

doi:10.3772/j.issn.2095-915x.2016.03.006

大数据时代学习平台的数据统计分析功能研究

——以 Canvas 和 Sakai 系统为例

黄贇茹

(北京邮电大学网络教育学院 北京 100876)

摘要: 随着大数据时代的到来,数据挖掘分析已经渐渐应用到了各行各业。大数据分析能够获得隐藏在庞大数据中的关系,趋势信息等等。通过数据挖掘的方式对互联网信息进行深层次解读将会得到更重要的信息而为之后的工作服务。目前教育者们已经重视到数据分析在教育中的应用,尤其在教学平台中如何设计数据分析功能已经是当今教学平台发展的一个重要环节。本文通过对 Sakai 和 Canvas 两个世界知名的学习管理系统的数据统计分析功能的比较分析,从平台数据分类方式,展示方式,数据内容,以及特点四个方面的研究,期望能使网络教育者对于平台数据统计分析功能的开发有所启示,促进学习分析在学习平台的应用,为学习平台实现自适应教学做铺垫。

关键词: 教学平台, 数据统计, 数据分析

中图分类号: G434

Research on Data Statistics and Analysis Function of LMS in the Era of Big Data

——Case Study of Canvas and Sakai

HUANG YunRu

(Institute of Network Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: With the advent of big data, technology of data mining and analysis has been applied to life gradually. Big data analysis can obtain information which are hidden in the huge data and analyze the relationship, and trends of information. The data mining can retrieve more important information from the Internet for the users.

作者简介: 黄贇茹 (1992-), 女, 蒙古族, 北京邮电大学网络教育学院研究生, 研究方向: 远程教育。

Recently, educators have paid great efforts to apply the data analysis in the education, especially for the teaching platform. However, how to design the function of data analysis is an important section in the development of teaching platform. In this research, we compared the function of statistics analysis of Canvas and Sakai, which are two of the most popular learning management systems. Comprehensive methods were applied to analyze the two platforms from four aspects: data classification, display model, data content and characteristics. We expected to inspire the network educational researchers on the development of data statistical analysis platform, to promote applications of learning analysis in learning platform, and to pave the way for the realization of adaptive teaching for learning platform.

Keywords: Learning management system, data statistics, data analysis

引言

随着教育信息化的不断深入, 纷繁复杂、功能各异的学习管理系统纷纷出现, 这些学习管理系统中存储着海量的学习者信息以及学习过程数据, 身处大数据环境的教育领域已经积累了海量的数据。如何将这些数据转变为有用的信息, 对教学提供帮助, 已经成为教育工作者关注的内容。利用海量数据进行挖掘分析, 对学习者的学习效果进行预测、分析, 合理有效地利用这些数据可以提升学习者学习效率, 提升教师的教学效果^[1]。目前各类教学系统的数据收集分析以及应用方式也是多种多样的, 为了将大数据环境下教学系统的数据分析功能设计的更加合理化, 对目前教学系统的数据统计分析功能进行研究是很有必要的^[2]。

1 平台数据统计分析的功能需求

2010 版地平线报告中提出了需要重视教育可视化数据分析, 2011 年版地平线报告提出学习分析的概念, 从 2012 年版至 2015 年版地平线报告中反复强调“学习分析”是影响当代学习科学技术发展的主要方法与技术并将全面介入教育

教学^[3]。可见学习分析与评价对于教学的发展有着非常重要的意义。高校的学习管理系统作为主流的学生在线学习的重要工具, 将学生在学习过程中产生的海量数据进行分析, 使得教育信息变为可量化的数字存在, 通过真实数据发现教与学的关系, 是目前教学管理系统性能提升的重要革命^[4]。对于大数据支持的学习分析与需求的调查研究结果中显示, 绝大多数的教师能够认同将大数据应用到教学中的分析与评价的价值。相信利用可视化数据和数据分析能够提升学习分析和评价的客观性, 透明度与公正性, 及时性。不仅能节省教师的时间和经历, 实现自动量化分析, 促进教师更加全面了解学生的学习需求和学习现状, 同时也促进了学生对自我学习过程的了解, 及时的调节学习方法, 更加积极的参与到学习中去^[5]。通过对目前已有的文献进行调查研究, 总结得出目前教育工作者对于教学平台能提供的数据统计与分析的功用需求有以下几点:

1.1 使学生及时获得学习反馈

学习平台是学生进行自主学习的工具, 学生在学习过程中大部分时间是没有教师陪伴的, 因此许多学生在学习过程中并不能很好的了解自己的学习进度, 调整自己的学习重点和方向, 常常

容易迷失自己。学习平台需要将学生的学习数据进行统计分析,让学生获得及时的学习反馈,学生可以及时的对自己的学习情况有所了解,对薄弱环节进行加深学习,对掌握的知识进行巩固,有效的调整自己的学习进度,改变学习方式。因此将学习数据进行统计分析,给予学生及时的反馈将有助于提升学生的自主学习意识和能力^[6]。

1.2 使教师快速了解学生,促进教学发展

将数据分析加入到学习平台中能够让教师快速准确的了解到学生的个性化特征,以及学生的学习行为,教师可以通过这些信息对学生进行个性化指导,因材施教。另一方面,教师也可以从学习数据分析中了解到自身教学方法的优劣,进一步对自己的教学实施效果进行了解,促进了教师对于教学过程中产生的问题作出及时调整,优化教学内容和教学方式,有利于自身教学素养的提升^[7]。

1.3 实现多元化评价功能

多元化的评价功能主要包括教师评价、学生自评与学生互评。实现多元化的评价功能有助于提升学生的学习绩效,优化学生的学习效果以及提升教师的教学效果。另外,第三方评价也有助于教学活动的进行,例如家长评价等。多元化的评价功能可以采集到更多考试以外的数据,将这些数据统计后提供给教师可以帮助教师全方位多元化的了解学生的学习状况,为教师的教学指导提供更多信息;提供给学生可以让学生更客观地了解自己的学习状况。

1.4 学习平台的统计分析工具具有实用性,便捷性。

学习平台的数据分析实用性主要表现为数据

采集可靠合理,分析全面。将学生在学习过程中产生的学习数据进行采集存储,经过科学设计的分析方法对数据进行全面分析,是学习平台数据分析功能的根本需求。学习平台的便捷性表现为分析结果直观与操作方式简单两个方面。分析结果的直观性影响了教师对教学过程的判断,操作方式的简单与否直接影响了平台的使用,因此学习平台的数据分析实用性和便捷性至关重要。

2 Canvas 与 Sakai 平台统计分析功能介绍

Canvas 是美国 Instructure 公司的一个学习管理系统,Instructure 公司创建于 2008 年,Canvas 开源于 2011 年。Canvas 平台是新一代教学平台,它的出现改变了传统教学平台的发展方向,它独特的云服务架构设计,教与学工具整合应用,强大的数据分析功能以及支持移动用户是它的四大特征^[8]。

Sakai 是由美国印第安纳大学,密西根大学,斯坦福大学和麻省理工学院于 2004 年发起的一项开放源代码的课程与教学系统开发计划。Sakai 是一个类似于 moodle 的课程管理学习系统,以及虚拟学习环境,是基于 java 的面向服务的应用程序开发,具有可靠性,协作性和可扩展性^[9]。

2.1 Sakai 数据统计

Sakai 平台数据统计功能在“站点统计”中展示。站点统计功能的使用权限只有管理员,站点统计是对某站点的数据进行统计。站点统计的三个功能是:观看概览,自定义报表,修改参数。

2.1.1 数据统计

数据统计分为一级分类和二级分类,展示方式分为图和表两种。具体内容如表 1 所示。

表1 Sakai 数据统计

数据一级分类	查看方式	数据二级分类	统计内容
访问	按日期	访问量	访问累计数
	按用户	访问人数	人数
		本站点成员	人数
		访问过本站点的站点成员	人数 + 占总数百分比
		没有访问过本站点的站点成员	人数 + 占总数百分比
活动	按日期	最活跃用户	用户 id+ 百分比
	按用户	事件	读取帖子, 读取资源, 发布聊天信息, 发布通知, 删除聊天信息, 添加资源, 新建帖子, 修改资源
	按工具	使用最多的工具	名字 + 百分比
资源	按日期	文件	名字
	按用户	文件打开次数	次数 + 百分比
	按资源	被打开次数最多的文件	名字
		打开文件最多的用户	用户 Id

2.1.2 站点统计

管理员可以自定义报表, 自定义内容包括: 报表、内容、时间、用户、展现方式。并且可以修改统计的参数, 主要是活动部分的事件或者工具部分的选择, 其中所有工具事件包括: 通知, 作业, 聊天答疑, 成绩册, 邮件列表, 新闻, 讨论区, 站内消息, 播客, 调查工具, 资源, 练习测验, 检测工具, 日程, 班组管理, 站点信息,

课程大纲, Wiki, 站点统计, 链接内容, 课程组织, 外部工具。

2.2 Canvas 数据统计

Canvas 数据统计分为管理员视角与教师视角。

管理员视角中分为数据统计和分析两个部分。统计中记录了一些常规信息, 具体内容如表2所示。

表2 管理员视角下的数据统计

数据分类	数据内容
常规信息	生成时间; 课程; 教师; 学生; 用户; 文件存储 (个数 + 大小); 媒体 (个数 + 大小)
最近开始的课程	名字
最近结束的课程	名字
最近登陆的用户	Id

分析中记录了学习者学习活动中产生的具体数据，具体内容如表 3 所示。所有部分的展现方式都是图表类型，当用户将鼠标移动到图表中时，会看到某节点的具体数据，非常清晰准确的记录了用户数据。

教师视角的数据统计分析功能在“查看分析”

表 3 管理员视角下的课程分析

分类	内容
按日期划分的活动	时间 + 活动项数 + 页面视图数
按类别划分的活动	类别 + 页面视图数
评分分布	百分比学生 + 分数
课程统计	总计，作业，学生，文件存储
讨论	讨论帖，新建讨论，新建讨论条目
作业	作业小组，课程标准说明，所有评分标准说明
活动学生	为接受的学生
测验	测验题，测验提交

中，指对一门课程的课程分析。其中所有数据的展示方式都是图表类型，用户可以将鼠标放置图中观看各个节点的详细数据。数据分析分类为一级分类和二级分类。具体内容如表 4 所示。教师可以查看学生的互动报告，并且可以在学生的评分中看到作业的具体信息，截止时间，评改内容，分数以及学习的掌握程度（某项内容的掌握程度体现在相关作业的最高分数）。

表 4 教师视角的课程分析

分析一级分类	分析内容	二级分类	分析内容
活动	时间 + 参与项 + 页面视图数		
提交文件	作业名称 + 到期日期 + 缺少百分比 + 准时百分比		
评分	名称 + 高分 + 中等分 + 低分		
学生	页视图; 参与; 提交文件; 当前分数	活动	时间 + 页面查看数
		通信	学生与老师通信数 + 时间
		提交文件	作业名称 + 按时提交 + 没按时提交 + 未交
			为接受的学生
		评分	课程名称 + 最高分 + 中等分 + 最低分 + 该同学该课程分

3 Canvas, Sakai 平台统计分析功能对比

3.1 数据分类方式对比

Sakai 平台和 Canvas 平台在数据统计的分类方式上有很大的不同。Sakai 平台将所有数据分为访问、活动和资源三类,在每一类数据中可以对其选择进行日期、工具或事件的不同展示方式。用户可以明确的查询到自己想要的的数据属于哪一种类型。Canvas 平台认为所有数据的产生都来自于“活动”,它将“活动”数据分别按照时间、类别划分。用户可以通过数据的展示来了解到自己在学习过程中的变化情况。Sakai 平台的分类方式属于树形结构,注重将数据的内容进行分类。而 Canvas 平台的数据分类是扁平化的,更注重学习者在学习过程中数据变化的展示,让教师了解课程的发展情况,学生学习的变化情况。

3.2 数据展示方式对比

Sakai 的数据展示方式是图表结合表格。图的类型有柱状图,饼状图,时序图三种,用户可以根据自己的喜好来选择展示的方式。Canvas 的数据展示方式几乎全部用表完成的。并且其中可以

通过将鼠标放置在表上来获取具体某节点的数据内容。Sakai 在数据展示方面更加注重数据内容和配图相结合,尽量把所有内容都显示在屏幕上,而 Canvas 在数据展示方面更加注重简洁直观性,不直接将所有数据都展示在界面上,但用户可以看到每个节点的详细数据。

3.3 数据统计内容对比

将数据统计分为访问数据,课程资源数据,教师活动数据和学生活动数据四个方面来看,其中数据收集量方面,Sakai 比 Canvas 能收集到的数据更多。Canvas 更重视学生学习行为的数据收集,教师的行为数据记录很少,即使是教师日志,在展示中也是和学生数据绑定在一起的,为了更好的了解学生的学习过程。而 Sakai 将教师和学生都看作用户,对所有用户的行为数据都进行收集。Canvas 统计数据的方式主要是按照学生在学习活动中会产生的学习行为进行分类统计,Sakai 是将学生和教师在课程中大部分可能产生的行为都记录下来。因此在数据统计方面 Sakai 的记录项更加详细,能涉及学生和教师的每一个具体行为,而 Canvas 更注重将活动数据分类,记录行为活动中需要的数据,而不是任何行为数据都进行记录。具体内容如表 5 所示。

表 5 数据统计内容对比

数据统计	Canvas 具体记录项	记录内容	Sakai 具体记录项	记录内容
访问数据	教师人数	数量	访问量	数量
	学生人数	数量	访问人数	数量
	用户人数	数量	本站点成员	成员名称
	最近登陆的用户	用户名称	访问过本站点的站点成员	成员名称
	学生浏览页面	按时间分布的浏览页面数量	没有访问过本站点的站点成员	成员名称
	教师浏览页面	按时间分布的浏览页面数量		

课程资源数据	1. 课程	数量	1. 文件	数量, 大小
	2. 最近开始的课程	名称	2. 文件打开次数	数量
	3. 最近结束的课程	名称	3. 被打开次数最多的文件	名称
	4. 文件存储	大小	4. 打开文件最多的用户	用户名称
	5. 媒体	大小		
教师活动数据	1. 通信	按时间分布的通信数量, 通信对象	1. 读取帖子	读取时间, 帖子名称
	2. 教师日志	按时间排列的教师日志内容	2. 读取资源	读取时间, 资源名称
			3. 发布通知	发布时间
			4. 添加资源	添加时间, 资源名称
			5. 删除聊天信息	删除时间
			6. 修改资源	修改时间, 资源名称
			7. 发布聊天信息	发布时间
			8. 新建帖子	新建时间, 帖子名称
	浏览视图	按时间分布的浏览视图数量、按活动分布的浏览视图数量	1. 读取帖子	读取时间, 帖子名称
	讨论	讨论时间, 参与比重	2. 读取资源	读取时间, 资源名称
学生活动数据	作业	作业分数, 递交时间	3. 发布通知	发布时间
	考试	考试分数, 排名	4. 添加资源	添加时间, 资源名称
	通信	按时间分布的通信数量, 通信对象	5. 删除聊天信息	删除时间
			6. 修改资源	修改时间, 资源名称
			7. 发布聊天信息	发布时间
			8. 新建帖子	新建时间, 帖子名称

3.4 两平台数据统计特点对比

Canvas 数据统计的特点在于简单直接的将教师和管理员需要的数据展现出来, 追求数据的精致, 而不追求数量。Canvas 将统计内容细化, 并且数据十分精确, 更加注重教学过程中的数据变化。因此在教学过程中, 管理员能够更加清晰地了解目前平台的使用状况, 教师能够更加直观地

对学生的学习行为进行了解, 对学生进行评价。而且 Canvas 还有数据分析的功能, 它可以将数据进行简单的分析, 让教师更好的对学生进行评价。

Sakai 数据统计的特点与 Canvas 相反, 它更加注重数据收集的全面性, 将每一个用户所能产生的行为都进行统计, 无论是管理员还是教师想要什么数据都可以从平台中调取出来。并且可以根据自己的需求来自定义报表, 定制出个性化的

展示方式与统计数据。

3.5 总体对比和建议

目前美国在学习平台的数据利用上处于领先的地位,早在2004年,有研究者利用聚类算法对课程中的用户行为进行研究,找出学习行为相近的学习者。将数据挖掘应用于测量学生在在线学习论坛中的贡献值。2005年,美国佐治亚州大学的研究人员对高中学生的GPA和SAT数学成绩通过判别式分析对他们的在线通识教育课程的完成情况进行预测。坎贝尔在普渡大学发表的论文中,对课程管理系统中的数据和人口统计信息数据利用因子分析和逻辑回归分析方法研究出一个能预测学生学习结果的预测模型。罗梅罗等人利用数据挖掘技术在Moodle平台上进行研究。近年来Bravo Agapito等人利用c4.5决策树规则来对在线学习系统中的学习不佳表现进行判别分析。2007年普渡大学就提出了通过对学生的素质评价和在线学习行为数据对学生的成绩进行预测的预测模型,并在10年基于普渡大学的Blackboard上实现了预测模型的开发部署,在实际使用中取得了非常好的结果^[10]。马里斯特大学在12年对普渡大学的预测模型进行了扩展研究,提出了OAAI预测模型,在对学生平台数据进行分析的同时能对学生进行自适应的引导,进一步利用大数据分析学习方法对学习数据处理,并将学习数据分析模型扩展到了Sakai平台,并且精准度进一步提升,目前已经可以部署在马力斯特大学的Sakai平台上,并且将源代码公开,所有的机构和学者都可以在此基础上进行扩展研究^[11]。目前国内在学习平台的数据统计分析方面也有所突破,例如立思辰公司推出了有效的数据分析和可视化展示相结合的学习平台,学生可以在平台上看到自己考试相关的详细数据统计和分析,了解知识点得分率和进步情况。国内外先进的大数据学习分析研究对目前学习平台的统计分析功能有很大的启示和帮助。

本文通过研究了解到Canvas和Sakai平台对于数据统计方面有自己的特点,两个平台在数据统计功能上各有优势和不足。Canvas在数据展示呈现方面更加简洁直接,它将学习过程分类进行数据统计,和学生的成绩紧密结合在一起,可以让教师更容易对学生的学习过程进行了解和评价。但Canvas对于数据的分类统计并不够全面,不能够适用于每一位教学者,只能说为大部分的教师和管理员提供了便利。Sakai的优势在于它的数据统计非常全面,并且可以定制,可以满足任何一位教师和管理员的需要。但它的缺陷也很明显,过分注重数据的全面而忽略了数据统计功能的实用性。Sakai的界面设计不够人性化,并且杂乱无章的数据内容很大一部分是用不到的,会造成空间浪费。用户体验并不好,数据之间也没有什么联系。两款平台有一个共同的不足在于数据统计后都没有得到有效的利用,平台不能将数据进行计算得出有效的结论提供给管理员和教师。管理员和教师只能凭数据了解学生的学习行为,按照经验对学生进行指导,而不能通过平台对数据的分析对学生的学习行为进行准确的定位。

从国内外目前的研究应用情况以及对两平台的对比分析来看,一个优秀的学习平台应该同时注重数据的统计和分析两部分功能的平衡发展。一方面能够将数据统计的更加全面,在展示中要有条理,从用户的视角出发来设计展示方式。另一方面应该加大分析力度,对统计的数据进行深度分析,统计数据经过计算,进行系统分析,对学习过程行为进行分析,预测,对教师和学生进行指导。

根据以上分析以及目前主流学习平台的现状得出,目前学习平台在大数据环境下的数据分析方面还有很大欠缺。根据学习平台的数据分析需求对比,目前学习平台的不足表现为以下几点:

1. 缺乏对学生的即时反馈。目前的主流学习平台数据分析功能主要为教师服务,忽略了对学

生自主学习的影响。

2. 缺乏对数据进行深度挖掘分析。目前学习平台纷繁复杂, 对学生的数据抓取存储也没有统一标准, 因此对数据处理学习分析的过程造成了一定的阻碍。大多数平台只能利用数据进行简单的分析, 没有经过科学的研究得出有效的分析方法。这样的分析结果对于教学的作用非常有限。

3. 缺乏对平台操作性的设计。目前的学习平台主要功能基本一致, 由于多数学习平台为学校使用, 而非商业化, 因此在系统的界面设计, 功能展示设计方面都有所欠缺。很多平台的界面传统, 没有新意, 让学生和教师提不起兴趣。另一方面, 一部分平台只注重功能的多样性, 忽略了平台的操作性, 导致软件非常难用, 甚至对教学都产生了影响。

可以借鉴国外的研究, 吸取可以利用的先进理论和成熟的研究方法, 结合本国国情以及学校学生的实际情况, 在研究中进一步改进, 得出新的发现, 创造出新的结果并对目前的学习平台进行改进。

针对目前学习平台的统计分析功能的不足, 提出以下建议:

1. 增加学生反馈的功能, 让学生能够及时了解自身的学习情况, 进行调整。

2. 增加对教师和学习者对于教学和学习过程的需求进行深入了解, 进一步将平台统计分析功能的需求细化, 让教师和学生能够得到有利于教学和学习的数据, 改善教学过程, 提高学习效率。

3. 更加深层次的研究数据挖掘在教育领域的应用, 提出更多科学的分析方法, 使得学习分析的结果更加全面、深入, 在教学过程中发挥更积极的作用。用可量化的数据对教学过程和学习过程进行引导, 最大程度上发挥数据的价值。

4. 请专业的设计人员对平台的页面布局、功能展示进行设计, 让学习者和教师更加便捷的试用学习平台。

4 结语

由于客观条件, 研究视野和实际研究水平有限, 在本研究中还存在很多不足之处。由于资源有限, 只针对 Canvas 和 Sakai 系统进行研究, 研究范围窄。目前存在的主流教学管理系统还有很多, 虽然都有共性, 可以将更多的教学管理系统纳入研究范围使研究更宽泛, 使对比更加强烈, 共性和异性更加鲜明; 主要对平台的基本功能进行对比, 可以增加更多相应实际应用的研究介绍来丰富本研究。对目前的大数据技术了解不够深入, 不能对大数据技术应用在教学系统中提出更明确的建议, 可以进一步对数据挖掘, 机器学习技术进行一定了解和研究, 对比大数据分析在商业化的产品中的应用, 提出相应的在教学系统的数据分析中使用的具体建议。目前国内对于利用学习平台的数据进行大数据分析的研究还处于初级阶段, 大部分的研究只是针对数据进行算法研究, 没有考虑到实际应用如何跟学习平台进行结合, 侧重点没有关注到教学平台的数据统计分析功能应该如何发展, 将什么样的数据怎样可视化呈现在用户眼前。但随着大数据的技术发展, 以及目前在教育中对数据分析的不断探索, 学习系统必将越来越智能化, 平台的统计分析功能将会越来越丰富, 对教学的作用也会越来越明显。

参考文献:

- [1] 顾小清. 教师教育评价系统提供诊断服务 [J]. 中国教育网络, 2011(1):33-35.
- [2] 葛道凯, 张少刚, 魏顺平. 教育数据挖掘: 方法与应用 [M]. 北京: 教育科学出版社, 2012:29-30.
- [3] New Media Consortium.Horizon Reports[EB/OL].[2016-05-15].<http://www.nmc.org/nmc-horizon/>.
- [4] 魏忠. 教育正悄悄发生一场革命 [M]. 上海: 华东师范大学出版社, 2014.

[5] 郑燕林, 李卢一. 对大数据支持的学习分析与评价的需求调查—基于教师的视角[J]. 现代远程教育, 2015(2):36-42.

[6] 顾小清, 张进良, 蔡慧英. 学习分析: 正在浮现中的数据技术[J]. 远程教育杂志, 2012, 30(1):18-25.

[7] 魏顺平, 韩艳辉, 王丽娜. 基于学习过程数据挖掘与分析的在线教学反思研究[J]. 现代教育技术, 2015, 25(6):89-95.

[8] 黄德群. 云服务架构下的 Canvas 开源学习管理系统研究[J]. 中国远程教育: 综合版, 2013(7):64-70.

[9] 王蓉. 自由、开放的 Sakai 平台[J]. China Education Network, 2007(6):63-64.

[10] Lauria E, Moody E, Jayaprakash S, et al. Open academic analytics initiative: initial research findings[C]// International Conference on Learning Analytics and Knowledge. 2013:150-154.

[11] Jayaprakash S, Moody E, Laur í a E, et al. Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative[J]. Journal of Learning Analytics, 2014, 1(1):6-47.