

基于 MOOC 数据的学习行为分析与预测

蒋卓轩 张 岩 李晓明

(北京大学网络与信息系统研究所 北京 100871)

(jzhx@pku.edu.cn)

Learning Behavior Analysis and Prediction Based on MOOC Data

Jiang Zhuoxuan, Zhang Yan, and Li Xiaoming

(Institute of Network Computing and Information System, Peking University, Beijing 100871)

Abstract With the booming of MOOC (massive open online course) in the past two years, educational data analysis has become a promising research field where the quality of teaching and learning can be and is being quantified to improve the educational effectiveness and even to promote the modern higher education. In the autumn of 2013, Peking University released its first six courses on the Coursera platform. Through mining and analyzing the massive data of learning behavior of over 80000 participants from the courses, this paper endeavors to manifest more than one side of learning activity in MOOC. Meanwhile, according to the characteristic of learning behavior in Chinese MOOC, learners are classified into several groups and then the relationship between their learning behavior and performance is thoroughly studied. Based on the above work, we find out that learners' performance, regarding whether he/she could get certificated eventually, can be predicted by looking into several features of their learning behavior. Experiment results indicate that these features can be trained to effectively estimate whether a learner is probably to complete the course successfully. Besides, this method has the potential to partially evaluate the quality of both teaching and learning in practice.

Key words massive open online course (MOOC); engagement style; learning behavior; data analysis; performance prediction

摘 要 随着近 2 年慕课(massive open online course, MOOC)的兴起,教育大数据分析正成为一个新兴的研究方向.2013 年秋,北京大学在 Coursera 上开设了 6 门慕课.通过分析挖掘约 8 万多人参与这 6 门课的海量学习行为数据,力图展现慕课学习活动多个侧面的风貌.同时,首次针对中文慕课中学习行为的特点,将学习者分类,以更加深入地考察学习行为与学习效果之间的关系.在此基础上,通过选择学习者的若干典型行为特征,对他们最后的学习成果进行预测的工作也尚属首次.数据表明:基于学习行为的特征分析能有效地判别一个学习者能否成功完成学习任务获得通过证书,并能找出潜在的认真学习者,这为今后更加精准的慕课教学测评提供了一种依据.

关键词 慕课;学习者类型;学习行为;数据分析;成绩预测

中图法分类号 TP181

在全球化的大背景下,教育领域的全球化趋势也越来越显著。随着互联网 Web 2.0 和云计算等技术的成熟,在过去 2 年里出现了为全球用户提供课程和教育服务的大型网络平台——慕课(massive open online course, MOOC)——这一新的事物。这不仅是互联网应用的创新,人们也认为它将引发高等教育和社会发展的一场变革。在短短 2 年时间里,以 edX, Coursera, Udacity 为代表,慕课理念与实践得到了国内外越来越多优秀大学的认同,包括哈佛大学、麻省理工学院、斯坦福大学、日本京都大学、新加坡国立大学、北京大学、清华大学、上海交通大学和复旦大学等等。他们在慕课平台上免费开放自己的课程,使全球任何希望得到优质教育资源的人都可以自由访问,目前已有近千万用户。这一新的教学模式相比单纯的视频公开课,具有正规课程的效能^[1],保证教学质量;而传统的远程教育相比,慕课又具有免费、课程丰富、自由度大、易于获取等优势。由于其符合教育学中的关联主义理论^[2-4],从而得到了广泛的认可。

慕课的主体是学习者,慕课最大的特点在于海量的学习者和多种多样的学习者群体。传统教育测量中已经假设:学生群体的学习动机基本相同,学生的心智发展和知识水平大致相当;另外由于学生数量较少,测评方法也相对固定和简单。这些在慕课情境下有了很大的变化:学习者的学习动机和知识背景差别呈多样性;在具有海量学习者的慕课环境中,传统测评方法采用的抽样数据评价方法的效用会大打折扣^[5-6]。与之相对,由于互联网技术的特点,学习者在慕课平台上的所有学习行为都会被以形式多样的数据形式记录下来^[5],因此想要对慕课环境中的主体进行测评,不能简单地沿用传统方法。实际上,研究基于大数据技术进行学习效果和课程价值的精准测评^[1]势在必行。

本文正是力图对北京大学在 Coursera 上开设的首批 6 门课程的学习者行为数据进行探索和定量分析,以达到更加深刻认识慕课的目的,从而为慕课课程的教学改进提供借鉴意义。

1 相关工作

慕课出现的时间还很短,国内高校的课程在 2013 年 9 月首次登陆慕课平台,目前有一些分析学习者反馈的课程报告^[7],尚没有形成相关的研究成果。即使从世界范围来看,相关研究也很少,并且都

没有取得大的进展^[8]。最早在 2012 年 12 月份的《科学》杂志上出现了研究人员对慕课的介绍,并展望它将改变未来教育^[9]。《自然》杂志 2013 年 3 月的一篇文章则详细介绍了慕课的发展、现状和趋势^[8],随后引起了越来越多研究人员的关注和参与,2013 年 NIPS(neural information processing systems)会议发起了一个专门针对慕课数据研究的研讨会^[10-15]。麻省理工学院的研究者们在对 edX 上的第 1 门课的完整数据进行研究后发现,学习者的构成、学习者的学习模式以及如何评估教学质量远比想象中复杂^[5],在 Coursera 上也得到类似的结论^[16-19]。文献[20]发现慕课的学习者几乎涵盖所有人类群体类别。文献[21]对 edX 上的首批全部课程进行了系统统计和分析。

总之目前为数不多的研究工作大致可以分为以下 3 类:

1) 通过对某个单维度数据进行统计分析,提出改进意见。如文献[10]专门研究学习者背景数据;文献[11]针对学习者提交的作业数据进行分析;文献[12]分析研究了论坛数据等;文献[22]对学习行为进行了分析,尝试在不同学期相同课程中进行平台改进并对比;文献[23-24]对学习者互评机制提出了改进;文献[25]试图建立对学习者的即时提醒机制以提高学习效果。

2) 从工程和平台的角度提出数据分析方法。如文献[12]为实现根据论坛帖子内容进行合理分类的目的开发了一个框架;文献[14]搭建了一个大数据分析平台 MoocViz,可以对来自不同课程、不同慕课平台的数据综合分析。

3) 对学习行为进行建模,从而预测学习行为。如文献[13, 15, 26-27]研究了学习者的退出率、课程完成率等模型。

本文在借鉴已有为数不多工作的同时,首次对中文慕课课程学习者的学习行为进行了刻画,并且首次根据中文学习者的行为特点进行分类和分析,对学习最终学习成绩的预测工作也尚属首次。

2 学习者类型分析

理解学习者和慕课的交互行为是很重要的问题。心理学认为:人的特定行为可以反映其对某个积极目标激励程度的价值评价,同时行为也是人内在本能推动和当时生活心理状态共同作用的结果。因此通过分析学习行为,可以区分他们的目标、动机、

背景、状态等有价值的信息,从而为个性化改进课程设置和合理评估学习效果提供依据.但同时人的动机无法被直接观察和测量,只能通过人们根据自身的经验对特定行为的“假定构想”而感知^[28-30].通过观察可以发现,我们很难一一准确判断学习者参与课程的目标、课程内容或者证书对其激励的程度、其当时的状态、内部本能推动、人生观和外部环境刺激,而这些都反映在学习行为的差异上.有的学习者会把慕课当作传统课程认真参与学习;有的学习者把慕课只是当作在线的参考或者工具,交互特点呈现不同步性和选择性;还有一些学习者只是把慕课当作检验其对相关知识的掌握程度的工具,因此他们很可能只做题而不看视频.仅在文献^[7]中列出的

学习者动机种类就多达 13 种,如果依据学习者动机分类,仅依靠行为数据既不能准确分类,类别也略显繁多.因此这里我们简化了问题,借鉴人们最直接的经验假设——学习行为越多成绩会越好——来对学习者进行分类.

表 1 对目前在 Coursera 上开设的首批 6 门课程的概况和数据记录数进行了统计,分别是人群与网络(P)、艺术史(A)、大学化学(C)、生物信息学(B)、数据结构与算法分析(D)、计算概论(I).各门课程的内容、要求都不尽相同,表 2 反映的是各门课程计入总分的项目,其中小测验和期末考试 2 项是每门课都有的,因此针对这个特点在设计分类标准时会有所体现.

Table 1 Information about Courses and Click Records
表 1 课程记录与相关信息统计

Course	Number of Registrants	Records of Assignment	Records of Quiz	Records of Video	Number of Threads	Starting Day	Cycle/week	Number of Certification (Ratio of Certification)
P	10 807	125	38 214	156 981	219	2013-10-20	14	149(0.013)
A	16 395	35	20 889	109 118	273	2013-10-20	14	237(0.014)
C	9 124	35	20 284	161 424	110	2013-09-30	11	116(0.013)
B	18 367	35	60 042	340 607	361	2013-09-30	13	581(0.032)
D	13 197	5 204	15 821	98 343	283	2013-10-20	14	57(0.004)
I	14 462	56 087	14 048	435 472	1 029	2013-09-30	15	285(0.020)

Table 2 Information about Course Grading Policy
表 2 课程计分政策

Course	Assignment	Quiz	Peer Grading	Mid-term Exam	Final Exam	Minimum Passing Score (Score with Distinction)
P	N	Y	N	N	Y	70
A	N	Y	Y	N	Y	70
C	N	Y	N	Y	Y	60(85)
B	N	Y	N	N	Y	60(85)
D	Y	Y	N	Y	Y	60(85)
I	Y	Y	N	Y	Y	60(85)

按照传统的教学评价方法,通常是根据成绩将学生评为“好、中、差”3 个档次.但在慕课环境下单纯依靠成绩评价学习者的努力程度和收获是片面的.以“人群与网络”课程为例,有成绩的学习者只占注册人数的 10%左右.表 3 和表 4 分别反映了“人群与网络”课程学习者的动机和知识背景分布^[7],可见慕课环境中学习者的动机和知识背景呈多样性,学习者利用平台的方式也各不相同,大量的学习者并不是为了获得证书而去学习,表 1 中表现的证书获得率均低于 edX 上的课程^[21].

在慕课众多类型的学习行为中,我们选取“看视频”和“提交测验”这 2 个最广泛也最基本的行为进行分析,其中“看视频”代表学习者的学习动机和主动性,“提交测验”表示学习者认真学习的程度,这个行为也是学习者最后获得证书的必要条件.图 1 显示的是 6 门课中“看视频+提交测验”次数对应的人数统计,总体上呈下降趋势,其变化可以总结为 3 个阶段:

1) 先迅速递减,均在次数等于 2 时出现一个较大拐点,然后继续下降;

Table 3 Survey about Learners’ Motivation in Ref [7]

表 3 文献[7]对学习动机统计

Motivation	Ratio of Agreement
For Interest, Curiosity or Fun	0.929
For Experiencing PKU Courses	0.551
For the Instructor	0.306
Communicate with Others with Similar Interests	0.320
Friends Recommendation	0.093
Related with My Major	0.400
Related with My Job or Business	0.392
For Getting a New Job	0.156
For MOOC Certificate form PKU	0.338
Upgrade or Expend Related Knowledge	0.924
Challenge Myself	0.649
For Graduate Admission Test	0.054
For MOOC-related Research	0.267

Table 4 Survey about Learners’ Background in Ref [7]

表 4 文献[7]对学习背景统计

Major Background	Ratio of Background
Natural Science	0.107
Humanity	0.142
Social Science	0.173
Information and Engineering	0.458
Medical Science	0.013
Others	0.040
No Subject Background	0.067

- 2) 随后保持相对平缓(增增减减,但总体变化不大,分离点大约在 $x=23$ 与 $x=30$ 之间的位置);
- 3) 接着是递减比较明显的阶段. 尽管慕课学习的特点是多样性,但将“看视频+提交作业”行为作为分类依据可以在不同课程间建立一致标准的同

时,也可以粗粒度地将不同动机和背景的学习者归类.

由于期望能把学习者分为“好、中、差”3 档,因此需要 2 个阈值. 这里简化了将视频下载观看的行为. 在“差”类别的学习者中,记录数已经较少,甚至是零记录,这部分学习者被归为“打酱油者”(Bystander); 在“中”类别学习者中,其特点是看视频加提交测验的总次数处于中等,可以认为他们没有完整地学习课程内容,这其中包括一部分是偶尔学习了几次就不再继续,或者中间间隔了较长时间没有学习,因此可以将他们再细分为“虎头蛇尾者”(Anti-climax)和“始终摇摆者”(Waverer); 在“好”类别学习者中,其特点是具有较多的学习行为记录,但是结合较少的通过率,可以认为其中很多学习者并不是为了拿证书而学习,因此我们将这类学习者细分为“坚定完成者”(Stalwart)和“只为学习者”(Just-learner),他们的差别在于有没有参加期末考试.

定义了 5 种基本学习者类型之后,接下来需要确定分类的阈值. 我们通过将图 1 的 x 轴分成 3 段,从左至右分别对应“差、中、好”3 个类别学习者,然后再在每个大的类别中细分. 对于某一门课程,定义 2 个阈值: c_1, c_2 ; 定义某学习者看视频记录数 l ,提交作业记录数为 q ; 定义某学习者的第 1 次记录和最后 1 次记录之间的间隔天数 t (记录指看视频或者提交测验),则学习者的记录密度 $p=(l+q)/t$; 定义 θ 为“中”类别的所有学习者的记录密度均值; 则对于一个学习者来说,他被称作 5 种情况:

- 1) 打酱油者,如果 $l+q\leq c_1$;
- 2) 虎头蛇尾者,如果 $c_1<l+q\leq c_2$ 且 $p\geq\theta$;
- 3) 摇摆不定者,如果 $c_1<l+q\leq c_2$ 且 $p<\theta$;
- 4) 坚定完成者,如果 $l+q>c_2$ 且参加期末考试;
- 5) 只为学习者,如果 $l+q>c_2$ 且未参加期末考试.

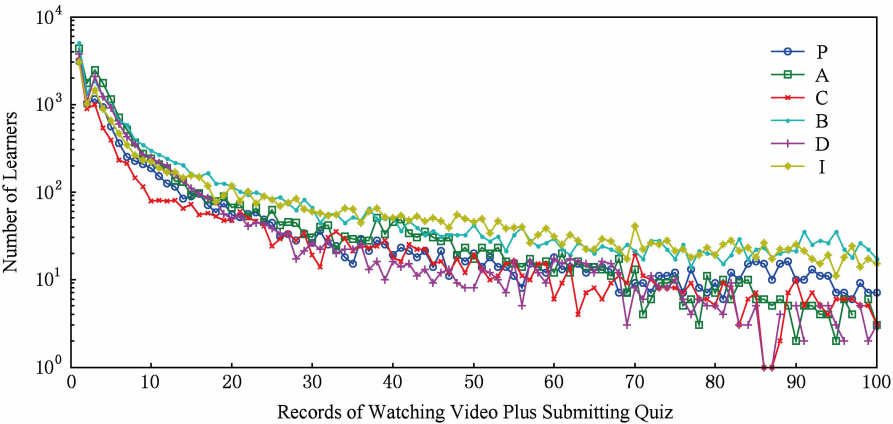


Fig. 1 Number of learners VS records of course engagement.

图 1 学习者参与课程记录的统计

根据之前的观察和图 1 反映的 3 个阶段趋势, 选择 $c_1=2, c_2$ 取 $x=2$ 到 $x=30$ 之间出现的所有局部波峰中位数的 x 值, 则各门课程的参数如表 5 所示. 表 6 是各门课程中不同学习者类型的数目和比例, 数据本身较为直观和合理地区分了学习者的行为类别, 其中“坚定完成者”的比例较小, 这也与获得证书率低的现象相互印证; 大约有一半的学习者只是尝鲜或者因为其他原因没有参与到课程中来; 在实验中对 c_2 选用了不同的阈值, 发现“坚定完成者”这一类别的数目和比例影响都不大, 这也反映了参加期末考试的学习者普遍有较高次数的看视频和提交作业记录. 另外由于课程不同, 各自在教学内容、

难度和要求上都有差别, 因此各课程中学习者类型的比重也各异.

Table 5 Parameters for Classification of Learners
表 5 学习者分类参数

Course	c_1	c_2	θ
P	2	26	4.71
A	2	22	3.19
C	2	16	4.79
B	2	25	4.21
D	2	22	3.92
I	2	20	4.60

Table 6 Number and Ratio of Learners in Different Groups
表 6 不同类型学习者人数和比例

Course	Number of Learners(Ratio of Learners)				
	Bystander	Anti-climax	Waverer	Stalwart	Just-learner
P	5 265(0.49)	1 395(0.13)	2 564(0.24)	248(0.02)	1 335(0.12)
A	8 533(0.52)	2 269(0.14)	4 079(0.25)	295(0.02)	1 219(0.07)
C	5 066(0.56)	865(0.09)	1 328(0.15)	210(0.02)	1 655(0.18)
B	8 218(0.45)	2 299(0.13)	4 462(0.24)	603(0.03)	2 785(0.15)
D	6 857(0.52)	2 321(0.18)	2 993(0.23)	63(0.00)	963(0.07)
I	5 496(0.38)	1 670(0.12)	2 842(0.20)	271(0.02)	4 183(0.29)

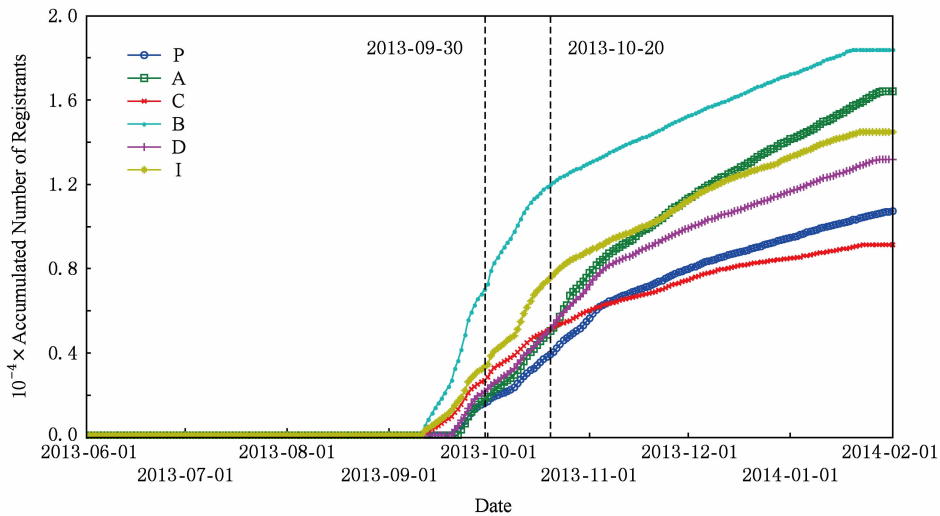
3 学习者参与行为分析

从学习者参与的角度,除了“看视频”和“提交测验”这 2 个最基本的学习行为之外,不同类型的学习者何时注册课程、学习者的成绩有什么规律、学习者

参与论坛的情况等等这些问题也是值得关注的. 下面就结合第 2 节提出的 5 类学习者分类,分别对学习 者与时间、成绩和在论坛相关的行为进行分析.

3.1 时间相关行为分析

学习者参与课程的时间是一个很重要的学习行为. 在已有的数据中发现,通常在开课前好几个月就



The two vertical dash lines denote the starting days of specific courses.

Fig. 2 Accumulated number of registrants.

图 2 累计注册人数

有注册课程的记录,而不同于传统课程之处是慕课在开课后甚至课程结束后仍然有大量学习者注册加入课程.图 2 显示了各门课程随时间变化累积注册人数的增长情况.以“人群与网络”这门课程为例,可以看到约有 36%的学习者在开课前注册,在课程结束后仍然有 2%的学习者新加入课程.与针对英文课程的研究相比,中文课程的特点是大多数学习者在开课附近和结束前的期间注册,而并不是英文课程所表现的:60%的学习者在开课前注册,有 18%的学习者在课程结束后注册^[22].中英文课程共同点是:均期望更多的学习者能从始至终紧跟课程学

习(在课程之前或开课前期注册),但更多的学习者在课程开始后的中后期甚至课程结束后才注册,这都与初衷不符;而中英文课程设置的差异可能在理念上:英文课程更注重教育的普及性和广泛性,学习难度不高,而中文课程更强调了学习者的努力过程和学习效果,教学内容更全,难度更大,但尽管如此中英文课程的通过率都较低.

不同期间注册的学习者在行为上也是不同的.图 3 是人群与网络课程中各学习者类型随时间变化的比例分布,其中打酱油者从开课前一个月的 100% 迅速下降至开课时的 35%,然后又在课程结束时达

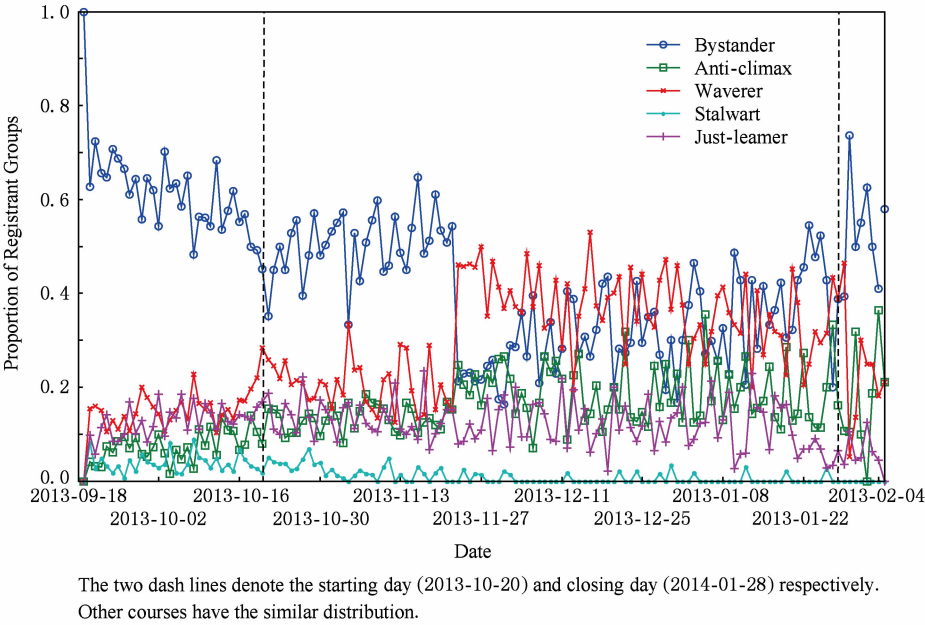


Fig. 3 Proportion of registrant groups in people and network.
图 3 人群与网络各类型注册者比例分布

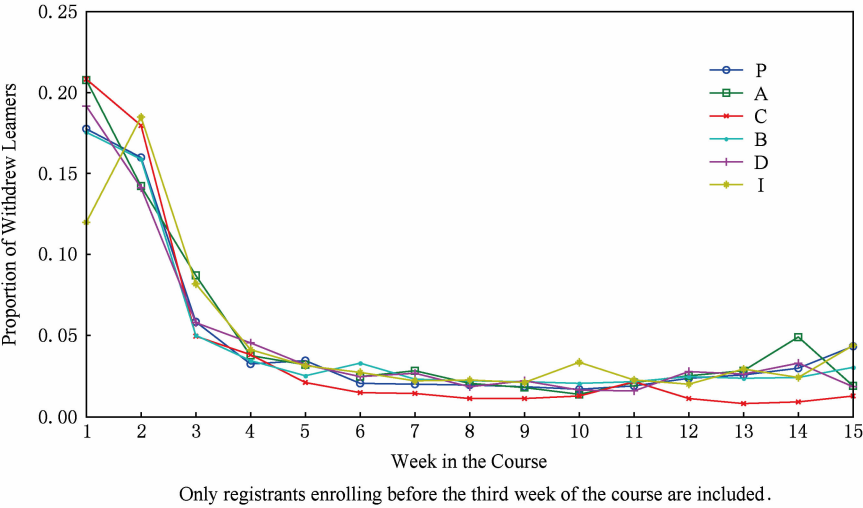


Fig. 4 Proportion of active registrants whose last action in a course is in a particular week x .
图 4 学习者在第 x 周退出比例

到了74%；“中等”学习者更多在课程的中后期参与进来,而只有尽早地参与课程才有可能获得“好”的成绩。

接着看看学习者退出课程的情况.图4是所有第3周之前注册的学习者在第 x 周离开课程的比例,例如人群与网络课程中在第5周退出者占了3.4%.大多数课程在前2周呈急速下降的趋势(计算概率课程相反),随后第3周进入平稳递减阶段,最后均在期末考试周有微小反弹.虽然理想情况是所有学习者均在期末考试周退出,但事实正好相反.另外平均约有33%的学习者在第3周之前注册,课程结束之后还有记录.他们并不一定是想关注课程后续信息或者自己成绩的“忠实粉丝”,也有可能是之前没有认真学习,在课程结束后想起来并返回课程的学习者,这

两者的共同点是当初都有想学好这门课程的愿望.

3.2 成绩相关行为分析

学生成绩是传统课堂中评价学习效果的重要指标.在慕课中虽然学习者的知识背景不同、学习动机不同、学习方法各异,但成绩毕竟决定了学习者最后是否能获得证书.因此结合学习者的行为分析其成绩分布,可以观察慕课成绩有着怎样的作用和价值.

从图5的各门课程学习者最终成绩分布图可以看出,6门课程虽然不同,但是学习者的成绩分布相似.以“人群与网络”课程为例,10 807个注册者之中,有9 677个学习者的成绩是0分,约占了90%;862个学习者成绩大于0但是不能获得证书(约8%);149个学习者获得了证书(约1%).

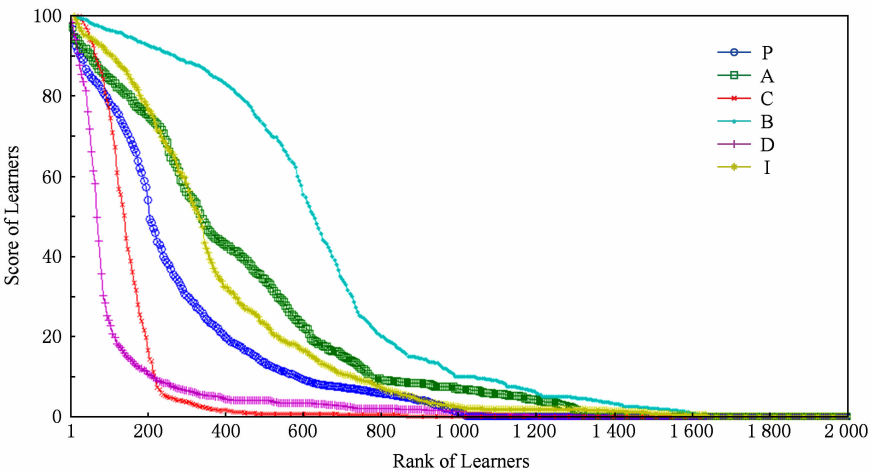


Fig. 5 Score of learners VS rank of learners.

图5 学习者成绩排名

表7是6门课程学习者成绩分布的统计,可以看出同一些英文的“人人拿证书”理念的课程相比,印证了首批的中文慕课普遍具有一定的难度,获得证书人数较少;同时绝大多数是0分,成绩不是0分又没有拿到证书的学习者则是今后需要辅导或者帮助的群体.

Table 7 Number and Ratio of Learners in Score Distribution

表7 不同成绩学习者人数和比例

Course	Number of Learners(Ratio of Learners)		
	0 Score	>0 but no Certificate	Certificated
P	9 677(0.90)	862(0.08)	149(0.01)
A	14 832(0.90)	1 084(0.07)	237(0.01)
C	7 884(0.86)	747(0.08)	116(0.01)
B	16 747(0.91)	1 039(0.06)	581(0.03)
D	11 983(0.91)	1 130(0.09)	57(0.00)
I	12 257(0.85)	1 348(0.09)	285(0.02)

当然获得0分的学习者不一定就说明他不努力,其中部分学习者是认真地并且用很长时间观看了视频,只是没有提交作业.图6(a)支持了这样的观点,显示了0分学习者观看视频的分布.以“人群与网络”课程为例,注册该课程的65%学习者至少观看了1次视频,21%学习者的至少观看了10次视频, $x=83$ 是所有视频数目,其他课程也有相似的分

布结果.

下面探索一下不同学习行为与成绩之间的关系.假设最终成绩是相互独立的,目标是试图用数据揭示低分学习者和高分学习者之间在行为上的差异.

图7显示了各门课程中不同分数段对应的行为分布,4种行为包括:看视频、提交测验、看论坛帖子、参与论坛(发帖、回帖或者评论).总的来看分数越高者各项行为的参与度也越高,大部分课程呈现了线性的关系.其中值得注意的是数据结构与算法分析课程D,虽然60~70分的学习者在论坛参与度

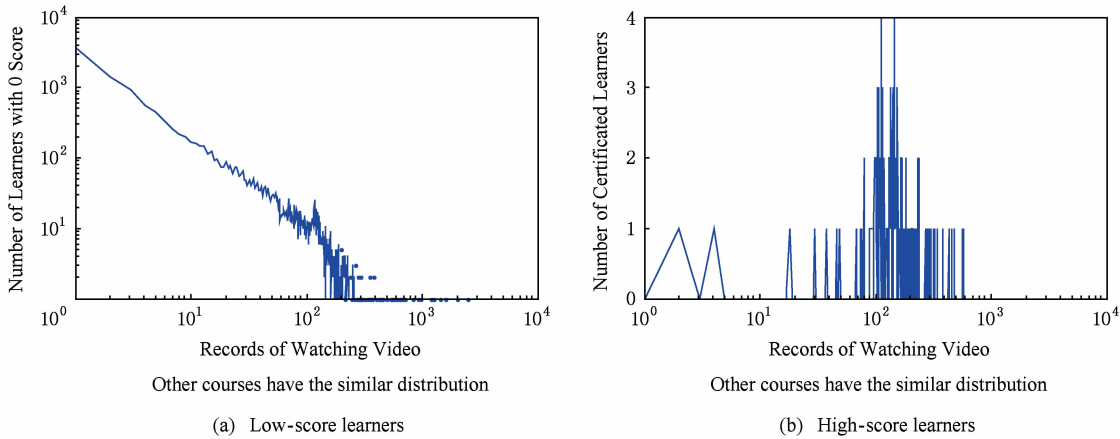


Fig. 6 Behavior of watching video of low-score and high-score learners in people and network.

图 6 人群与网络学习者看视频行为统计

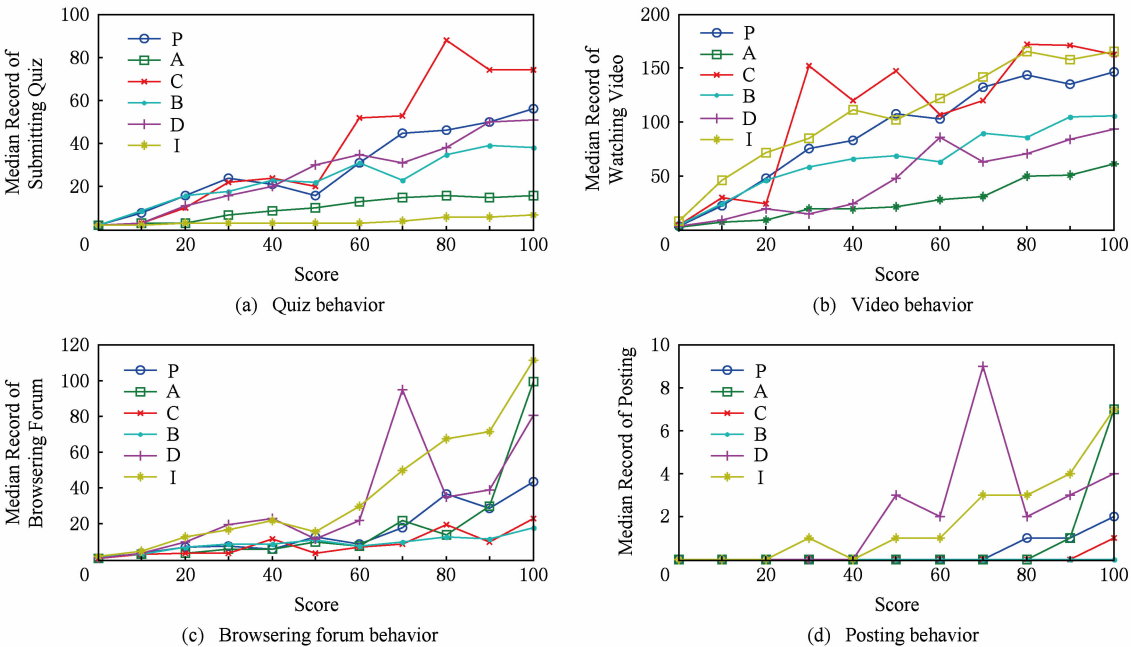


Fig. 7 Distribution of four kinds of behavior VS score in the course.

图 7 4 种参与行为在不同分数段的分布

上比其他分数的同学都高,但是他们看视频和提交测验的次数均低于 50~60 分(不及格)和更高分数的同学;大学化学 C 在看视频和提交测验 2 个行为上波动较大,特别是相当多的学习者提交了测验却只得到了不及格的分数;在论坛相关的行为上,各门课程的学习者在被动参与(看帖)上和主动参与(发帖)上都保持了一致的趋势,可见“需要努力才能取得好成绩”是北京大学首批 6 门课程的特点,而论坛的足够参与也是取得好成绩的有力保证。

接着再对获得证书的学习者群体进行分析,图 6(b)显示获得证书的学习者普遍观看了很多次视频,而且大多数超过了课程的总视频数目 83,说

明他们很多都会对一个视频学习多次,当然也有少部分学习者看的视频次数非常少也获得了证书,可能他们要么已经学习过相关课程只是为了获得证书或检验自己,要么他们将视频下载观看学习而没有在线观看的记录,总之这部分人数较少。从图 8 可以看出,人群与网络课程中获得证书的学习者提交 14 次的人数最多(13 次课后作业和 1 次期末考试),更多的学习者有多次提交的情况,甚至最多的达到了 111 次。由于视频中也有一些不计分数的测验,因此次数为 14 的学习者可能是下载视频进行学习,也可能是完全没有看视频直接做题。

下面看看学习者类型与成绩之间的关系,图 9

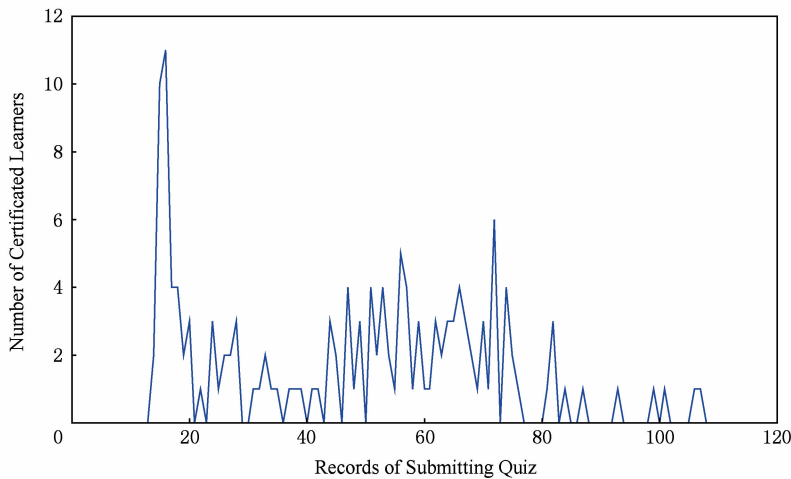


Fig. 8 Number of certificated learners VS records of submitting quiz in people and network.

图 8 人群与网络获证书者提交测验次数

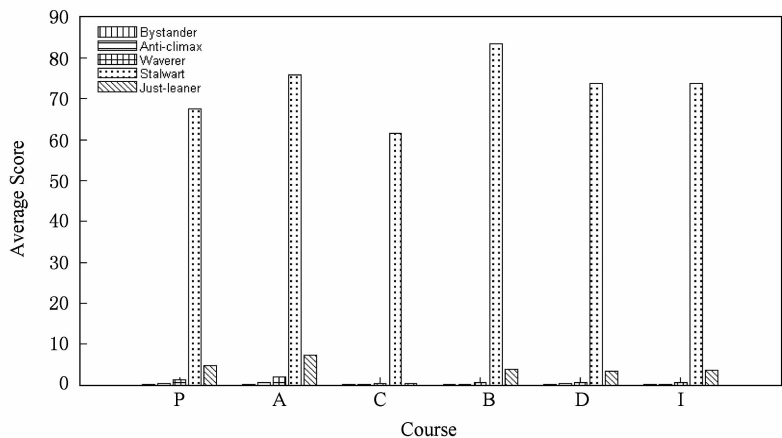


Fig. 9 Average score of different learner groups in the course.

图 9 不同学习者类型成绩均分

Table 8 Mutual Distribution of Certificated Learners and Learner Groups

表 8 获证书者与学习者类型相互分布

Course	Proportion	Bystander	Anti-climax	Waverer	Stalwart	Just-learner
P	$P(S C)$	0.000	0.000	0.013	0.987	0.000
	$P(C S)$	0.000	0.000	0.001	0.593	0.000
A	$P(S C)$	0.000	0.000	0.038	0.962	0.000
	$P(C S)$	0.000	0.000	0.002	0.773	0.000
C	$P(S C)$	0.000	0.000	0.000	1.000	0.000
	$P(C S)$	0.000	0.000	0.000	0.552	0.000
B	$P(S C)$	0.000	0.000	0.015	0.969	0.015
	$P(C S)$	0.000	0.000	0.002	0.934	0.003
D	$P(S C)$	0.000	0.000	0.000	0.930	0.070
	$P(C S)$	0.000	0.000	0.000	0.841	0.004
I	$P(S C)$	0.004	0.000	0.021	0.772	0.204
	$P(C S)$	0.000	0.000	0.002	0.812	0.014

Note: $P(S|C)$ denotes the distribution of groups in certificated learners.
 $P(C|S)$ denotes the ratio of certificated learners in groups.

显示的是各门课程中每个类型学习者成绩的平均分,坚定完成者均分高于及格线,而其他类型学习者的成绩几乎是 0 分,可以认为坚定完成者都是潜在证书获得者.表 8 显示每个课程中不同学习者类型(用 S 表示)在获得证书学习者群体(用 C 表示)中分别所占的比例和不同类型学习者中各自获得证书的比例.可以认为打酱油者和虎头蛇尾者几乎完全没有获得证书的可能,但是计算概论课程中有非常少的打酱油者获得了证书,可能原因是与其课程成绩评定方式有关,其测验只有 3 次共占成绩的 45%,重点是编程作业部分占 50%,因此该课程的打酱油者可能恰好满足:没有看视频记录(下载观看或直接做题)、3 次以内的测验提交记录、分数在及格分以上.另外始终摇摆者和只为学习者中获得证书的比例也微乎其微,坚定完成者几乎可以和获得证书群体重合.

3.3 不同学习者参与论坛情况

Coursera 的论坛提供了一个学习者之间以及与教学人员之间交流沟通的平台,参与论坛与否完全是学习者的自主行为,当然教学人员制定最终成绩的评价标准也可以与学习者的论坛表现建立关联,但目前基于海量学习者的情况下,还没有提供一个自动的机制根据学习者论坛表现来打分,所以本

文中讨论的参与课程论坛均是学习者的独立行为.

结合第 2 节提出的学习者分类方式,先看看不同类型学习者在论坛中的表现,表 9 显示了不同类型学习者的论坛参与情况,分别用 $P(S|F)$ 和 $P(F|S)$ 表示,其中 S 表示不同类型的学习者, F 表示参与过论坛的学习者.各门课程中参与论坛学习者类型的分布 $P(S|F)$ 展现了截然不同的情况:有的课程打酱油者超过了 10%(人群与网络、艺术史、数据结构与算法分析、大学化学),其余的则低于 10%;有的课程始终摇摆者占多数(艺术史、数据结构与算法),其余的是只为学习者占多数;有的课程“差”和“中”类别学习者总数超过了半数(人群与网络、艺术史、数据结构与算法分析),其余的则低于半数.它们的共同点是虎头蛇尾者和坚定完成者均不占多数.中文课程的打酱油者与英文课程相比展现了近似的参与度^[22],以人群与网络课程为例,总注册人数中占了约 50%的打酱油者在参与论坛的学习者中只占了 14%,实际上参与论坛的只是注册者中很少的一部分.“中”类别的学习者中,虎头蛇尾者的比重普遍偏低(数据结构与算法分析除外),但始终摇摆者比重均比其高.各门课程的坚定完成者在论坛中的比重普遍较低,可能跟这类学习者绝对数量较少有关.

Table 9 Mutual Distribution of Forum Participants and Learner Groups
表 9 论坛参与者与学习者类型相互分布

Course	Proportion	Bystander	Anti-climax	Waverer	Stalwart	Just-learner
P	$P(S F)$	0.140	0.125	0.235	0.136	0.364
	$P(F S)$	0.044	0.148	0.151	0.903	0.450
A	$P(S F)$	0.284	0.100	0.339	0.096	0.181
	$P(F S)$	0.098	0.129	0.244	0.956	0.437
C	$P(S F)$	0.152	0.072	0.140	0.182	0.454
	$P(F S)$	0.029	0.082	0.103	0.848	0.269
B	$P(S F)$	0.085	0.099	0.274	0.169	0.373
	$P(F S)$	0.032	0.134	0.191	0.869	0.416
D	$P(S F)$	0.197	0.236	0.361	0.033	0.173
	$P(F S)$	0.051	0.179	0.212	0.921	0.317
I	$P(S F)$	0.066	0.066	0.131	0.092	0.645
	$P(F S)$	0.034	0.111	0.130	0.959	0.435

Note: $P(S|F)$ denotes the distribution of groups in forum participants.
 $P(F|S)$ denotes the ratio of forum participants in groups. Forum participation means reading or submitting at least one record.

从各类别学习者参与论坛的比例情况来看 $P(F|S)$ 表现了一定的规律,坚定完成者与参与论坛学习者的重合度最高,范围从 85%~96%,结合表 8 的数据,可以认为获得证书的学习者与参与论坛的

学习者几乎重合,但参与论坛不是必要条件.其他类型的学习者参与比例按从高到低的顺序一致地表现为:只为学习者>始终摇摆者>虎头蛇尾者>打酱油者.总的来看,注册比例越少的类型参与论坛的比例越高,即“好”>“中”>“差”,因此中文课程中论坛的价值得到了体现,但其潜力也还有进一步挖掘和提升的空间.

学习者参与论坛通常会是因为学习上遇到困难或者寻找学习伙伴.值得注意的是人群与网络、艺术史和数据结构与算法分析这3门课程,其“差”和“中”类型学习者相加的比重均高于其“好”类别的学习者,考虑到之前图7表现的其论坛参与度高的学习者分数并不高,似乎可以认为这3门课的论坛并没有起到该有的作用,其中以数据结构与算法分析课程为甚.“差”和“中”类型学习者相加的比重高也可以认为是该课程难度大.

4 学习效果预测

对于慕课来说,提高通过率是其平台效用最重要的体现,而已有相关工作均没有用学习者的行为特征进行成绩预测.本节将试图根据前面分析的内容和结论对学习者的成绩进行预测,作为基准测试.

4.1 问题描述和数据集

我们期望在1门课程进行的过程中能实时预测学习者的学习状况——最终能否获得证书.本文的课程具有11周至15周不等的时间跨度,而且从图2和图4可以看出,各门课程均在开课前2周左右时间里出现注册人数和退出人数的较大波动,并且学习人数在大约2周后逐渐稳定.我们假设分别只拥有每门课程前5周、前6周和前7周注册的学习者及其期间的行为数据,并且过滤了数据中非学习者的行为记录,表10是数据样本量统计.定义

Table 10 Number of Samples
表10 数据样本量

Course	1 st 5 Weeks	1 st 6 Weeks	1 st 7 Weeks
P	7 491	7 927	8 332
A	10 566	11 280	12 019
C	6 158	6 459	6 720
B	13 206	13 738	14 189
D	9 401	9 868	10 277
I	9 042	9 543	9 990

每门课程有n个学习者,每个学习者具有p个特征值,即

$$\mathbf{X}=\{\mathbf{X}_1,\cdots,\mathbf{X}_p\}\in\mathbb{R}^{n\times p},$$

预测值为

$$\mathbf{Y}=f(\mathbf{X})\in\mathbb{R}^n,\ y\in\{0,1\},$$

其中,y=0表示未获得证书,y=1表示获得证书.

4.2 特征选择与实验方法

基于第3节的分析,选择对成绩有影响且课程共有的特征:看视频次数(L)、提交测验次数(Q)、记录密度(ρ)、论坛发帖次数(FW)、论坛看帖次数(FR)、注册时间距离开课日期的天数(D).

针对每门课程,将数据集按3:1:1的比例随机分成训练集、验证集和测试集.每次实验用训练集训练参数,用验证集选择最优参数,然后用测试集计算各项指标.采用3种分类模型:线性判别分析(linear discriminant analysis, LDA)、逻辑回归(logistic regression, LR)和线性核支持向量机(linear support vector machine, l-SVM).实验结果如下:表11是对各门课程用不同分类器预测的结果,可以看出准确率均较高,而且在不同的课程中,3种分类器表现出

Table 11 Prediction Result of Certification with Registrants
in the 1st 7 Weeks

表11 前7周学习者获证书的预测结果 %

Course	Classifier	Accuracy	Precision	Recall	F-score
P	LDA	97.2	40.3	90.6	55.8
	LR	97.9	47.7	63.6	54.5
	l-SVM	96.3	34.4	93.9	50.4
A	LDA	98.3	50.0	82.9	62.4
	LR	98.4	55.6	68.2	61.2
	l-SVM	98.3	54.7	61.7	58.0
C	LDA	99.6	77.3	100	87.2
	LR	99.7	85.7	94.7	90.0
	l-SVM	99.7	89.3	96.2	92.6
B	LDA	96.7	58.4	82.8	68.5
	LR	97.3	64.7	68.8	66.7
	l-SVM	97.0	62.1	83.3	71.2
D	LDA	99.6	50.0	88.9	64.0
	LR	99.5	65.0	76.5	70.3
	l-SVM	99.7	64.3	81.8	72.0
I	LDA	97.1	51.9	68.9	59.2
	LR	97.0	57.6	54.3	55.9
	l-SVM	96.8	50.0	66.7	57.1

一致的性能,例如大学化学课程中,3 种分类器的查全率和 F 值相比其他课程均较高.图 10 统计了各门课程预测 F 值的时序变化,显示课程间变化较

大,课程内变化较小.这个现象也证实了慕课具有“以小见大”的特点,但也值得进一步分析通过 F 值进行课程间对比的方法.

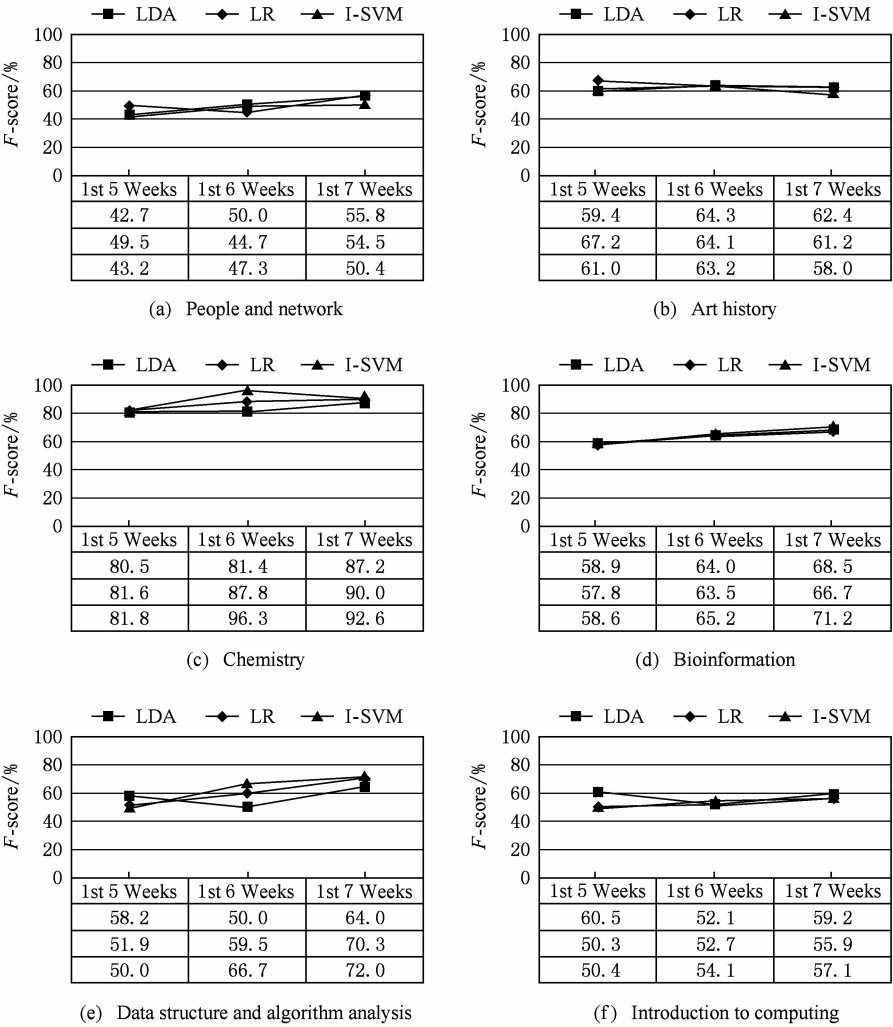


Fig. 10 Sequential variation of F -scores.

图 10 F 值时序变化

4.3 实验结果分析

由于已有研究并无涉及这方面的预测工作,因此本节的实验可以作为基准测试.实验中达到相对较高的查全率可以通过调节分类器的参数来实现,事实上这是有意义的,因为这样不仅可以准确找出“最终”获得证书的学习者,同时也找出了“有可能”但“最终”没有获得证书的学习者,从而可以通过教师或者平台进一步的个性化课程辅导与提示将“有可能”获得证书者转变成为“最终”获得证书者.因此通过计算学习者获得证书的概率这个过程性指标,不仅可以自动监测课程学习者的实时学习状况,也可以通过其他配套措施,提高课程通过率.

从心理学的角度来看,学习慕课作为人的一种

高级行为,学习者都受到目标最终达成的激励因素影响.从通过率来看,可能证书对学习者的激励程度不高,另外也可能某些外部环境——如新的学习模式不利于促进学习,这需要进行进一步地分析和挖掘.从预测的各项指标结果来看,选取的 6 项行为特征似乎还不能完全准确地刻画学习行为,要做到这一点,还需要考虑结合学习者的目标、对目标的期待和价值估计、学习时的状态、内性格、人生观和外部环境刺激等更加复杂的因素.心理学家 Hull^[31]将行为趋势描述为习惯、欲望、激励 3 种强度的综合,因此大量学习者在第 2 周退出现象可能解释为:新的学习模式还不为大众所接受而没有养成习惯、人们学习的欲望不强以及拿证书的激励价值不高.另一

方面,认真学习者的人数在前期就表现比较稳定的现象可能源于他们接受了新的学习模式,而且(或者)学习和拿证书的欲望强烈.

5 结论与未来工作

2013 年秋,北京大学在 Coursera 上开设了 6 门慕课,长度在 11 周至 15 周不等. 本文对超过 8 万参与那些课程学习者的行为数据进行了多侧面分析,力图能从首批开设的部分中文慕课中发现一些具有普遍意义的学习行为特点和规律. 同时,我们也尝试基于学习行为特征来预测学习者的最终学习效果(即能否获得证书). 慕课的学习行为与传统课堂的学生行为有着很大的差别,基于这 6 门课程的数据我们给出了一种将学习者分为 5 种类型的方法,并且结合学习行为分析了每种类型的特点. 最后,选取若干典型学习行为作为特征,对学习者成绩进行了预测,实验结果表明选取的学习行为特征不仅能在课程进行中实时有效地区分出潜在的证书获得者,而且能为学习者实时的学习状况自动监测、课程教学效果的实时反馈提供获取方法,为课程教学的改进提供参考意见. 从数据中发现的现象和主要结论如下:

1) 约 60% 以上的学习者是在课程开始后注册进来的(这可能与这批课程发布时间和开课时间相距较近有关,今后我们会进一步对比观察这个数据);大约有一半的学习者没有有效参与到课程中来(打酱油者),坚持参与学习的人数在第 3 周基本稳定,大约占最高峰注册人数的 20%.

2) 一直坚持到最后参加期末考试的学习者看视频和提交作业次数较多,分数较高者各项学习行为的参与度也较高,但的确能看到有些人参与活动较多,分数不高.

3) 大部分获得证书的学习者参与了论坛,但总的来说参与论坛的学习者不多(接近 20%),从不同成绩段对应的论坛参与度分布可以分析课程教学效果(例如分数低的学习者参与论坛多说明论坛没有解决他们的问题,或者他们没有掌握讲课的要点),结合通过率可以进一步地分析课程难易程度(例如难度大的课程会表现为:低分(不及格)学习者论坛参与度高、课程通过率低等).

4) 与国外慕课的情况相似,课程绝对通过率(获得证书人数/最高峰注册人数)虽然都较低(本文

涉及的这批课程不到 4%),但同时也有相似占比的积极学习行为,例如 60% 以上的学习者观看过视频,40% 左右的学习者提交过测验等.

5) 根据学习者在前半学期的学习行为,可以较准确地预测其最终的学习成果(是否能获得通过证书). 这也许表明了:在慕课中,一个学习者若在前半学期表现很好,说明他/她既有能力,也很坚定,于是最终获得证书的可能性就很高. 预测结果的时序变化不大反映了慕课“以小见大”的特点,同时分析预测结果的时序变化趋势也是今后探索建立课程间评价指标的一个思路.

随着越来越多的大学和学习者加入,更多的优质课程将会使更广泛的人群受益,慕课的意义和价值也将越来越大. 但是由于慕课发展时间不长,目前还面临着许许多多的挑战,进一步认识慕课以及研究学习者行为和学习者目标、动机、背景等之间的关系是一个有意义的基础问题. 我们考虑接下来的研究将会是:学习者行为和成绩的预测模型;帮助学习者学习的个性化推荐机制;从海量学习者中自动化地精准定位每一位学习者对知识点的掌握情况或者是否需要辅导;更加深入地多层次理解学习者在论坛上的行为和内容;找到刻画课程中“教”与“学”状态的多维度指标并研究其效用等等.

参 考 文 献

- [1] Li Xiaoming. MOOC, showcase or shop? [J]. Communications of the China Computer Federation, 2013, 9 (12): 24-28 (in Chinese)
(李晓明. 慕课: 是橱窗? 还是店堂? [J]. 中国计算机学会通讯, 2013, 9(12): 24-28)
- [2] Siemens G. Connectivism: A learning theory for the digital age [EB/OL]. (2005-04-05) [2014-06-01]. http://www.ingedewaard.net/papers/connectivism/2005_siemens_ALearningTheoryForTheDigitalAge.pdf
- [3] Osvaldo R C. Moocs and the AI-Stanford like courses: Two successful and distinct course formats for massive open online courses [J/OL]. European Journal of Open, Distance and E-Learning, 2012(2): 1-13. [2014-06-01]. <http://files.eric.ed.gov/fulltext/EJ982976.pdf>
- [4] Downes S. New technology supporting informal learning [J]. Journal of Emerging Technologies in Web Intelligence, 2010, 2(1): 27-33
- [5] Breslow L, Pritchard D, DeBoer J, et al. Studying learning in the worldwide classroom: Research into edX's first MOOC [J]. Research & Practice in Assessment, 2013, 8(1): 13-25

- [6] McAndrew P, Scanlon E. Open learning at a distance: Lessons for struggling MOOCs [J]. *Science*, 2013, 342 (6165): 1450-1451
- [7] Team of PKU MOOC Teaching Research. 2013 first MOOCs of Peking University: People, crowd and network course investigation and analysis report [R/OL]. Beijing: Peking University, 2014 [2014-06-01]. <http://pan.baidu.com/s/1c0vUMm0> (in Chinese)
(北大慕课教学研究组. 2013 北京大学首批慕课(MOOC)人群与网络调查分析报告[R/OL]. 北京: 北京大学, 2014 [2014-06-01]. <http://pan.baidu.com/s/1c0vUMm0>)
- [8] Waldrop M M. Online learning: Campus 2.0 [J]. *Nature*, 2013, 495(7440): 160-163
- [9] Stein L A. Casting a wider net [J]. *Science*, 2012, 338 (6113): 1422-1423
- [10] Nesterko S, Dotsenko S, Hu Q, et al. Evaluating geographic data in MOOCs [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://nesterko.com/files/papers/nips2013-nesterko.pdf>
- [11] Han F, Veeramachaneni K, O'Reilly U. Analyzing millions of submissions to help MOOC instructors understand problem solving [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://groups.csail.mit.edu/EVO-DesignOpt/groupWebSite/uploads/Site/ProblemAnalytics.pdf>
- [12] Stump G S, DeBoer J, Whittinghill J, et al. Development of a framework to classify MOOC discussion forum posts: Methodology and challenges [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. http://tll.mit.edu/sites/default/files/library/Coding_a_MOOC_Discussion_Forum.pdf
- [13] Yang D, Sinha T, Adamson D, et al. "Turn on, tune in, drop out": Anticipating student dropouts in massive open online courses [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://lytics.stanford.edu/datadriveneducation/papers/yangetal.pdf>
- [14] Dernoncourt F, Do C, Halawa S, et al. MoocViz: A large scale, open access, collaborative, data analytics platform for MOOCs [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://groups.csail.mit.edu/EVO-DesignOpt/groupWebSite/uploads/Site/MoocViz.pdf>
- [15] Ramesh A, Goldwasser D, Huang B, et al. Modeling learner engagement in MOOCs using probabilistic soft logic [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://lytics.stanford.edu/datadriveneducation/papers/rameshetal.pdf>
- [16] Boxtel T. Interesting participant demographics from a Coursera course [EB/OL]. 2013 [2014-06-01]. <http://tim.vanboxtel.ca/wp/interesting-participant-demographics-from-a-coursera-course/>
- [17] Severance C. Visualizing the geographic distribution of my Coursera course [EB/OL]. 2012[2014-06-01]. <http://www.dr-chuck.com/csev-blog/2012/09/geographic-distribution-of-my-coursera-course/>
- [18] Yager J. Massive open online courses with a global outreach [EB/OL]. 2013 [2014-06-01]. http://www.aahcdc.org/Portals/0/mtgs/ifl3/Yager_Jim_Massive%20Open%20Online%20Courses%20with%20Global%20Outreach.pdf
- [19] Balch T. MOOC student demographics [EB/OL]. 2013 [2014-06-01]. <http://augmentedtrader.wordpress.com/2013/01/27/mooc-student-demographics/>
- [20] Klapp A. MOOCs open doors for diverse student body [EB/OL]. 2013 [2014-06-01]. <http://www.diversityjournal.com/10107-moocs-open-doors-for-diverse-student-body/>
- [21] Ho A D, Reich J, Nesterko S O, et al. HarvardX and MITx: The first year of open online courses, HarvardX and MITx Working Paper No.1 [R]. Cambridge: Harvard University and MIT, 2014
- [22] Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses [C] //Proc of the 23rd Int Conf on World Wide Web. New York: ACM, 2014: 687-698
- [23] Diez J, Luaces O, Betanzos A A, et al. Peer assessment in MOOCs using preference learning via matrix factorization [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://lytics.stanford.edu/datadriveneducation/papers/diezetal.pdf>
- [24] Shah N B, Bradley J K, Parekh A, et al. A case for ordinal peer evaluation in MOOCs [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://lytics.stanford.edu/datadriveneducation/papers/shahetal.pdf>
- [25] Williams J J, Williams B. Using interventions to improve online learning [C/OL] //Neural Information Processing Systems Workshop on Data Driven Education, 2013 [2014-06-01]. <http://lytics.stanford.edu/datadriveneducation/papers/williamswilliams.pdf>
- [26] Balakrishnan G. Predicting student retention in massive open online courses using hidden markov models, UCB/EECS-2013-109 [R/OL]. Berkeley: University of California at Berkeley, 2013 [2014-05-01]. <http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-109.pdf>
- [27] Kizilcec R, Piech C, Schneider E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses [C] //Proc of the 3rd Int Conf on Learning Analytics and Knowledge. New York: ACM, 2013: 170-179
- [28] Heckhausen J, Heckhausen H. Motivation und Handeln [M]. Berlin: Springer, 2006
- [29] Heider F. The Psychology of Interpersonal Relations [M]. New York: Psychology Press, 1982

- [30] Weber H, Rammsayer T. Das problem der motivation [G] // Handbuch der Psychologie, Band 2. Göttingen: Hogrefe, 2005: 414-464
- [31] Hull C L. A Behavior System: An Introduction to Behavior Theory Concerning the Individual Organism [M]. New Haven: Yale University Press, 1952



Jiang Zhuoxuan, born in 1987. PhD candidate from Peking University. His current research interests include machine learning, MOOC data analysis and online education.



Zhang Yan, born in 1970. Professor and PhD supervisor in Peking University. His current research interests include information retrieval, text analysis and data mining (zhy@cis.pku.edu.cn).



Li Xiaoming, born in 1957. Professor and PhD supervisor in Peking University. His current research interests include search engine and Web data mining, parallel and distributed system (lxm@pku.edu.cn).

2015 年起《计算机研究与发展》双月将固定领域专题

致广大读者和作者：

本刊从 2015 年起将双数期约 1/2 版面固定为某个领域，每年将策划该领域的一个热点主题进行集中报道。具体的征文通知将在专题发表前 6 个月发布，请关注期刊网站！

此外，本刊依然欢迎自由来稿。谢谢！

具体领域分布及执行领域编委如下：

刊期	领域	领域编委	投稿方式	截稿日期
2 期	软件技术(含数据库)	孟小峰 xfmeng@ruc.edu.cn 中国人民大学	期刊网站投稿， 备注填写“年+专题名称”	上一年 10 月 1 日左右 (以具体征文通知为准)
4 期	网络技术	林闯 chlin@tsinghua.edu.cn 清华大学	期刊网站投稿， 备注填写“年+专题名称”	上一年 12 月 1 日左右 (以具体征文通知为准)
6 期	体系结构	刘志勇 zyliu@ict.ac.cn 中国科学院计算技术研究所	期刊网站投稿， 备注填写“年+专题名称”	当年 2 月 1 日左右 (以具体征文通知为准)
8 期	人工智能	周志华 zhouzh@nju.edu.cn 南京大学	期刊网站投稿， 备注填写“年+专题名称”	当年 4 月 1 日左右 (以具体征文通知为准)
10 期	信息安全	曹珍富 zfcao@sjtu.edu.cn 上海交通大学	期刊网站投稿， 备注填写“年+专题名称”	当年 6 月 1 日左右 (以具体征文通知为准)
12 期	应用技术	郑庆华 qhzheng@mail.xjtu.edu.cn 西安交通大学	期刊网站投稿， 备注填写“年+专题名称”	当年 8 月 1 日左右 (以具体征文通知为准)