

Analysis of MOOCs Courses Dropout Rate Based on Students' Studying Behaviors

Fang-jie LIU^{1,a}, Lu WANG^{2,b,*}, Yi-jun QIAN^{3,c} and Yu-jie WU^{4,d}

^{1,2,4}School of Mathematics, Southwest Jiaotong University, Chengdu, China

³School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China

^a530024997@qq.com, ^bwanglu@home.swjtu.edu, ^cqyjboy@126.com,
^dJaywu1996@outlook.com

Keywords: Analysis of Variance, Logistic Regression, Social Learning Networks, Dropout rate.

Abstract. This paper investigates course-dropout rate based on users' various behaviors-video-watching, discussion and question-answering behavior in MOOCs. We mainly analyze two problems: one is the comparison of average duration of studying on MOOC among users who either drop out or complete the course, and the other is to identify the exact relationship between these three behaviors and course completion rate, which in other words to predict the completion rate by means of students' studying behaviors. By applying analysis of variance and logistics regression, we draw several conclusions:(1) In high dropout rate courses, the difference between the behavior of all users regardless of whether they have dropped out or completed the course is not statistically significant, however, in contrast, among low dropout rate courses, there exists magnificent difference between the behaviors among those two kinds of users. (2) It is common both in high and low dropout rate courses that all users tend to spend less time studying as time goes on. (3)Considering all these behaviors, only video-watching behavior significantly and positively determines the completion rate: Considering the findings above, we suggest that it is useful to make improvement on videos to keep students motivated in the course, thereby boosting the completion rate.

Introduction

Over the past few years, Internet technology development has greatly influenced the way we learn. One of the most representative innovation is the Massive Open Online Courses (MOOCs). MOOCs have been mentioned frequently in the scientific literature recently as a new way of providing training which is attracting millions of students across the world and forcing universities to reformulate their online education courses.[1]Considered as a revolution to the education, it provides a widespread, convenient and flexible study method, which could help students absorb advanced knowledge globally.

However, although the MOOCs are developing flourishingly, the problem it faced is remarkable. It is noticed that MOOCs not only has high enrollment rate, but also has high dropout rate. There are quite few students who could stick to completing the courses. [2] Research revealed that one salient feature of MOOCs is high dropout rate, with typically less than 10% of students that initially enroll in a course receiving a certificate of completion in the end. [3]Considering the current situation, we maintain that it is necessary to study the current courses' completion rate in the MOOCs platform. Through analyzing the differences of learning behaviors between drop-out

students with those who successfully complete the same course and studying the differences of learning behaviors among different courses with various dropout rates, we could know about the courses' features and thus proposing effective measures to improve completion rate for the sake of promoting further development.

Research Method

Various work has been done analyzing student forum participation in MOOCs, for example, Sinha processed social network analysis (SNA) to identify students who were actively participating in course discussions and those who were potentially at risk of dropping off. [4] However, there are few articles that analyze the exact relationship between users behaviors and course completion rate. Therefore, the aim of this study is to analyze the learning behavior of users who either drop out or complete the curriculum. The questions addressed are:

- A) What is the difference between the behavior of students who either drop out or complete the same course?
- B) What is the difference between courses with distinct dropout rate based on users' behaviors?

To investigate these questions, we will use data from one of the typical MOOCs platform *xuetangzaixian*. The data consists of over 120,000 logs records of 39 different courses. By analyzing the data, we pick up video-watching time, discussion time and problem-answering times as our indexes reflecting a student's extent of attention. For Question A, we first select two courses with different completion rate, and categorize students into two groups according to whether they have completed the course or not. Then we calculate each student's video watching time, discussion time, question-answering time according to the time that they have logged in. Finally, by employing analysis of variance, we successfully obtain the results. Moreover, we apply logistic regression to these two courses in an attempt to predict the dropout rate based on the three indexes that we have picked up. For Question B, because only the video watching behavior is statistically significant in the logistic regression, we thus choose the time students spent watching videos as factor that contribute to the different dropout rates among the courses, and apply analysis of variance to reach our conclusion.

Analysis and Results

Question A

In this part, we will focus on Question A. Firstly, we will select two typical courses through calculating their respective dropout rate— one with a high dropout rate of 91% and the other with low dropout rate of 64%. Then we calculate the time students spent watching videos, joining discussions and answering questions, as well as calculating the average duration of these behaviors according to the number of time they logging in to study. The average time of behaviors of the two courses are given in Figure1

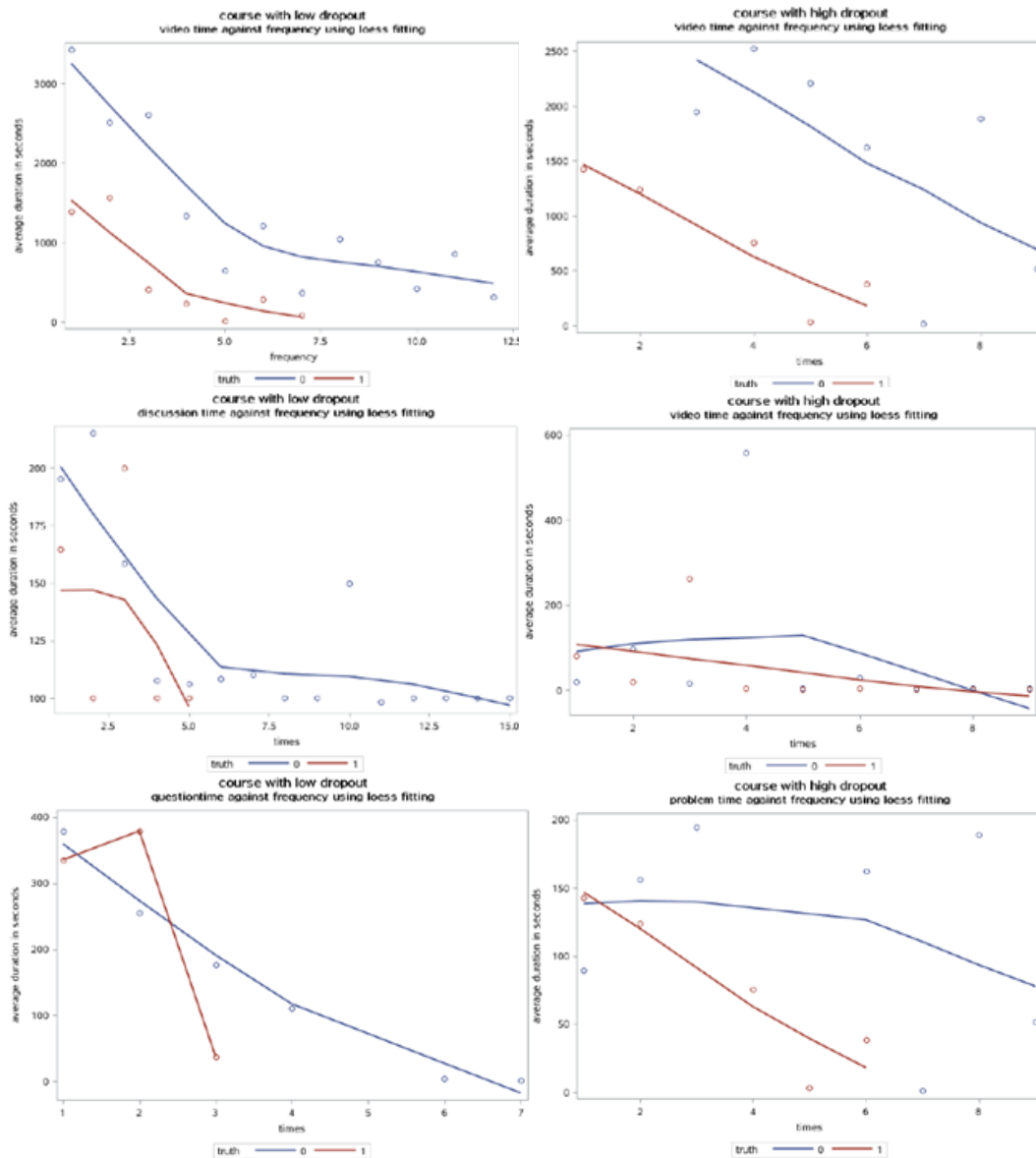


Figure1. Various studying behaviours against frequency using loess fitting

The figures listed in the first column represent behaviors of the course with low dropout rate while the figures listed on the right side represent the course with high dropout rate. And 0 represents those who complete the course while 1 represent those who drop out in all pictures.

For the three behaviors of students in the course with low dropout rate, we can clearly see that the time students spend watching videos are significantly different between those who completed the course with those who did not, moreover as time went on, students from both groups had a tendency to spend less time watching videos which is in correspondence with common sense. However, as for the time that students spent joining discussions and answering questions, there exists less significant difference between these two groups of students. Furthermore, the time that students spent in completing discussions and questions was far less than that in watching videos, indicating that there time students took to watch videos play a main role in determining the course's dropout rate. In order to test whether there exists magnificent difference between the time students of different groups spent participating in these three behaviors, we apply analysis of variance to test our

hypothesis. Shown in Table 1 is the statistic of ANOVA, where video time is the variable.

Table 1. Analysis of variance

Source	Freedom of degree	Quadratic sum	Mean square	F	Pr>F
Model	2	8405086784.2	4202543392.1	30.8	<.0001
Deviation	1240	168783811061	136115976.6		
Total	1242	177188897845.2			

The test result confirms our hypothesis that it is indeed the case students who complete the course spend more time watching videos. However, as for the time of students spent discussing and answering questions, there is no trace of difference, which further confirms our guessing that the time students spent watching videos is the main contributing factor. In order to precisely predict the completion rate (which equals to 1 minus the dropout rate), we employ logistic regression and the result is as follows:

$$\ln\left(\frac{P}{1-P}\right) = -2.0427 + 0.00706\text{video} + 0.000019\text{discussion} + 0.00083\text{problem}$$

$$t = (22.937) \quad (4.3278) \quad (1.2398) \quad (2.3291) \quad (1)$$

As for the situation of high dropout rate courses, again, we recognize similar patterns with the course of low dropout rate, i.e. students who complete the course spent significantly more time watching videos while the time that students spent discussing and answering questions are not that significant compared with the time that students spent watching videos. The ANOVA test further confirms our hypothesis that the time that students took to watch videos is different under the confidence level of 5% but the time that students spent discussing and answering questions are not significant under the confidence level of 5%.

In conclusion, students from different courses all show similar studying patterns: tending to attaching much more time to watching videos while ignoring discussing and answering questions. Moreover, students who drop out of the course half way inevitably invest less time in studying especially watching videos regardless of the dropout rate of the course. Considering the discovery above, we further investigate whether there are different study behaviors between these two courses according to whether the students have dropped out or not.

Question B

In this section, we mainly investigate whether the study pattern is different between courses of different completion rate. In other words, the question remains to be answered is whether the students who complete the course from these two courses possess different study habit and those who dropout has distinct study pattern.

Likewise, we employ ANOVA to answer this question. The result is shown in Table 2, Table 3 and Table 4:

As can be seen from the table above, the average time that students spent on watching videos, answering questions and joining discussions are not statistically significant when the course itself if the factor in ANOVA test (with an F-value of only 0.5399).

In conclusion, although the courses show very distinct completion rate, the study patterns of students who drop out and those who complete are actually similar, which indicates that it is the nature of the course itself that cause the difference of completion rate.

Table 2. Variance analysis of watching videos

Source	Freedom of degree	Quadratic sum	Mean square	F	Pr>F
Model	2	3405086571.4	1702543285.7	11.24	<.0001
Deviation	1240	187783811061	151438557.31		
Total	1242	191188897632			

Table 3. Variance analysis of answering questions

Source	Freedom of degree	Quadratic sum I	Mean square	F	Pr>F
Truth	1	3348157659	3348157659	22.11	<.0001
Course	1	56928913	56928913	0.38	0.5399

Table 4. Variance analysis of joining discussions

Source	Freedom of degree	Quadratic sumIII	Mean square	F	Pr>F
Truth	1	2292100992	2292100992	15.14	<.0001
Course	1	56928913	56928913	0.38	0.5399

Remark: Truth denotes whether the students have dropped or not.

Conclusions

In this paper, we study the association between users' video-watching behavior, discussion behavior, and answering-problem behavior with course completion rate. In order to explore the relationship, we formalize two factors to describe their study behavior: one is frequency (i.e. the times that students log in to study), the other is average duration of watching video, joining discussion, answering questions. With datasets from MOOCs platform *xuetangzaixian* we accomplished two main goals:

- 1) Identify the difference between the behavior of students who either drop out or complete the same course.
- 2) Identify the difference between courses with distinct dropout rate based on users' behaviors.

Through this study, we reach several conclusions:

- 1) In Both low dropout rate courses and high dropout rate courses, students are inclined to spend more time watching videos while attaching less significance to discussion and questions. Moreover, students who completed the course significantly spend more time watching videos than those who did not regardless of the dropout rate of the two courses.
- 2) The study patterns of courses with different dropout rate are actually similar, which indicates that the dropout rate is more than determined by the nature of the course itself than by the students.

Suggestions

Considering the statistics listed above we propose the suggestions listed below

- 1) Because the behavior of watching course videos play a determining role in whether the student complete the course or not, along with the fact that Udacity, Coursera, Edx all use videos as their key study resources, it is of great significance that course designers should make the videos not only as attractive as possible but also as diversified as possible. Given the fact that the time that students spent on answering questions and joining discussions are lower than that spent on watching videos, designers are also supposed to stitch the videos and after-course questions together perfectly in order to help students grasp the knowledge better by answering questions after each video they watch, which consequently form a smooth process of studying online [5].

2) According to the analysis, the pattern of the time of student's watching video is mainly based on the internal characteristics of the lesson, so while allocating classes the MOOC platform may choose more charming ones.

3) Students spend significantly less time on discussing and asking problems, so the MOOC platform should improve the Q&A system. Providing more opportunity for the students to discuss with teachers and students of their age.

4) According to the terminal data recording students' action while having MOOC, many students can not bear the temptation to browse other irrelevant information on the internet, so the platform may introduce a system in which browsing irrelevant information may cause penalty.

Acknowledgement

This research was financially supported by the National Science Foundation, Projects in Sichuan Education Development Research Center of Key Research Base for Humanities and Social Sciences of Sichuan Provincial Education Department (Grant NO.CJF14014), Student Research Training Program in Southwest Jiaotong University (GrantNO.201610613074).

References

- [1] Castaño C, Maiz I, Garay U. Design, Design, Motivation and Performance in a Cooperative MOOC Course [J], Online Submission. (2015) 22.
- [2] Wang Y, Zhang J A, Zhang B, MOOC: Characteristics Analysis Based on Typical Projects and Its Enlightenment [J], Journal of Distance Education. 2013.
- [3] Brinton C G, Buccapatnam S, Chiang M, et al. Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance [J]. IEEE Transactions on Signal Processing, 64 (2016) 3677-3692.
- [4] T Wu, Y Yao, Y Duan, X Fan, H Qu, Network Seer: Visual analysis for social network in MOOCs, 2016 IEEE Pacific Visualization Symposium (PacificVis) (2016).
- [5] Gao D. A Coolheaded Response to Hot MOOCs: Reflections on the Six Problems of MOOCs [J]. Journal of Distance Education, (2014).
- [6] Brinton C G, Chiang M. MOOC performance prediction via clickstream data and social learning networks [J], (2015) 2299-2307.
- [7] Kellogg S, Edelman A. Massively Open Online Course for Educators (MOOC-Ed) network dataset [J]. British Journal of Educational Technology, 46 (2015) 977-983.