

基于大数据学习分析的在线学习绩效 预警因素及干预对策的实证研究

赵慧琼¹, 姜强¹, 赵蔚¹, 李勇帆², 赵艳³

(1. 东北师范大学 计算机科学与技术学院, 吉林 长春 130117;
2. 湖南第一师范学院 信息科学与工程学院, 湖南 长沙 410205;
3. 长春教育学院, 吉林 长春 130061)

[摘要] 利用学习分析技术挖掘教育大数据,能够发现潜在价值,并使其转换成有意义的教学信息,进而优化学习过程、提高教学效果,成为教师、学生及教育研究者的共同诉求。从学习分析技术的教育应用视角,追踪、积累并筛选在线学习行为数据,利用多元回归分析法判定影响学生学习绩效的预警因素,在此基础上构建了干预模型,将其应用于教学实践中,对产生的学习行为数据进行二元 Logistic 回归分析,并结合问卷调查和访谈法对该模型在学习活动、知识习得等方面的有效性进行验证。结果表明,通过学习过程中实施的干预模型识别出存在学习危机的学生,及时向其发出预警信号并提供个性化干预对策,有利于增强学习动机,培养学习毅力,提高学习质量。

[关键词] 教育大数据; 学习分析; 预警; 干预对策; 个性化学习

[中图分类号] G434 [文献标志码] A

[作者简介] 赵慧琼(1993—),女,河南许昌人。硕士研究生,主要从事大数据学习分析研究。E-mail:zhaohq071@nenu.edu.cn。

一、引言

利用数据挖掘算法、机器学习、关联分析、回归分析等学习分析技术,收集、测量、分析与报告大量的学生数据,理解与优化教学过程及其情境,为教学决策、学业预警提供支持,真正实现个性化学习,提高教学效果,这是大数据学习分析在教育领域的价值所在。随着 MOOCs 在全球范围内的普及,教育大数据来源、种类的不断丰富,学习分析技术的应用及发展越来越受到广大教育者及研究者的重视。从 2011 年到 2015 年连续举办五届的学习分析技术与知识国际会议(International Conference on Learning Analytics & Knowledge, LAK)主题来看,其关注的焦点由首届提出的学习分析在技术、社会与教育维度的整合^[1]到大数据

背景下利用学习分析技术改善教与学^[2],再聚焦到利用大数据技术实现对高等教育的实质性影响研究^[3],该会议的持续召开促使教育大数据学习分析的不断发展与创新,有助于利益相关者获得更为丰富、有意义的教与学信息。美国新媒体联盟(NMC)与北京师范大学智慧学习研究院合作的《2016 新媒体联盟中国基础教育技术展望:地平线项目区域报告》指出,大数据学习分析技术将在未来两至三年成为极具影响力的教育技术,并表明有效运用学习分析技术可以设计更好的教学活动,让学生积极主动地参与学习,准确定位处于危险中的学生群体,评估预测影响学生成功的因素^[4]。

二、研究现状

国外学术界研究大数据学习分析技术的时间较

基金项目:教育部人文社会科学研究规划基金项目“基于知识图谱的开放学习资源自主聚合研究”(项目编号:14YJA880103);基础
教育信息化技术湖南省重点实验室(项目编号:2015TP1017)

长、力度较深,并通过大量的实证研究获得丰富的研究成果。George Siemens 提出了包含数据收集、数据分析、数据预测、个性化调整等四个阶段的学习分析应用模式,通过收集分析学生数据可以判断学生的学习情况、及时发现教学中存在的问题,利用数据分析结果帮助教师预测、调整教学,以尽可能满足不同学生的学习需求,促进个性化学习的实现^[5]。Leah P. Macfadyen 等学者利用英属哥伦比亚大学本科生基于 BlackBoard 的完全在线生物课程追踪数据进行了学习分析的实证研究,通过对采集的学生数据进行简单相关分析、多元回归分析、Logistic 回归分析与社会网络分析,最终证明了挖掘 LMS 数据可以为教师开发“早期预警系统”的可行性,使教师能够依据预警系统识别处于危险中的学生并及时提供干预与支持^[6]。José A. Ruipérez-Valiente 等人开发了 Khan Academy 平台中学习分析扩展工具插件 ALAS-KA (Add-on of the Learning Analytics Support in the Khan Academy),利用 ALAS-KA 既能够分析整个班级的学习情况,也能够单独分析个体学生的学习风格,并且可以使分析结果可视化,这些可视化信息有助于教师与学生作出合理的教学决策与学习选择,以达到更好的理解与优化学习过程的目标^[7]。比利时鲁汶大学研发了角色展示平台工具 (ROLE Showcase Platform Tool) Student Activity Monitor,该工具可以提供诸如学生网上学习时长、文件使用数量、任务花费时间及学习者之间的比较等多种数据的可视化分析,通过数据分析及可视化呈现,帮助教师监控学习进程、发现教学中存在的问题,有助于学生了解自己和同伴的学习情况,提高教师与学生的自我反思能力^[8]。美国普渡大学为了应对其逐渐下降的新生保留率和日益延缓毕业的毕业生危机,于 2007 年启动了课程信号项目,其目的在于帮助学生尽快融入学习、获得学业成功,为此开发的 Course Signals 系统通过利用预测学生成功的算法 (Student Success Algorithm, SSA) 对学生进行分析预测,以确定其存在学习失败的可能性,由教师结合学生的学业表现对可能失败的学生给予及时的干预支持^[9];学习管理平台 Desire2Learn 开发的学生成功系统 (Student Success System) 可以基于学生在平台中留下的原始学习数据,利用语义分解、整体预测建模、学习分析技术与数据可视化技术来预测学生的学习状况,确定影响学生取得成功的重要因素,根据输出的可视化结果教师能够识别存在学业危险的学生并向其提供适当的干预服务,从而降低学生的辍学率,提高学生获得课程学习成功的概率^[10]。

相比于国外研究的多样化与实践性,国内大数据学习分析技术的研究在分析工具研发及实证研究方面还有待提高,大多集中于对学习分析技术的引介述评与模型搭建上。徐鹏等人通过解读美国国家教育部 2012 年发布的《通过教育数据挖掘和学习分析促进教与学》报告,对美国关于教育大数据的应用与挑战作出引介,进而指出我国应从技术、管理、法律三个层面加强对教育大数据学习分析的应用研究^[11]。李青等人利用文献分析法对学习分析技术的基本概念、组成要素、应用模型、技术来源、分析方法与分析工具进行了述评,并通过案例分析法整理了国外高校对学习分析技术的具体应用情况,剖析了学习分析技术的优势与不足,以期国内应用学习分析技术提供参考^[8]。孟玲玲、顾小清等人依据学习分析工具面向的对象和类型,将 24 种分析工具划分为学习网络分析工具、学习内容分析工具、学习能力分析工具、学习行为分析工具、综合学习分析工具,并对每类学习分析工具作出详细的比较,指出各类工具的优势与不足,以方便研究者和教师选择合适的工具加以运用,充分发挥学习分析技术的应用潜力及价值^[12]。姜强、赵蔚等以大数据分析为基础,从数据与环境、关益者、方法、目标等四个方面搭建了个性化自适应在线学习分析模型,阐述了数据产生的环境、相关的参与人员、大数据学习分析的应用方法及要实现的分析目标,期望通过该模型向每个学生提供自适应个性化的学习^[13]。武法提等学者通过梳理总结国内外学习分析模型存在的问题,根据个性化学习模式建立了个性化学习行为分析模型,并在此基础上设计了学习结果预测框架,旨在利用学习数据进行学习分析与结果预测,以发现存在学习问题与困难的学生,从而能够使教师为其提供个性化指导与支持,提高学生的学习质量^[14]。鉴于此,本研究利用大数据学习分析技术进行实际教学的实证分析,判定学习过程中影响学生学习绩效的预警因素,进而为其提供适当的干预支持,提高学习质量,促进教学效果的改善。

三、基于大数据学习分析的在线学习绩效预警因素判定

利用学习分析技术发现影响学生学习绩效的预警因素,提供及时的干预支持,是理解优化教学效果、实现个性化学习的有效途径。在线学习过程性数据的追踪与积累为学习分析技术的实施提供有力的依据。本研究是以 Moodle 平台中的一门大学课程为切入点,通过采集学生信息与学习行为数据形成初始数据

表 1

教学活动流程

周次	教学内容	学 习 活 动
第 1 周	html 标记语法基础 &html 文档基本结构	授课资源:PPT 课件;任务:使用 font、body、hr 及 br 等标记完成第一个网页;讨论:同学们经常访问的网站;自我评价:对提交的作业进行自评,给定成绩,写下自我反思
第 2 周	html 编辑器及 body 详解 &html 标记语义应用上集	授课资源:PPT 课件、颜色工具;任务:基于 html 基本骨架做一个类似百度检索后的效果;同伴评价:对同学提交的作业进行互评,给定成绩,写出评语;小测验:第 1~2 周的知识点
第 3 周	html 标记语义应用下集 &html 列表与图片应用	授课资源:PPT 课件、讲课案例、颜色工具、faststone 小工具;任务:以 163 为例,任选一个标题,模拟制作新闻信息页;讨论:运用颜色、faststone 等工具时的问题及解决方法;自我评价:对提交的作业进行自评,给定成绩,写下自我反思
第 4 周	html 表格应用基础	授课资源:PPT 课件、讲课案例;任务:以“百度搜索风云榜”为题,以 table 形式制作网页;同伴评价:对同学提交的作业进行互评,给定成绩,写出评语;小测验:第 3~4 周的知识点
第 5 周	html 超链接应用 &html 表单应用基础	授课资源:PPT 课件、讲课案例;任务:利用超链接制作类似 hao123 效果的网页;讨论:hao123 网页与淘宝首页的区别(制作设计方面);自我评价:对提交的作业进行自评,给定成绩,写下自我反思
第 6 周	html 表单详解 &html 多媒体应用	授课资源:PPT 课件、讲课案例;任务:制作一个注册表单信息;同伴评价:对同学提交的作业进行互评,给定成绩,写出评语;小测验:第 5~6 周的知识点
第 7 周	html 框架技术 &CSS 理解表现	授课资源:PPT 课件、讲课案例、字幕滚动;任务:综合 photoshop、flash 及 html 标签应用框架 frame;讨论:讨论网页制作过程的趣事及问题;同伴评价:对同学提交的作业进行互评,给定成绩,写出评语
第 8 周	总结复习	总任务:应用所学知识简单制作一个表达某主题的网站;交流讨论:初步学习“网页设计与开发”的感受;教师评价:对学生提交的综合作业进行点评,并给出成绩,写下评语;综合测验:前 7 周的知识点

集,利用散点图绘制、简单二元相关性分析、多元回归分析得到最终的预警因素判定模型,进而确定影响学生学习绩效的关键变量。

(一)问题提出

随着学习管理系统(Moodle、BlackBoard、Sakai)的广泛应用,国内越来越多的高校注重建设在线学习课程,采用在线学习方式进行教学,如清华大学、南开大学、吉林大学、东北师范大学等高校开设的部分必修课、选修课与通识课的在线课程。为了优化在线教学效果,有效提高学生学习成绩,需要解决三个主要问题:如何从学生产生的大量学习行为数据集中筛选影响学生学习绩效的活动变量?如何确定哪些活动变量最能影响学生的学习绩效?教师应该对哪些学习活动给予及时的关注、修改与优化?

(二)研究对象

选取东北师范大学计算机科学与技术学院 2015 级教育技术学的 40 名学生作为研究对象。为了保证最终形成的预警因素判定模型的准确性,需要提供合理有效的数据集,去除未完成全部或部分学习活动的学生,仅保留完成所有学习活动的学生样本,最终得到的有效样本为 38 人。

(三)教学活动设计

以教育技术专业 Moodle 平台中的“网页设计

与开发”必修课为切入点展开研究,具体教学活动流程见表 1。依据教学安排可知,学生的初步学习成绩评定包含在前八周的学习活动中,其中前七周的学习活动基本上由阅读授课资源、任务提交、实施评价及测验等四个环节构成,但是由于前七周教学内容的整体连贯性,分别在第一、三、五、七周授课结束时组织了课后交流讨论,以帮助学生解决面临的学习问题,而在第二、四、六周为学生设计了阶段测试题以评估学生的学习成效;而第八周则是总结复习先前所学内容,学生需要完成的学习活动包括综合任务的提交、实施评价、综合测验及交流讨论学习心得。在实施评价阶段,学生需要依据评价要求进行自我评价或者同伴互评,给定成绩,并写下自我反思或是评语,而教师则在第八周学习活动全部结束后依据学生提交的综合任务及学习过程中的学习表现进行点评,给出成绩;学生学习成绩的评定是按照 30%的自我评价、40%的同伴评价、30%的教师评价比例计算。

(四)数据处理与分析

学生在课程学习过程中会留下多种学习痕迹,如阅读授课资源次数、任务提交情况和实施评价情况等,这些学习痕迹自动保存在服务器日志中。因此,通过收集这些日志数据并进行数据过滤与筛选,最终选定了含有 21 个学习活动变量的初始数据集,分别是

总在线时间、浏览授课资源次数、在线测验次数、阅读讨论帖子量、发布讨论帖子量、回复讨论帖子量、总讨论帖子量、查看网页链接、使用搜索工具、访问成绩单、访问学习网站、访问聊天室、浏览新闻、查看新事件、查看课程动态、自我评价、同伴评价、教师评价、阅读任务要求、提交任务和完成任务时间;并运用散点图绘制方法得出这些变量与学生学习成绩是呈正相关的。

为了进一步发现影响学生学习绩效的关键变量,需要将各变量与学生的学习成绩进行简单的二元相关性分析,结果表明,选定的21个变量中有13个变量的Sig值小于.05,表明这13个变量(总讨论帖子量、浏览授课资源次数、总在线时间、同伴评价、自我评价、回复讨论帖子量、在线测验次数、提交任务、阅读讨论帖子量、发布讨论帖子量、教师评价、访问成绩单、完成任务时间)与学生学习成绩是呈显著正相关的,由此将Sig值大于.05的其他变量从初始数据集中剔除,便得到含有13个与学生成绩显著相关的变量数据集。在新得到的变量数据集中,有6个变量的Pearson相关系数是包含在(0.30~0.50)区间内,表明这6个变量(总讨论帖子量、浏览授课资源次数、总在线时间、同伴评价、自我评价、回复讨论帖子量)与学习成绩之间的相关程度是中等强度的;而余下7个变量的Pearson相关系数是包含在(0.10~0.30)区间,表明这些变量与学习成绩之间的相关程度是弱相关。

尽管获得的13个变量与学生学习成绩之间是呈显著正相关的,但是仅依靠简单二元相关性分析来判断影响学生学习绩效的预警因素是不可靠的,容易出现差错,而且所实施的相关性分析也并不能完全解释学生复杂的学习行为,因此需要利用多元回归方法对含有13个变量的数据集进行回归分析,依据分析结果中的回归系数,最终确定影响学生学习绩效的预警因素,见表2。

由表2可以看出,影响学生学习绩效的预警因素包括讨论区总发帖数量、在线测验次数、同伴评价、自我评价、提交任务与浏览授课资源次数。其中,讨论区总发帖数量对学习成绩的影响最显著,说明讨论交流活动促进学生之间的思维碰撞与信息共享,有助于激发学生学习的积极性;学生通过在线测验检查所学知识的掌握程度,发现遗漏的学习内容与问题,及时进行补充以保证学习的完整性与有效性;同伴评价能够使学生通过评价他人提交的任务弥补自己任务中的不足之处,取长补短以改善完成任务的质量;自我评价在于积极推动学生进行自我反思,促进学生养成自我反省的习惯,以改进提交的任务质量;以任务驱动

的方式促使学生完成任务并及时提交,有利于巩固学生习得的知识与技能,了解学生对学习内容的掌握程度;浏览授课资源可以帮助学生更好地理解学习内容,娴熟地运用所学知识完成学习任务,进而提高学习效率,促进学生积极完成学习活动。因此,最终得到影响在线学习绩效的回归方程是 $Y = -3.468 + 0.443X_1 + 0.219X_2 + 0.182X_3 + 0.176X_4 + 0.151X_5 + 0.098X_6$ 。其中,方程中因变量Y为学生的学习成绩,自变量X为各类学习活动。

表2 多元回归相关系数

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
6	(Constant)	-3.468	.261		-1.904	.000
	总讨论帖子量	.443	.087	.443	5.073	.000
	在线测验次数	.219	.083	.219	3.249	.000
	同伴评价	.182	.066	.182	2.618	.000
	自我评价	.176	.059	.176	2.057	.000
	浏览授课资源次数	.151	.052	.151	1.862	.000
	提交任务	.098	.037	.098	1.306	.000

注:a. Dependent Variable: 成绩。

四、基于大数据学习分析的在线学习干预模型构建

依据上述提出的影响在线学习绩效的回归方程,本研究基于数据挖掘算法和学习分析技术,设计构建了在线学习干预模型,提出了有针对性的干预对策,实施个性化指导,解决学习危机,提高学习质量,如图1所示。

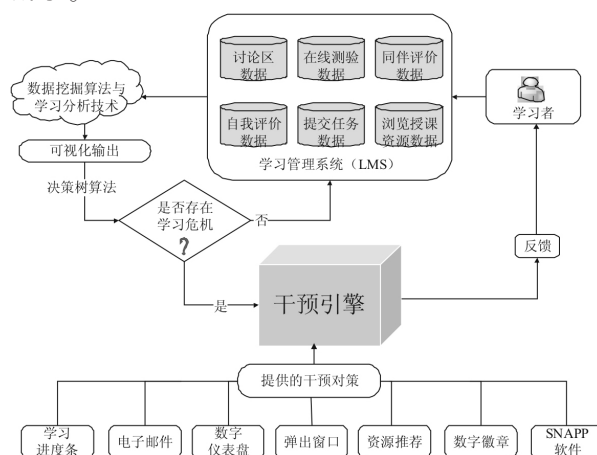


图1 在线学习干预模型^[15]

该模型旨在利用数据挖掘与分析工具对学习管理系统存储的大量影响学生学习成绩的讨论区总发帖量、在线测验次数、同伴评价、自我评价、提交任务与浏览授课资源等相关学习行为数据进行整合分析,以了解学习者当前阶段的学习情况,并运用可视化技术将分析结果以图表、数字等形式输出;依据分析结果呈现的学习情况,结合学生在学习管理系统的学习表现,利用决策树算法进行危机诊断,以判断当前阶段学习者是否存在学习危机;如果当前阶段的诊断结果显示学习者可能存在学习危机,则会为学习者提供适当的干预对策;如果诊断结果显示的是学习者不存在学习危机的可能性,那么会利用数据挖掘算法与学习分析技术持续对学习者的相关学习行为数据进行收集分析,对分析结果进行可视化输出,时刻监控学习者的学习过程、明晰学习情况,以防遗漏可能在下一阶段的学习中存在学习危机的学生。

此外,所设计建构的干预模型中,对于学生学习危机的诊断是伴随着学习活动的持续开展而循环进行的,即通过学习管理系统存储数据的不断更新,可以实时分析学生的学习情况,时刻监督学生的学习进展。一旦诊断出学生在某一阶段可能存在学习危机,将会及时为其提供适当的干预对策,帮助学生顺利完成课程的学习,进而取得课程学习的成功,保证课程学习的质量。

在干预引擎中,分别从学习进度条、电子邮件、数字仪表盘、弹出测试窗口、资源推荐、数字徽章、SNAPP 软件等七个方面提取干预对策,为学生提供学习支持。其中,学生通过查看自己的学习进度条可以起到自我监控的作用,访问他人的学习进度条则可以激励自己加快学习速度,积极完成学习任务,从而使其快速融入学习过程以保证学习的可持续性、持久性,增强学生的学习毅力;以电子邮件的形式向学生发送警示通知信息,提醒学生按时登录学习平台,及时提交学习任务并进行讨论交流与任务评价,从而帮

助学生完成阶段性课程内容的学习,为顺利开展下一阶段的学习提供支持;数字仪表盘中呈现的学习活动完成情况,既可以帮助学生准确地评估自己的学习成效并进行自我反思,又可以作为学生评价他人的依据为其提出客观合理的学习建议,促进学生之间的相互学习,加强学习经验的交流,以共同提高学习的质量;以弹出窗口的形式向学生提供当前学习内容的相关测试题,促使学生进行自主测试,发现存在的学习盲区并及时给予弥补,以保证学生的学习质量,同时起到缓解学习疲劳的作用;以资源推荐的形式向学生推送能够帮助其脱离学习困境的各类资源,包含学习伙伴的选择、课程学习资源的补充及学习工具的利用等,通过推荐给学生适当的资源供其有效运用来优化学习过程、提高学习效率;以数字徽章的形式认可学生付出的努力,借以激发学生的学习兴趣,鼓励学生充分发挥学习主动性来完成课程内容的学习,增强学生的学习动机,进而提高学生取得课程成功的概率;利用 SNAPP 社会网络分析软件实时收集分析学生的讨论区数据,以社会网络结构图的形式真实展现学生参与讨论交流的情况,有助于学生明晰其在讨论活动中的角色,激发学生参与讨论的积极性与主动性。

五、干预效果实证分析

将建构的在线学习干预模型应用于第 9 周至第 12 周的教学实践中,对学生的学习情况展开实时的监控追踪,以识别可能存在学习危机的学生,并及时向其提供适当的干预对策。模型实施过程通过收集在线平台中学生第 9~12 周的相关学习行为数据,利用二元 Logistic 回归分析法对干预模型的可行性与有效性进行验证,并结合具体数据统计结果来证明在线平台提供的干预对策有助于学生学习质量的提高、教学效果的改善。此外,还以学习该课程的学生为调查对象进行题为“在线学习干预模型效果评价问卷调查”,并在调查结束之后与学生进行谈话,从而主观说明干

表 3 最终方程中的变量

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a	总讨论帖子量	.544	.098	31.090	1	.000	1.723
	在线测验次数	.362	.057	24.611	1	.000	1.425
	同伴评价	.083	.021	15.728	1	.000	1.086
	自我评价	.069	.023	9.027	1	.000	.933
	提交任务	.051	.016	6.841	1	.000	.587
	浏览授课资源次数	.028	.009	3.052	1	.000	.410
	(Constant)	-1.181	.275	18.505	1	.000	.307

注:a. Variable(s) entered on step 1:总讨论帖子量,在线测验,同伴评价,自我评价,提交任务,浏览授课资源。

预模型的有效性。

(一)二元 Logistic 回归分析

二元 Logistic 回归分析法可以验证建构的在线学习干预模型的可行性与有效性。依据学科成绩的规定标准,对学生的学习成绩进行编码赋值:将低于 60 的分数赋值为 0,表示学生存在学习危机;高于或等于 60 的分数赋值为 1,表示学生不存在学习危机;利用经过编码整理的数据进行二元 Logistic 回归分析,得到最终方程中的变量(见表 3)和最终的预测结果分类。

由表 3 方程中变量可以得出 Logistic 回归模型的表达式为: $P(Y)=1/1+e^{-(-1.181+0.544* \text{总讨论帖子量}+0.362* \text{在线测验次数}+0.083* \text{同伴评价}+0.069* \text{自我评价}+0.051* \text{提交任务}+0.028* \text{浏览授课资源次数})}$,这表明利用 Logistic 回归分析最终确定的活动变量与判定的在线学习绩效预警因素是相一致的。

最终的预测结果分类显示,在 38 位学生中有 17 位学生是存在学习危机的,对 17 位存在学习危机的学生进行预测,其中成功预测了 12 位学生,有 5 位学生预测失败,预测的成功率达到 70.6%;而对余下的 21 位不存在学习危机的学生进行预测,其中成功预测了 16 位学生,有 5 位学生预测失败,预测的成功率达到 76.2%。从总体情况来看,达到的预测成功率是 73.7%,也就是说,二元 Logistic 回归模型能够以 73.7%的正确率来证明所建构的在线学习干预模型具有一定的可行性与有效性。

鉴于此,本研究提取学生第 9~12 周的学习数据进行分析统计,将统计结果与前 8 周的学习数据分析结果加以比较,得出在线平台实施干预对策前后的学

习任务完成情况:提交任务由 300 增至 379;同伴评价由 218 增至 336;自我评价由 240 增至 364;讨论交流由 88 增至 312;浏览授课资源由 123 增至 358;在线测验由 162 增至 339。

对比结果显示,向学习者提供的干预对策能够推动其积极参与学习,表现在提交任务、同伴评价、自我评价、讨论交流、浏览授课资源及在线测验等数量的提升上。

此外,对学生前 8 周和后 4 周习得的知识量进行统计比较,如图 2 所示。

根据布鲁姆的教育目标分为记忆、理解、应用、分析、评价和创造等 6 个级别,每个级别学习内容由领域专家划分为事实性知识、概念性知识、程序性知识和元认知知识。由图 2 可知,学生在后 4 周习得的知识量虽然并不明显多于前 8 周,但由于后 4 周的学习时间较短,从而说明了学生的学习效果有所提升,也证明了所提供的干预对策有助于提高学习质量。

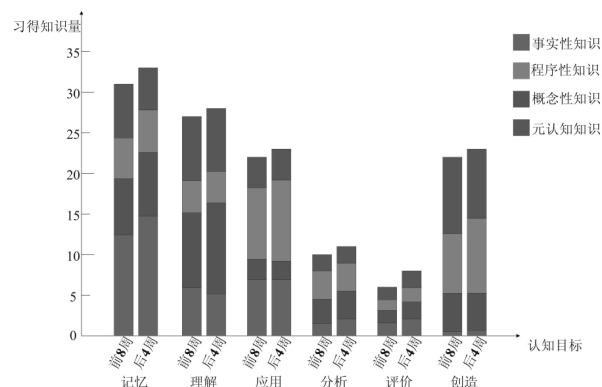


图2 在线平台学习成效统计

表 4

在线学习干预效果评价的问卷反馈结果

问 题	等 级				
	非常同意	同意	无所谓	不同意	非常不同意
1.在线平台中显现的学习进度条能够促使我积极参与学习活动,增强我的学习动机	10(27.78%)	19(52.78%)	4(11.11%)	3(8.33%)	0(0%)
2.在线平台发送的电子邮件能够提醒我及时登录平台、提交任务、进行交流讨论与任务评价	8(22.22%)	18(50%)	7(19.45%)	3(8.33%)	0(0%)
3.在线平台中的数字仪表盘能够使我了解自己的学习问题,及时进行查漏补缺	9(25%)	18(50%)	8(22.22%)	1(2.78%)	0(0%)
4.课程学习中时常弹出的测试窗口有助于了解知识的掌握情况	5(13.89%)	12(33.33%)	12(33.33%)	7(19.45%)	0(0%)
5.在线平台推送的学习资源能够满足我的学习需求,优化学习过程	13(36.11%)	17(47.22%)	4(11.11%)	2(5.56%)	0(0%)
6.在线平台中数字徽章的获取能够激发我的学习兴趣,激励我快速有效完成任务	5(13.89%)	17(47.22%)	13(36.11%)	1(2.78%)	0(0%)
7.在线平台中 SNAPP 软件呈现的社会网络结构图能够激励我积极参与讨论交流	8(22.22%)	20(55.56%)	5(13.89%)	3(8.33%)	0(0%)

(二) 问卷调查反馈

为了获得学生对所提供干预对策的主观评价,本研究实际发放问卷调查 38 份,回收 36 份(一份问卷中同一题目填写了 2 个选项,另一份则全部选择了无所谓选项,均视为无效问卷),有效回收率为 94.7%,统计结果见表 4。

从反馈结果看,学生基本满意所提供的干预对策,有利于激发学习兴趣,增强学习动机,培养学习毅力。其中 80.56% 的学生认为查看学习进度条能够增强学习动机,积极参与学习活动。72.22% 的学生相信发送的电子邮件能够起到警告提醒的作用,比如避免出现逾期未提交任务情况。75% 的学生认为通过查看仪表盘能够发现存在的 learning 问题,及时进行查漏补缺以保证学习质量。77.78% 的学生认可运用 SNAPP 软件有利于积极主动参与讨论交流活动。83.33% 的学生认为平台推送的资源能够满足学习需求,优化学习过程。正如学生 E 说:“我认为平台推送的有些学习资源是没用的,但大多数情况下我还是采用的,因为我坚信它们能够帮助我顺利完成学习”。61.11% 的学生坚信数字徽章的赚取有利于激发学习兴趣,提高学习效率。学生 C 说:“我对平台中数字徽章的获取是非常感兴趣的,它不仅督促我加快速度完成学习任务,还可以激发我潜在的学习竞争力”。值得注意的是,针对课程学习中时常弹出测试窗口,仅有 47.22% 的学生认可。学生 D 说:“我认为弹出的测试窗口会打断我的学习,分

散我的注意力,而且还会影响我进行后续的学习”。

六、结 论

学习分析技术可以将大量学生数据转换成有价值的教学信息,通过对最终分析结果的可视化展示,能够发现影响学生学习成绩的关键因素,识别可能存在于学习危机的学生,从而为进行学业预警、教学干预、教学决策等服务提供有力支持,以促使学生取得学习成功,提高学习质量,达到优化教学效果的目标,实现真正意义的个性化学习。本研究从 Moodle 平台中“网页设计与开发”课程前 8 周学生产生的大量学习行为数据中筛选出 21 个与学习成绩呈正相关的活动变量,并利用多元回归分析法确定了 6 个影响学生学习绩效的预警因素。在此基础上设计构建了在线学习干预模型,将其应用于第 9~12 周的教学,利用二元 Logistic 回归分析法验证了可行性与有效性。此外,在前 8 周与后 4 周的学习活动情况与习得知识量的比较中,发现后 4 周的各项学习活动完成情况均有提高,而且最终习得的知识量也有所增多,进一步证明了本研究提出的干预对策有助于提高学生的学习质量。此外,通过问卷调查与访谈的反馈结果,表明了学生基本满意所提供的干预对策,从而引证了干预对策的有效性。然而,由于特定教学活动的限制,且样本数较少,因此下一步重点将研究成果应用到大规模在线课程学习体系中。

[参考文献]

- [1] LAK11. 1st international conference on learning analytics and knowledge [EB/OL]. (2010-07-22)[2011-02-27]. <https://tekri.athabasca.ca/analytics/>.
- [2] LAK13. Third conference on learning analytics and knowledge [EB/OL]. (2013-01-23)[2013-04-08]. <https://lakconference2013.wordpress.com/>.
- [3] LAK15. The 5th international conference on learning analytics and knowledge [EB/OL]. [2015-03-16]. <https://lak15.solaresearch.org/>.
- [4] JOHNSON L, LIU D, HUANG R, ADAMS BECKER S, GAO Y, CUMMINS M, DAVIS A, ESTRADA V. 2016 NMC technology outlook for Chinese K-12 education: a horizon project regional report[R]. Austin, Texas: The New Media Consortium, 2016.
- [5] SIEMENS G. What are learning analytics [EB/OL]. (2010-08-25)[2011-01-16]. <http://www.elearn.space.org/blog/2010/08/25/what-are-learning-analytics/>.
- [6] MACFADVEN L P, DAWSON S. Mining LMS data to develop an “early warning system” for educators: a proof of concept[J]. Computer & Education, 2010(54): 588-599.
- [7] RUIPEREZ-VALIENTE J A, MUNOZ-MERINO P J, LEONY D, DELGADO KLOOS C. ALAS-KA: a learning analytics extension for better understanding the learning process in the Khan academy platform[J]. Computers in human behavior, 2015(47): 139-148.
- [8] 李青,王涛.学习分析技术研究与应用现状评述[J].中国电化教育, 2012(8): 129-133.
- [9] PISTILLI M D, ARNOLD K E. Purdue signals: mining real-time academic data to enhance student success [J]. About campus: , 2010, 15(3): 22-24.
- [10] 移动 LABS.学习管理平台 Desire2Learn 野心不小,想改变大学生学习生涯规划方式 [EB/OL]. (2013-05-11)[2015-12-10]. <http://>

labs.chinamobile.com/news/94091.

- [11] 徐鹏,王以宁,刘艳华,等.大数据视角分析学习变革——美国《通过教育数据挖掘和学习分析促进教与学》报告解读与启示[J]. 远程教育杂志,2013(6):11-17.
- [12] 孟玲玲,顾小清,李泽.学习分析工具比较研究[J].开放教育研究,2014(4):66-75.
- [13] 姜强,赵蔚,等.基于大数据的个性化自适应在线学习分析模型及实现[J].中国电化教育,2015(1):85-92.
- [14] 武法提,牟智佳.基于学习者个性化行为分析的学习结果预测框架设计研究[J].中国电化教育,2016(01):41-48.
- [15] WU F, HUANG L, ZOU R. The design of intervention model and strategy based on the behavior data of learners: A learning analytics perspective [C]//Simon K S, Cheung. Hybrid learning: Innovation in educational practices. Switzerland:Springer International Publishing,2015:294-301.

Empirical Research of Predictive Factors and Intervention Countermeasures of Online Learning Performance on Big Data-based Learning Analytics

ZHAO Huiqiong¹, JIANG Qiang¹, ZHAO Wei¹, LI Yongfan², ZHAO Yan³

(1.School of Computer Science and Information Technology, Northeast Normal University, Changchun Jilin 130117;2.College of Informational Science and Engineer, Hunan No. 1 Normal College, Changsha Hunan, 410205; 3.Changchun Institute of Education, Changchun Jilin 130061)

[Abstract] To Mine educational data by learning analytics will be helpful to discover the potential value of the data and then transfer them into meaningful information, optimize learning process and improve teaching effects, which becomes the common demands of teachers, students and educational researchers. From the perspective of educational application of learning analytics, an intervention model has been constructed based on tracing, accumulating and selecting the online learning behavior data in Moodle, and using multiple regression to determine predictive factors affecting students' learning performance. After that, this model is applied to teaching practice, and the learning behavior data produced is analyzed by binary logistic regression. The validity of the model in learning activities, knowledge acquisition etc. has been verified through questionnaire and interview. The results indicate that through the intervention model, the students with learning crises can be identified. A timely warning sign and personalized intervention can be helpful to enhance students' learning motivations, to cultivate their learning perseverance and improve their quality of learning as well.

[Keywords] Big Data in Education; Learning Analytics; Warning; Intervention Countermeasures; Personalized Learning