作为从节点，构建小型计算集群。基于Hadoop环境，构建了HDFS+Spark Mlib的在线网络课程质量评估系统，其中HDFS用于多节点分布式存储，通过Sqoop实现HDFS与本地SQL数据库的导入。将采集到的图片、文字等学习数据进行统一存储，利用Spark实现流式数据的实时分析。基于Mlib中的机器学习算法库构建统计分析算法，利用Python编写前端，实现人机交互可视化。

以《网络技术与应用》在线课程为例，整个处理过程是一个线下存储、线上分析、结果输送的整体过程。对学生学习过程中产生的大量数据进行解释，以评估学业进步、预测未来表现和识别潜在问题。数据来源于学生的显性行为，如平时作业和考试，以及隐藏的学生行为，如在线社交网络、课外活动和论坛发帖，以及其他未直接评估为学生教育进步的活动。

该系统通过采集学生在线学习数据，将学生成绩表、单元测试与出勤率进行分类汇总，结合Spark Mlib中的数据分析算法计算得到综合评价报告，最后在平台界面以图表的形式展示出学生的学习参与度、单元测试、薄弱环节、课后反馈等评估数据。

三、结语

通过对线上教学监测采集获得的数据，结合大数据分析算法，可以更合理判断课程教学评价的影响因素、变化趋势和规律。本文提出了基于大数据的线上教学信息综合分析模型，可以准确把

握课程教学质量监测评估的现状和特征。以《网络技术与应用》授课质量评估为例，通过大数据平台的质量监测系统，对线上学习平台提供更详细的课程质量评估。可以帮助教师全面了解课程在线学习情况，及时发现问题，深入分析原因，为其他在线课程的质量评估提供科学的理论与实践依据。

参考文献：

- [1] 廖非凡, 黄园媛, 王静, 等. 双循环互促的计算机专业课程在线教学与评估研究[J]. 计算机教育, 2021(06): 11-14.
- [2] 李国强. 高校内部质量保障体系建设的成效、问题与展望[J]. 中国高教研究, 2016(2): 1-11.
- [3] 石建国, 刘涛, 修宇, 等. 云计算课程在线教学策略及实践探析[J]. 计算机教育, 2020(09): 34-38.
- [4] 郑石明. 大数据驱动创新创业教育变革: 理论与实践[J]. 清华大学教育研究, 2016, 37(03): 65-73.
- [5] 王博丞, 刘胜娟, 闫燕勤, 等. 基于新型因果关系的高校实验教学质量评估系统的设计与实现[J]. 教育现代化, 2020, 7(37): 8-10.
- [6] Kinser K, Lane J E. Managing the oversight of international branch campuses in higher education[J]. Higher Education Management and Policy, 2014(3): 161-176.
- [7] 吴艳丽, 武创举, 阴钰娇. 证据理论和神经网络的互联网教学质量评估模型[J]. 现代电子技术, 2020, 43(19): 175-178.
- [8] 蒋家琼. 我国科学素质教育政策的目标、主体及类型[J]. 湖南师范大学教育科学学报, 2016, 15(03): 47-50.
- [9] 程海涛, 李世勋, 徐鹤, 等. 基于知识图谱的慕课课程学习成效分析[J]. 计算机教育, 2021(04): 107-110+115.
- [10] 廖非凡, 黄园媛, 王静, 等. 双循环互促的计算机专业课程在线教学与评估研究[J]. 计算机教育, 2021(06): 11-14.

基金项目：南京邮电大学通达学院教学改革项目（JG30620010）。