

MOOC 学习者课程学习行为分析

——以“电路原理”课程为例

李曼丽¹ 徐舜平¹ 孙梦嫻²

(1. 清华大学 教育研究部, 北京 100084; 2. 清华大学 计算机系, 北京 100084)

[摘要] 以往研究发现,大量 MOOC 学习者中途退出,课程完成率非常低,这个现象一直困扰着 MOOC 授课教师和 MOOC 平台所有者,低课程完成率不是期望的目标,也不利于 MOOC 的持续发展,许多授课教师和研究者试图改变这种状况,这个现象也因此成为学者们关注的焦点。为了分析在此背景下哪些因素对学习者的课程参与和课程完成产生影响,并基于分析结果给出建议,本研究以“学堂在线”(xuetangX)平台的“电路原理一”课程数据为基础,使用 Tobit 和 Logit 两个定量分析模型,分别对 MOOC 学习者的课程参与和完成情况进行深入分析。研究发现:学习者的课程学习动机越强,学习者在课程讨论区和 Wiki 表现越活跃,即课程参与度越高,同时更有可能学完课程;课程注册时间越早,学习者越容易完成课程学习,也会较多地参与课程讨论;相对于非高校学习者,高校的学习者在讨论区中表现更活跃;个体课程参与度对其他学习者具有明显的正向影响。这些发现对于 MOOC 的持续发展具有一定指导意义。

[关键词] MOOC; 学习行为; 退出行为; 课程参与

[中图分类号] G434

[文献标识码] A

[文章编号] 1007-2179(2015)02-0063-07

一、前言

2012 年,大规模公开在线课程,即 MOOC 逐渐在美国顶尖高校中兴起,名校在网络学习平台上提供大量免费精品课程,为更多学生自主学习高校课程提供了可能。此后,越来越多美国以外的大学通过与 Coursera、Udacity、edX 三大 MOOC 平台建立合作而加入到 MOOC 潮流中。2013 年 5 月,清华大学正式加盟 edX,成为 edX 首批亚洲高校成员之一。为了促进国内学习者更便于通过 MOOC 学习,清华大学还特别设计搭建了学堂在线(XuetangX)。

MOOC 在世界范围内备受关注的同时,在中国的发展也成为 MOOC 研究者关注的焦点。在 MOOC 出现前,中国已有远程网络教育,它是广域分配教育资源的重要途径。中国在网络课程建设实践的过程中,出现了诸如课程数量多但使用率低,重视

教学内容传授而轻视学习环境设计,教学内容呈现形式单一等问题(林君芬等,2001)。但是,之后出现的 MOOC,丰富了学习内容、教学方式和学习方式,为学习者提供了更多的机会(Yuan & Powell, 2013),这使得在线教育进入了新的阶段。在 MOOC 平台上,学习者可以学习由世界一批一流大学提供的高质量精品课程。

MOOC 被认为是一种有别于传统教学模式的新型知识传播模式,然而它的诸多问题也开始被人们关注,比如,主要是哪些类型的学习者在学习?学习者的个人属性(性别、所在地区、高校学生、已受教育年限等)对他们的 MOOC 学习会产生怎样的影响? MOOC 是一种完全自主的学习模式,大部分学习者中途退出及课程完成率低,如何解释这种普遍存在的情形?上述问题对 MOOC 的发展有非常重要的影响,值得研究者予以关注。

[收稿日期] 2015-02-04

[修回日期] 2015-03-06

[DOI 编码] 10.13966/j.cnki.kfjyyj.2015.02.007

[作者简介] 李曼丽,清华大学教育研究院教授,博士生导师,研究方向:课程与教学(marylee@tsinghua.edu.cn);徐舜平,清华大学教育研究院 2013 级硕士研究生,研究方向:在线教育(xsp13@mails.tsinghua.edu.cn);孙梦嫻,清华大学计算机系本科四年级学生,教育研究院 MOOC 研究课题组研究助理(summl_thu@163.com)

二、文献综述

虽然中国已有多所高校积极推进 MOOC,但总体来说,MOOC 在中国仍处于萌芽阶段。根据调查,学习者在年龄、地区和职业上有明显的集中性,多以城市的年轻人和大学生为主,且具有完整学习体验的 MOOC 深度学习者数量十分有限;在地区分布上,学习者主要集中在东部沿海经济发达和中西部经济比较发达、高等教育资源相对丰富和人口较多的省市,以市区为主(李曼丽等,2013)。

作为一种完全自主的学习模式,MOOC 学习更多的是自我驱动型的,即内在的学习动机推动学习者选修课程。国外对 MOOC 课程学习者的学习动机研究发现,主要的学习动机包括:对主题感兴趣(87%)、职业发展的需要(44%)、对授课方式感兴趣(Belanger & Thornton, 2013),能够学到知识和技能(55.4%),作为个人的挑战(Breslow et al., 2013)等。国内调查问卷也呈现相似结果,79.8%受访者表示对课程内容有强烈兴趣,75.76%受访者认为能够接触到高质量的课程或好的学校,45.96%受访者认为课程对工作和学习非常有用,兴趣成为 MOOC 学习的基本前提(刘杨等,2013)。

远程教育的高退出率是被频繁关注的重要现象之一,MOOC 也呈现出相同趋势。国内对远程教育的研究发现,引起远程学习者退出是多个因素相互作用的结果,这些因素可以分为两大类:

- 1)与学习者本身相关(学习者的教育、环境、心理背景等);
- 2)与教育环境相关(教师、学校、学习系统支持等)。

其中,退出者的性别、年龄、课程负担是影响远程教育学习者在学时间的重要因素。据调查,在总体上,男性学生退出率高于女性,退出人数随年龄增大而呈阶梯状增多,在某个年龄段达到峰值,再随年龄的增大,退出人数又呈阶梯状减少;几乎全部退出者在第一学期都选了较多课程,课程负担过重,极易退出学习(刘永权等,2012)。MOOC 由于缺乏传统教育模式下的督促措施和教学环境,也出现了类似特征。MOOC 高退出率的原因也很多,如难以实现学习的自我调控,淹没在海量信息和复杂网络中,难以适应协作交流学习方式等,这些压力都会导致学

习者无法继续学习(樊文强,2012)。发现 MOOC 学习者退出的主要原因,并提供针对性解决措施,是 MOOC 平台所有者和课程制作者面临的一项重要任务。

针对 MOOC 学习者的高退出率问题,MOOC 平台的讨论区可以发挥一定的作用。在 MOOC 讨论区中,世界各地的学生都可以从自己的角度就某一问题进行讨论(苏芑等,2013),有些授课教师和助教也会通过网络社交平台积极地参与到学习者的讨论中,帮助学习者更好地学习(李曼丽,2013)。

对于讨论区的效果,有学者研究了课程论坛的使用对课程通过率的影响。库切等(Coetzee et al., 2014)2013年3月对一门 MOOC 课程的调查结果显示,5985个注册学生中,仅有532人通过,其中使用课程论坛的1572个注册学生中,442人(占通过人数的83%)通过了课程学习,近30%,远高于不使用论坛的2%的学习通过率。

根据上述相关文献的回顾,在 MOOC 课程学习中,学习者选修某门课程主要是由其学习动机驱动的,这成为本文分析的基本出发点。而对普遍存在的学习者高退出率问题,虽然有学者对其原因进行了说明,但鲜有文献对其进行更为精准的量化阐释,这也成为了本文研究的重点。许多文献证实了讨论区有助于降低退出率,但这些文献仅仅是简单的描述统计,本文试图更加精准地分析讨论区对课程完成的影响。

三、研究问题和研究方法

(一)研究问题

本文主要研究两个问题:

- 1)哪些因素影响课程参与,以及影响的方向和程度如何?
- 2)哪些因素影响课程完成,以及影响的方向和程度如何?

本文以清华大学电机系于歆杰老师在学堂在线(xuetangX)上开设的“电路原理一”课程为例,对上述研究问题进行实证研究。

关于学习者的课程参与和课程完成情况,本文研究方向是分析哪些因素会影响课程的完成(用课程完成度衡量),以及如何影响。对于课程的参与情况,本文拟用课程参与度衡量,衡量指标是课程讨

论区的参与情况和 wiki 的词条编辑情况。本文对课程参与情况的分析,主要是为了通过提高学习者的参与度,间接地提高学习者的课程完成度。

课程参与度的影响因素(即自变量)包含在课程完成度的影响因素中,在课程参与度不存在外部性的情况下,它仅仅作为中间变量,不会对课程完成度带来新的内容。但是,由于网络产品的特殊性(即使用人数越多,价值越大),就 MOOC 课程而言,某一个体的课程讨论区和 wiki 的参与程度越高,对其他学习者的帮助越大,存在显著的正外部性,所以对于课程参与度的分析具有必要性。

(二) 研究意义

本文期望通过这两个问题的研究,发现 MOOC 课程学习者退出学习的影响因素,并且通过对影响因素的分析,为解决 MOOC 课程学习普遍存在的高退出率提供解决思路或方案,以更好地帮助学习者从 MOOC 中获得更多的价值。

(三) 研究方法

1) 对课程参与情况,本文以课程参与度(讨论区和 wiki 的参与情况)为因变量进行分析。对课程参与度的衡量,本文将学习者在讨论区中参与讨论、在 wiki 中编辑词条,作为课程参与意愿的度量指标,但是在两个模块中都没有参与的学习者占比较高,数据存在被左截取(left-censored)的情况,所以本文将使用 Tobit 模型进行估计。

2) 对于课程完成情况,本文以课程完成度(学习者参加了课程的期末考试,即表示学习者完成了课程学习)为因变量进行分析。课程完成度只有两种情况:完成、未完成,属二值因变量,所以本文用 Logit 模型进行估计。

四、数据和描述统计

本文分析数据是清华大学电机系于歆杰老师 2013 年 10 月 17 日在“学堂在线”上开设“电路原理一”课程过程中学习者学习所产生的。由于涉及对学习者的退出行为的分析,本文提取了 2013 年 12 月 23 日该课程期末考试结束时及结束前所生成的学习者的数据(因为期末考试结束后再加入课程的学习者,无法参加考试完成课程的学习,所以本文分析没有使用这部分数据),共 5027 个样本。

下面对几个主要分析变量说明如下:

1) 学习动机是对课程学习分析的基本影响因素,是出发点。但是,由于“学堂在线”所记录的数据没有关于学习动机的度量,所以本文用数据中最能体现学习者动机的学习时长的对数(\ln 学习时长)作为学习动机的代理变量。

2) 高校集中度表征的是高等院校的相对集中程度,用以分析学习者是否因高校资源比较集中,而不太倾向学习 MOOC 课程,本文的高校相对集中地区包括:京、沪、苏、鄂、陕、粤、辽、鲁、津。

3) 课程完成度以学习者是否参加课程期末考试为判断标准,如果参加了考试,则表示学习者完成了课程学习。

4) 课程注册时滞表示学习者的课程注册时间和课程开放注册时间(2013 年 10 月 9 日)的差值,单位为天,差值越大,表示注册越晚,该变量用以分析学习者学习课程时间的早晚,是否会对学习者的学习产生影响。

5) 课程参与度则用学习者在讨论区的发帖数和在 wiki 中编辑词条的次数之和度量。本文分析所用的其他变量见表一和表二。

表一 学习者的课程学习描述统计

变量名称	变量定义	观测值	均值	标准差	最小值	最大值
课程完成度	二值因变量:0=未完成课程学习,1=完成课程学习	3010	0.036	0.187	0	1
课程参与度	讨论区发帖数和 wiki 词条编辑次数之和	3010	0.531	6.763	0	339
性别	虚拟变量,0=男性,1=女性	3010	0.087	0.282	0	1
受教育年限 ^①	受教育年限(年)	3010	15.491	2.613	6	23
高校集中度 ^②	对高等院校相对集中地区的度量,虚拟变量,0=相对集中,1=相对不集中	3010	0.488	0.500	0	1
学习者类型	虚拟变量,0=非高校学习者,1=高校学习者	3010	0.132	0.339	0	1
课程注册时滞	课程注册时间与开放注册时间之差(天)	3010	26.660	17.668	0	63
学习时长 ^③	作为学习动机的代理变量,度量学习动机	3010	229.270	693.454	1	14573

注:①受教育程度为小学、初中、高中、大专、学士、硕士和博士,我们将其折算为受教育年限(年),依次为 6、9、12、14、16、19、23。

②高等院校相对集中地区:京、沪、苏、鄂、陕、粤、辽、鲁、津;其他地区为高等院校相对不集中地区。

③学习时长表示学习者实际学习电路原理课程的时间(分钟)。

表二 学习者地区分布

变量名称	变量定义	观测值	东部地区占比	中部地区占比	西部地区占比	其他地区占比
地区	东、中、西部地区和其他地区	3010	0.598	0.224	0.156	0.022

注:东部地区包括京、津、冀、鲁、辽、沪、苏、浙、闽、粤、琼;中部地区包括晋、豫、皖、赣、鄂、湘、吉、黑;其余省份构成西部地区;其他地区包括港澳台和国外。

表一和表二是“电路原理一”课程学习以及学习者分布的描述统计。女性占 8.7%, 该门课程的学习者主要为男性。学习者平均受教育年限 15.5 年, 图 1 显示了学习者学历分布情况, 表明该课程的学习者主要为本科及以上学历群体(本科学历占 51%), 符合课程主要面向本科及以上学历学习者的特点。高校学习者占 13.2%, 表明该课程的学习者主要不是来自于高等院校, 而是校外群体居多。东部地区的学习者占 59.8%, 表明学习者主要来自于经济发达地区, 而经济落后的西部地区学习者比例不到 16%。学习者平均课程注册时滞 26.7 天, 从图 2 也可以看出, 课程注册时间较分散, 开课后的第一周有

一个小高峰, 课程首次上线(2013 年 10 月 17 日)的时滞为 8 天, 第一次练习的截止时间(2013 年 10 月 24 日)的时滞为 15 天, 表明许多学习者是在开课一段时间后加入课程学习的。图 3 是学习者年龄分布直方图(由于数据填写不规范, 该图去掉了前 2% 和后 1% 的年龄数据), 从图中可以发现学习者主要集中在 19-25 岁年龄段内, 青年人是该课程的主要学习人群。

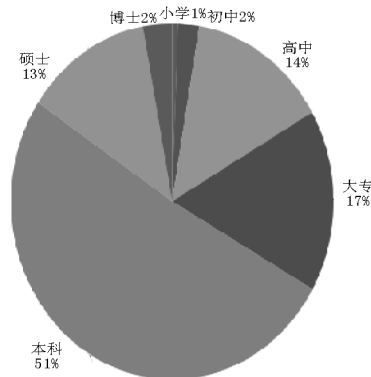


图 1 学习者学历分布

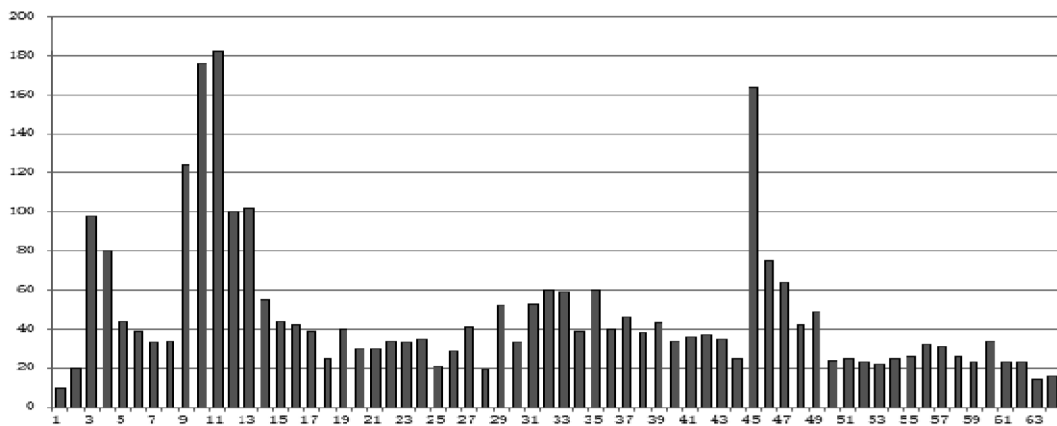


图 2 课程注册时滞分布

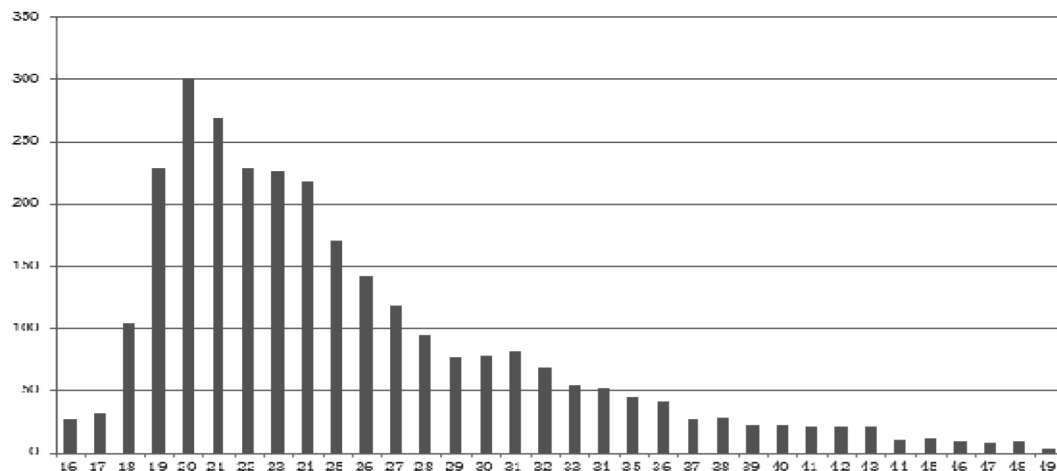


图 3 学习者的年龄分布

五、计量结果及分析

表三中的模型 1 是课程参与度的 Tobit 回归结果,其中 $\ln(\text{学习时长})$ 显著为正,表明该变量所代表的学习者的学习动机越强,学习者会越积极地参与到课程互动中(即在讨论区中与其他学习者交流讨论或在 wiki 词条中做贡献),表明学习动机是课程参与的重要影响因素;高校学习者显著为正,表明高校学习者比非高校学习者课程学习的活跃程度更高;课程注册时滞显著为负,体现了学习者的学习时间越短,参与程度越低,可能是由于时间少而降低了学习者的参与意愿,也可能是较早开始学习的学习者在讨论区中的讨论已经解决了该学习者的许多问题,或者 wiki 词条已被其他学习者编辑较多而该学习者很难再去完善词条。其中,女性学习者在 10% 的显著性水平上为正,体现了女性学习者课程参与程度相对男性学习者更高些。其他的系数都不显著。

表三中的模型 2 是课程完成度的 Logit 回归结果,其中课程参与度显著为正,表明学习者的课程参与越多,越倾向于完成课程的学习,也体现了学习者对于学习付出的珍惜; \ln 显著为正,且系数值较大,表明学习动机仍然是影响课程完成情况的最重要影响因素,MOOC 课程的学习主要是学习动机驱动型的;课程注册时滞显著为负,表明学习者的学习时间越短,越难将课程学完,主要原因可能是,由于离课程期末考试时间短,一方面是学习者认为自己学习的内容少,没有自信能够通过期末考试,另一方面是平时成绩少,即便参加了期末考试,也无法保证总分及格,这些都让学习者没有动力参加课程期末考试。其他系数都不显著。

六、结论与讨论

学术界对 MOOC 学习者的学习退出已经开展了一些研究,试图找出学习者退出的原因以及如何减少学习者中途退出的比例。本文结合“电路原理一”课程,得出以下结论。

1) 首先需要关注的是学习者的学习动机。学习动机是课程完成的最重要影响因素,可以从知识、技能、主题甚至从学习方法上激发学习者的兴趣。我们通过学习时长体现学习动机,因为长期和深入

表三 课程参与度和课程完成度的模型回归

	模型 1	模型 2
	Tobit	Logit
	课程参与度	课程完成度
课程参与度		0.077 * * * (0.022)
$\ln(\text{学习时长})$	8.672 * * * (0.659)	0.852 * * * (0.079)
受教育年限	-0.382(0.394)	0.007(0.047)
高校集中度: 相对不集中	-0.661(2.333)	-0.119(0.273)
学习者类型: 高校学习者	6.513 * * (2.774)	0.156(0.342)
课程注册时滞	-0.675 * * * (0.075)	-0.036 * * * (0.008)
性别:女性	6.392 * (3.415)	0.535(0.379)
地区:		
东部地区	2.401(2.828)	0.067(0.330)
西部地区	-0.002(3.565)	0.185(0.410)
其他地区	0.695(6.627)	0.322(0.658)
常数项	-60.308 * * * (7.742)	-7.243 * * * (0.960)
观测值数	3010	3010
卡方	471.51	337.74
Pseudo R2	0.1398	0.3603

注:1) *、**、*** 分别表示在 10%、5%、1% 水平上显著,括号内为标准差,表中报告的是估计的边际效应;2) 高校相对集中度、非高校学习者、男性、中部地区为对照组。

的 MOOC 学习,即较长的学习时长,是由包含兴趣在内的学习动机驱动的(刘杨,2013)。同时我们发现,学习者较早地加入到课程学习中,有助于他们完成课程学习,MOOC 平台所有者和课程制作者可以采取某些方法,例如通过平台推送提醒信息,吸引潜在的学习者更早地加入到课程学习中,让其有更加充足的时间学习课程。

2) 其次需要关注课程参与度的正向影响。MOOC 强调学习的主动建构性、社会互动性和情境性,重视学习共同体与合作学习(MOOC 中的交流论坛、同伴评价)(李曼丽,2013)。MOOC 的这一教学设计原理,说明了课程参与度在 MOOC 学习中的重要性。我们的研究也验证了课程参与度与学习完成率之间有着极密切的关系,有不可忽视的正向影响。这给授课老师的启示是,在教学方法上进行升级创新,积极鼓励学习者参与到课程中去;给 MOOC 平台所有者的启示是,技术平台要能够更好地辅助教师的教学,更好地支持学习者的课程参与或课程互动。

本文还想就本研究 and 基于 MOOC 平台所产生的海量数据进行如下讨论:

1) 本文对于课程参与度的模型分析,关注MOOC作为一种网络产品所带有的社交性,与传统的属性有着明显的不同。网络产品的一个重要属性是,产品使用人数越多,该产品对个人的价值越大。但是,本文对于课程参与度的分析仅局限于某一学习者在课程参与中的付出对自己课程完成情况的影响,而对其他学习者的课程参与情况的影响未能提供更多的信息。尽管如此,就本研究的结论而言,仍然可以支持MOOC课程互动越强,讨论区或者其他社交性的网络板块越活跃,会给该课程所有学习者带来显著的正影响等结论。

2) 在教育 and 互联网相结合的时代背景下,数据越来越重要,价值越来越大,利用课程产生的学习行为数据改善学习者的学习表现和结果,将是未来发展的必然趋势。MOOC平台基本具备学习行为数据的记录功能,平台运行过程中产生的海量数据应该成为MOOC相关机构的重要信息资产(魏顺平,2014)。由于“学堂在线”数据记录的局限,本文的实证分析未能充分考虑学习者更多的背景信息和学习行为,比如学习者的专业背景、所处行业、是否在其他地方修习过相关课程、为何学习本课程、课程难度、课程价值等,这些信息对学习者的课程学习,以及中途退出都会产生影响。因此,更加完备的数据记录和分析,更有利于推动MOOC向前发展。平台系统利用数据挖掘技术对学习者的数据进行更为深入的挖掘分析,通过预测模型自动提出相应的课程内容和教学环节顺序的调整方案,教师和研究者可以基于数据挖掘结果和学习分析,对学生干预(张羽等,2013),是技术应用更为深入的方向。这些都有赖于计算机技术的进一步发展,通过计算机技术的发展,把教育推向更高的层级。

[参考文献]

- [1] Belanger, Y., & Thornton, J. (2013). Bioelectricity: A quantitative approach Duke University's First MOOC[EB/OL]. http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/6216/Duke_Bioelectricity_MOOC_Fall2012.pdf.
- [2] Breslow, L. B., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment*, 8:13-25.
- [3] Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? [A]. In *Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing*[C]. 1176-1187. ACM.
- [4] 樊文强(2012). 基于关联主义的大规模网络开放课程(MOOC)及其学习支持[J]. *远程教育杂志*, (3):31-36.
- [5] 李曼丽(2013). MOOCs的特征及其教学设计原理探析[J]. *清华大学教育研究*, (4):13-21.
- [6] 李曼丽,张羽,叶赋桂(2013). 解码MOOC——大规模在线开放课程的教育学考察[M]. 清华大学出版社:163-164.
- [7] 林君芬,余胜泉(2001). 关于我国网络课程现状与问题的思考[J]. *现代教育技术*, (1):55-77.
- [8] 刘杨,黄振中,张羽,李曼丽(2013). 中国MOOCs学习者参与情况调查报告[J]. *清华大学教育研究*, (4):27-34.
- [9] 刘永权,牛健,李莹(2012). 国内外远程开放教育辍学研究之比较[J]. *现代远程教育研究*, (5):57-63.
- [10] 苏芃,罗燕(2013). 技术神话还是教育革命?——MOOCs对高等教育的冲击[J]. *清华大学教育研究*, (4):6-12+21.
- [11] 魏顺平(2014). 大数据时代的MOOC与学习分析[EB/OL]. *中国远程教育青年学者论坛*, <http://ouchn.edu.cn/News/ArticleDetail.aspx?ArticleId=3584a87b-7534-4ffc-bb8a-13236990b2b1&ArticleType=1>.
- [12] Yuan, L., & Powell, S. (2013). MOOCs and Open Education: Implications for Higher Education[EB/OL]. <http://publications.cetis.ac.uk/wp-content/uploads/2013/03/MOOCs-and-Open-Education.pdf>.
- [13] 张羽,李越(2013). 基于MOOCs大数据的学习分析和教育测量介绍[J]. *清华大学教育研究*, (4):22-26.

(编辑:徐辉富)

Analysis of Learning Behaviors in MOOCs ——A Case Study of the Course "Principles of Electric Circuits"

LI Manli¹, XU Shunping¹ & SUN Mengliao

(1. Institute of Education, Tsinghua University, Beijing 100084, China;

2. Computer Science & Technology Department, Tsinghua University, Beijing 100084, China)

Abstract: Many researches indicate the rate of completion of MOOCs (Massive Open Online Courses) is typically low. A large number of MOOC learners drop out during the learning process, which has challenged many MOOC courses instructors and administrators. Since the low rates of course completion are undesired and it will restrict the sustainable development of MOOC, administrators, researchers, and instructors are all trying to change the situation.

Within such a context, we decided to analyze factors which impact learners' course participation and course completion, and intended to provide advice to increase the course completion rates based on these findings. We use the data of the course "Principles of Electric Circuits" instructed by Xinjie Yu on the xuetangX platform to analyze course participation and course completion situation in-depth. We found that learners with stronger learning motivation have higher course participation. That is, if students are active in the forum and the wiki of the course, more likely they tend to complete the course. Learners who registered in the course during the earlier stage are more likely to complete the course and to participate in the course. Learners who are in colleges and universities are more active than those who are not in colleges and universities in the forum of the course. Additionally, we also found that a learner's course participation will have some significantly positive effects on his/her peers, which is worth further attention. The findings from this research study may have some significance for the sustainable development of MOOCs

Key words: MOOC; learning behavior; dropout behavior; course participation; positive externalities