

Moodle 平台师生访问行为日志统计与挖掘研究

郭涵阳,高曼如,沈良忠
(温州商学院,浙江 温州 325000)

摘要:信息技术的快速发展催生了网络教学平台的不断涌现,其中 Moodle 平台因其开源特征被认为是当前构建 MOOC (大规模开放在线课程)的有力平台之一。网络课程教学的有效开展使得 Moodle 系统日志表中存储了大量的师生访问行为数据,有效的日志统计挖掘有助于发现大量数据背后潜在的访问规律。通过统计分析、数据挖掘等技术实现了对日志数据的深入研究,并针对分析结果提出了 Moodle 课程使用中存在的问题及相应的改进意见,不仅有助于理解师生在 Moodle 平台的课程访问学习规律,而且能为 Moodle 课程的教学评价提供有价值的参考建议,有助于后期 Moodle 课程自动评价系统的研究。

关键词:Moodle;统计分析;数据挖掘;聚类

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2016)11-0168-04

doi:10.3969/j.issn.1673-629X.2016.11.037

Research on Statistics and Mining of Log Data about Visiting Behavior for Both Teachers and Students from Moodle

GUO Han-yang, GAO Man-ru, SHEN Liang-zhong
(Wenzhou Business College, Wenzhou 325000, China)

Abstract: The rapid development of information technology has led to the emergence of the online teaching platforms, among which Moodle system is considered to be one of the powerful platform to build MOOC (Massive Open Online Course) because of its characteristic of open-source. The effective development of online course teaching makes the Moodle system accumulate a lot of visit behavior data of both teachers and students. An effective analysis on the log data would surely help discover the underlying pattern. It makes a deep study on these log data by means of statistical analysis and data mining technologies. Based on the result, the existing problems and corresponding improving suggestions are presented which not only help to understand behaviors of both teachers and students, but provide valuable suggestions for Moodle course's teaching evaluation which is also helpful for the later research on the automatic evaluation system of Moodle course.

Key words: Moodle; statistical analysis; data mining; clustering

0 引言

随着教育信息化理念的深入人心,教育技术与课堂教学的结合日益紧密,越来越多的高校在积极构建基于网络的课程教学平台。目前,国内外高校使用的课程管理系统主要包括 WebCT、BlackBoard、Top-Class、Claroline、Moodle 等等^[1]。其中,基于社会建构主义教学理念而开发的 Moodle (Modular object-oriented dynamic learning environment, 模块化面向对象的动态学习环境) 由于其开源性在国内外被广泛使用^[2]。Moodle 课程开发与设计的最大特点就是模块化,教师可以根据课程的需要自行设定包括以下十多

种教学活动模块,如作业、测验、讨论区、博客、聊天室、程序教学、在线调查等,从而实现教师个性化的教与学生个性化的学^[3]。

该校自 2011 年起,完成建设 Moodle 课程总计 90 多门,每学期使用近 40 门,平均涉及学生数 4 000 人左右。Moodle 平台累计了几百万的师生访问日志记录,有效利用统计分析、数据挖掘等技术实现日志分析,不仅能发现师生的访问规律,同时能为 Moodle 课程评价等提供有价值的建议。在国外,Romero 等^[1]在 2007 年采用数据挖掘的方法对 Moodle 平台的各种数据进行了统计分析和挖掘等研究。Jovanovic^[4]根据课

收稿日期:2015-10-28

修回日期:2016-02-24

网络出版时间:2016-10-24

基金项目:浙江省 2015 年度高等教育教学改革项目(jg2015249);浙江省 2015 年度省教育技术研究规划课题(JA021)

作者简介:郭涵阳(1994-),男,研究方向为计算机应用技术;沈良忠,副教授,通讯作者,研究方向为教育数据挖掘、计算机应用技术。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20161024.1105.010.html>

程中学生作业完成数,测验试题正确数、错误数等信息建立分类模型预测学生表现。Lavigne 等^[5]根据学生点击量、登录次数、会话时间等信息实现学生聚类,并根据两组学生的访问路径差异推理不同的网络学习策略。而在国内,魏顺平^[6]在 2011 年仅以一门培训课程为例,进行了相关的数据分析。金杰^[7]也以一门课程为例,主要做学生参与度与平时成绩、作业完成次数与成绩、教师参与度与学生参与度之间的相关性研究。张国荣^[8]通过模糊 C 均值方法对一门课程的学习数据进行模糊聚类研究,使得班级成员在划分小组进行个性化学习时,更好地理解每个学生所属的类别。文中通过对学校 2013 年、2014 年 Moodle 平台使用课程运行日志数据的比较分析,发现了隐藏在数据背后的师生访问规律及课程建设中存在的一些问题,帮助教学管理部门加深对 Moodle 课程使用情况的宏观认识,并有针对性地提出了教师改进课程建设的若干意见,为 Moodle 课程教学评价标准的制定提供了强有力的数据支撑。

1 Moodle 日志统计分析

1.1 日志表预处理

Moodle 日志记录了教师、学生与 Moodle 课程各教学模块之间交互的所有数据,甚至包括师生的每次登录,显然系统日志所积累的大量数据对于师生访问偏好、课程建设问题、课程教学评价等方面都提供了非常有价值的信息。通过对某门课程师生交互日志信息的详细分析,不仅可以真实再现师生教与学的多维立体过程,而且可以实现针对师生教与学的过程性考核和评价。Moodle 日志表中每条操作记录的具体信息如表 1 所示。

表 1 日志表字段说明

字段名称	说明
Id	日志流水号
Userid	用户编号
Ip	访问地址
Course	课程编号
Module	访问模块
Action	访问行为
Url	访问链接
Logtime	日志时间

Moodle 日志表在进行分析的时候,其用户信息需要参考用户表(md1_user),以便判断该用户是教师还是学生;课程信息参考课程表(md1_course),以便获取课程性质。日志数据选取 2013 年 2 月至 2013 年 6 月以及 2014 年 2 月至 2014 年 6 月这两个时间段作为研

究对象。其中,2013 年 2 月至 2013 年 6 月的日志记录共 84 万;2014 年 2 月至 2014 年 6 月的日志记录共 88 万。为了实现教师、学生的分组查询,日志表中增加 isteacher 字段以便区分用户身份;为了实现日志信息的时间访问,日志表中增加了 logyear、logmonth、logday、loghour,每条日志信息最终归属到某小时。

1.2 模块访问统计

要实现对 Moodle 课程教学成效的评价,必须获得师生经常使用的课程教学模块。2013 年和 2014 年 Moodle 课程各教学模块的使用统计情况如图 1 所示。



图 1 Moodle 课程各教学模块统计

由图可知,测验、课程、讨论区、资源、作业等模块是师生最经常访问的,将被作为评价的主要模块;而其他模块如用户、上传、博客、投票、wiki 等使用较少。比较发现,讨论区的访问量较高,表明师生能积极利用该模块进行互动交流或者知识库建立;资源模块的访问量基本与讨论区持平,但是资源模块的性质暗示了课程在教学材料建设不高,需要更好地完善以便更多地吸引学生进行自主学习。测验和作业两大模块需要循序渐进加以完善。测验更多的是以客观题的方式出现,系统可以自动判分;而作业更多的是以文档提交的方式上传,作业相对较低的访问量及上传量说明课程作业少、上交率低的情况在课程中普遍存在,其原因可能是教师对作业模块的批改和反馈不够及时。

1.3 行为访问统计

模块统计分析结果表明测验、课程、讨论区、资源、作业等模块最为常用,每个教学模块又具有不同的操作行为。通过对各个教学模块不同操作行为的统计发现,对于课程、资源模块,98% 以上的操作行为是 View 或 View All。在作业模块中,View 占比 80% 左右,Upload 占比 13% 左右,Update Grades 占比仅为 3% 左右,说明教师在作业的批改部分确实存在较大问题,教师作业批改的学生覆盖率和反馈及时性均不到位,使得学生日后作业的上交率也随着教学进程的逐步推进呈现降低的趋势。在测验模块中,行为 Continue Attempt 平均占比达到 30% 左右,说明大部分课程都开放了“多次测验,最高取分”的策略,学生也是积极参与到测验模块进行自我学习成效的评价,这也很好地说明了图 1 中测验模块的统计占比高的情形。与其他教学模块不同的是,讨论区教师隶属的特有行为占比比其

他教学模块都高,例如“新帖”行为占比大约 10% 左右,充分说明大部分教师都有在讨论区积极的组织主题讨论活动。

2 Moodle 课程时间访问

课程的学习具有一个时间连续性,Moodle 课程的学习亦是如此。为了深入理解师生访问行为在时间上的特点,分别统计了按月、星期及小时的访问情况。

2.1 按月访问统计

通常,每年 2 月份为高校第二学期的开始,Moodle 平台数据访问量较低,随着授课进度的前移,访问数据量显著上升,随后访问数据可能增加或降低,主要取决于该学期使用 Moodle 课程的学生人数、课表安排等其他因素。临近考试的月份,Moodle 访问数据又会明显上升,因为教师通常将 Moodle 课程学习作为一项成绩记录期末总分,较大部分学生会会在考前做一些突击性工作。

2.2 按星期访问统计

同一门课程,由于师生的随机性,每学期的授课时间安排可能不同,例如 A 课程 2013 年安排在周二,2014 年可能安排在周三。相应的 Moodle 平台该课程的访问时间也肯定存在差异,为了研究每周可能存在的课程访问规律,2013 年和 2014 年每周的访问量统计通常呈现“M”型,即两头访问量较低,中间又存在两个峰值和一个峰谷。两头访问量低的情况可能是由于周末的原因,师生使用 Moodle 平台的积极性不高;另外周三下午学生通常会参与一些社团类的活动,教师参与教研室的一些教学活动或者会议,使用量也会相对较低。

2.3 按小时访问统计

为了更好地掌握教师、学生每天的 Moodle 课程访问规律,2013 年和 2014 年每天每小时的访问统计结果如图 2 所示。

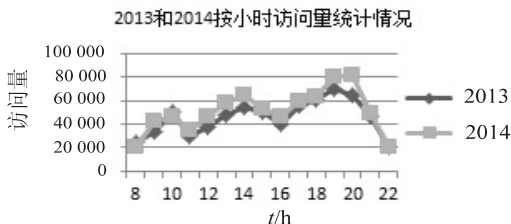


图 2 按小时访问统计

由图可知,每天的访问峰值基本出现在以下三个时间点。第一是早上 10-11 点,有些学生刚上完上午 2 节课,需要完成 Moodle 平台的相关作业等;第二是下午 4 点左右,大部分学生刚上完下午 2 节课;第三是晚上 9 点左右。通常,学生主要选择下午和晚上进行 Moodle 课程学习,另外也有小部分比例数据表明即使

在凌晨也有用户在使用网络平台,但是由于在校学生使用网络的时间通常在晚上 11 点之前,这说明相当一部分教师深夜依然利用 Moodle 平台进行资料上传、作业批改、测验查看等课程准备类工作;当然也不排除有少部分同学在校外利用 Moodle 平台进行学习的可能。

3 Moodle 日志挖掘研究

数据挖掘作为一种数据利用的有效手段,已经在许多领域得到了广泛应用,例如金融业、电子商务等^[9]。数据挖掘的过程主要包括数据收集、数据预处理、数据挖掘和结果评价四个必要环节^[10]。当前,教育信息化的发展使得教育领域的各类数据得以迅速增长,因此如何从海量的教育数据中挖掘出对教育者、学习者和管理者有用的信息,促使了教育数据挖掘研究的出现^[11]。Moodle 平台作为一个网络学习平台,虽然不能像传统课堂中通过面对面交流得到情真意切的及时反馈,却能够完整记录学生的大多数学习行为,通过对系统中师生交互的日志数据的挖掘,还能有效地发现数据背后所隐藏的师生教与学的规律,同时对 Moodle 平台数据的挖掘同样需要经过以上四个必要的环节。

3.1 师生讨论区回归分析

Moodle 平台基于社会建构主义学习理念而开发,社会建构主义认为教师、学生应该是平等的主体,在教学活动中他们相互协作,相互交流,并逐步根据自己已有的经验共同建构知识^[12]。因此 Moodle 设计目标不仅仅停留在教学内容的管理与呈现,而是更加关注对教学过程中各种“活动”的支持。Moodle 对教师角色的定位不是知识的提供者,而是一个学习社团中有影响力的榜样,是各种学习活动的引导者,是与学生平等对话、共同实现预定学习目标的组织者^[13]。基于此,通过对 Moodle 课程各教学模块的线性回归分析,发现与其他模块相比,讨论区的建设最能凸显教师作为组织者(引导者)与学生之间的关系,教师与学生在讨论区操作帖子数量之间的相关系数达到 0.84,属于高度相关,如图 3 所示。通过对若干门课程比较发现,教师在论坛操作量大的课程,其学生的平均操作量也较大,反之亦然。通过对论坛数据的观察,可及时提醒教师进行论坛的师生互动交流。

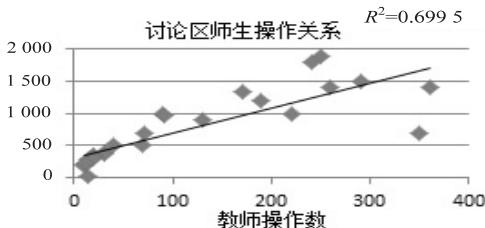


图 3 讨论区师生操作线性关系

3.2 学生聚类分析

数据聚类是一个将物理或者抽象对象的集合分组成由类似对象组成的多个类或簇的过程^[14]。通过学生聚类可以将利用 Moodle 平台进行学习的学生根据其操作行为实现相似群体的划分,从而实现针对不同群体的针对性引导,例如针对作业提交少的群体及时提醒其进行作业提交。根据图 1 的模块访问统计结果,测验、课程、讨论区、作业、资源在模块访问上排名前五。因此,主要通过学生在以上五个模块的访问数据对学生进行聚类,发现相似学生的访问共性。选取 1 518 名在各个模块上均有操作的记录,使用 SPSS 的 K - means 进行聚类,设置聚类个数为 3,聚类结果和案例分布如表 2 所示。

表 2 K - means 聚类结果

聚类	案例数	讨论区	资源	测验	作业	课程
1	113	361	59	205	37	151
2	344	70	20	304	23	55
3	1 061	63	21	68	27	44

为了更好地表明聚类结果之间的差异,依据学生操作行为聚类结果绘制的雷达图如图 4 所示。

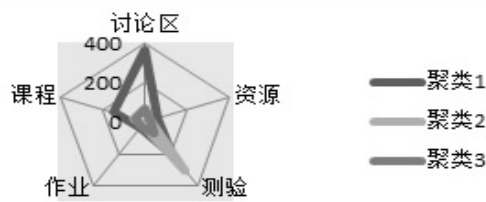


图 4 聚类结果雷达图显示

由图可知,从第 1 类到第 3 类,学生的活跃程度逐渐降低,并主要表现在讨论区、测验两个模块。从第 1 类到第 3 类,作业、资源、课程模块的访问量都很低,这值得教师反思。课程、资源的较低访问量充分说明课程教学材料以及其他相关的自主学习材料存在数量较少,类型单一,不能吸引学生的注意或者学生认为从中难以获得自己所需内容的问题。作业资源的低访问量亦表明学生在作业提交的环节缺乏监管,教师没有及时对网络作业进行批阅,对于学生作业交或不交的态度是模糊的,这种态度直接影响了学生再次递交作业的积极性。教管部门、教师应该及时针对这些表现不活跃的学生,开展针对性教学干预。

4 结束语

Moodle 平台存储了大量的日志数据,对以上日志数据的统计分析和挖掘研究是理解 Moodle 平台运行状况的有效途径。文中通过对 2013 年和 2014 年两个学期 Moodle 课程日志数据的统计分析和挖掘研究,掌

握了 Moodle 平台的总体访问情况,数据背后隐藏的潜在师生访问偏好,并依据学生平台访问数据对学生进行聚类和分类分析,有助于改进课程教学和学生干预工作。统计分析结果给教管部门、一线教师以启迪和深思。例如,前五的模块访问明确预示着 Moodle 课程教学评价的主要方面;课程、资源的较低访问预示着课程负责教师必须在 Moodle 课程的教学材料等方面进一步完善;讨论区师生操作的高度相关性暗示有些教学模块需要师生共同的经营等等。

课题的下一阶段工作就是根据日志分析结果,制定 Moodle 课程的评价标准,实现课程评价结果的自动生成,最终实现师生成长档案袋的设计与开发工作。

参考文献:

[1] Romero C, Ventura S, García E. Data mining in course management systems: Moodle case study and tutorial[J]. Computers and Education, 2008, 51(1): 368-384.

[2] 熊小梅. 基于 Moodle 平台的网络课程设计及应用[J]. 教育与职业, 2014(2): 172-173.

[3] 沈良忠, 李莹, 万晓洁, 等. 基于 Moodle 的实习管理模式探究[J]. 中国电力教育, 2013(2): 145-146.

[4] Jovanovic M, Vukicevic M, Milovanovic M, et al. Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study[J]. International Journal of Computational Intelligence Systems, 2012, 5(3): 597-610.

[5] Lavigne G, Ruiz G G, McAnally-Salas L, et al. Log analysis in a virtual learning environment for engineering students[J]. Revista De Universidad Y Sociedad Del Conocimiento, 2015, 12(3): 113-128.

[6] 魏顺平. Moodle 平台数据挖掘研究——以一门在线培训课程学习过程分析为例[J]. 中国远程教育, 2011(1): 24-30.

[7] 金杰. 基于平湖电大课程平台(Moodle)的数据挖掘研究[J]. 电脑编程技巧与维护, 2014(22): 57-58.

[8] 张国荣. Moodle 平台数据挖掘方法设计与实现[J]. 计算机技术与发展, 2014, 24(5): 231-234.

[9] 毕晓东. 私有云模式下基于 Moodle 的课程协作学习平台设计[J]. 软件导刊, 2014, 13(1): 85-86.

[10] 卢永艳. 数据挖掘在网络教育中的应用[J]. 现代计算机, 2007(11): 56-57.

[11] 李婷, 傅钢善. 国内外教育数据挖掘研究现状及趋势分析[J]. 现代教育技术, 2010, 20(10): 21-25.

[12] 黄旌. Moodle 在“结构与计”教学中的应用探究[J]. 教育研究与评论: 技术教育, 2012(1): 80-81.

[13] 李春晖. 基于 Moodle 的课外自主学习研究[J]. 时代报告: 学术版, 2012(9): 349.

[14] Han Jiawei, Kamber M. 数据挖掘: 概念与技术[M]. 北京: 机械工业出版社, 2001.