Investigating Interaction of Study Group Based Massive Open Online Courses

Li-Ling ZHOU¹, Lu-Ying Li², Wen YIN³ and Ming-Hua JIANG⁴,*

Wuhan Textile University, Wuhan, China
¹Email:905824058@qq.com, ²liluying46@163.com, ³753086438@qq.com,
⁴Email: jmh@wtu.edu.cn
Corresponding author

Keywords: MOOC, Study group, Team activity, Survival analysis.

Abstract. In the past few years, the rapid development of Massive Open Online Courses (MOOC) has attracted wide attention from the society, and the high dropout rate of learners has become one of the issues of wide concern[1]. A major limitation of the current generation of MOOCS is ignore the advantage for students to make use of study group as resource. In this paper, we have built MOOCROOM platform to extract data and find out the team activity that affect the interest of learners through survival analysis[2]. We valuate an intervention that makes more interactivity in MOOCROOM. From the impact of the biggest team activities to improve the teaching management and study group building to reducing the dropout rate. And this will be helpful for students to complete the study process. We discuss implications for design, limitations of the current study, and directions for future research.

Introduction
Over the past few years, MOOCs have led the new revolution in e-learning, by providing limitless opportunities for thousands of learners to participate in free higher education courses online. Massive Open Online Courses (MOOCs) are online platforms that have gotten a huge amount of press, both positive and negative. MOOCs have scant affordances for social interactivity, and arguably, that social interactivity is not a major part of the participation experience of the majority of participants. One argument in favor of social-oriented MOOCs is that there is accumulating evidence that team interactivity enhances commitment to the course and thus addresses the biggest concern in the negative press about MOOCs, namely the high drop-out rates[3].

To leverage team interaction or team-based learning in MOOCs, it is important to understand the impact of aspects of the design of social affordances in MOOC contexts on the properties of social interactivity that occurs therein and, ultimately, how those properties feature in the overall participant satisfy activity and success. In this paper, we explore properties of study groups social interactivity in a small sample of MOOCROOM and how it relate to aspects of current designs. We suggest directions for future work designing intelligent support for social interactivity in MOOCs.

In this paper, in order to do better research, our team built the MOOCROOM. Many team leaders don't know how to manage their teams, in which case the team will fall apart. We conducted a quantitative study of the team's activities from both live and peer reviews to predict team performance at the end of the semester. Through this article, our findings help to broaden the study of online groups and provide some design advice for MOOCROOM and social MOOCs.

MOOCROOM Based Datasets
In the new MOOCROOM, students work together in a group that encourages positive and peer learning through a team's live stream. However, not all research teams are successful. In the MOOCROOM course we built, more than half of the teams did not submit final team projects, although the teachers were committed to supporting teamwork.
Our MOOC dataset use of the Java programming course which is more focus on. The statistics are shown in Table 1.

<table>
<thead>
<tr>
<th>MOOCROOM Course</th>
<th>student</th>
<th>visitor</th>
<th>forum</th>
<th>period</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java program</td>
<td>7739</td>
<td>10708</td>
<td>128</td>
<td>8</td>
<td>32</td>
</tr>
</tbody>
</table>

Based on the platform of MOOCROOM learners as the research object, according to learners of all activities in the platform, the learner activity can be divided into four categories: Team Leader activity, Member activity, Interactivity activity and Others[4].

1. Team Leader activity: Team leader take part in number of activities, team leader drop-out
2. Members activity: members take part in number of activities, member drop-out
3. Interactivity activity: Interactivity between teachers and students, interactivity in study group. Interactivity between groups and visitor
4. Others: platform rectification, teaching preparation and some other activity

Figure 1 shows the proportion of four types of activity, which account for the largest proportion of activity. It also shows that the core of MOOC is the part of the course that best reflects the learner's activity. Therefore, this paper will extract the learner's activity from the course activity as model input.

Survival Analysis of Study Group based MOOCROOM

Survival Analysis Method

Survival models are a form of proportional odds logistic regression, and they are known to provide less biased estimates than simpler techniques that do not take into account the potentially truncated nature of time-to-event data[5,6]. The survival model we used was based on the existence of some predictors, and the probability of participating in the course was predicted. The estimated weights on the predictors are referred to as hazard ratio. Survival data are generally described and modelled in terms of two related probabilities, namely survival and hazard. The survival probability (which is also called the survivor function) S(t) is the probability that an individual survives from the time origin (e.g. diagnosis of cancer) to a specified future time t. It is fundamental to a survival analysis because survival probabilities for different values of t provide crucial summary information from time to event data. These values describe directly the survival experience of a study cohort.

The hazard is usually denoted by h(t) or l(t) and is the probability that an individual who is under observation at a time(t) has an event at that time. Put another way, it represents the instantaneous event rate for an individual who has already survived to time t. Note that, in contrast to the survivor function, which focuses on not having an event, the hazard function focuses on the event occurring. It is of interest because it provides insight into the conditional failure rates and provides a vehicle...
for specifying a survival model. In summary, the hazard relates to the incident (current) event rate, while survival reflects the cumulative non-occurrence.

More formally, the probability of being online at the period $t_j$, $S(t_j)$, is calculated from $S(t_{j-1})$ the probability of being online at $t_{j-1}$, $n_j$ the number of student online just before $t_j$, and $d_j$ the number of events at $t_j$, by formula:

$$S(t_j) = S(t_{j-1})(1 - \frac{d_j}{n_j})$$

(1)

where $t_0=0$ and $S(0)=1$. The value of $S(t)$ is constant between times of activity, and therefore the estimated probability is a step function that changes value only at the time of each activity. This estimator allows each student to contribute information to the calculations for as long as they are known to be event-free. Were every individual to experience the event (i.e. no censoring), this estimator would simply reduce to the ratio of the number of individuals events free at time $t$ divided by the number of people who entered the study.

Confidence intervals for the survival probability can also be calculated. The Kaplan–Meier(KM)survival curve, a plot of the KM survival probability against time, provides a useful summary of the data that can be used to estimate measures such as median survival time.

**Survival Estimate**

In this section, visualizing the student’s learning process as a timeline[7]. We use survival analysis to validate that when team has no interaction, the team member is more prone to dropout, but the effect of no interaction is stronger.

**Dependent Variable:**

We consider a student to drop out if the current period is his/her last period of active participation unless it is the last course period (i.e. the 32nd course period).

**Control Variables:**

Activity: total number of activities (team blogs, blog comments or messages) a student participated in that period, which is a basic effort measure of engagement of a student.

Group Activity: total number of activities the whole team participated in that course period. Since it is correlated with Activity ($r >0.5$). In order to avoid multicollinearity problems, we only include Activity in the final survival models.

**Independent variables:**

Live: 1 if the group organizes regular online communication in previous periods, 0 otherwise.

Peer evaluation: 1 if members of the group evaluate each other and give directions in the current period, 0 otherwise.

Table 2. Survival analysis results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Haz.ratio</td>
<td>Std.Err.</td>
<td>$P&gt;</td>
<td>z</td>
<td>$</td>
<td>Haz.ratio</td>
</tr>
<tr>
<td>Group Activity</td>
<td>0.652</td>
<td>0.091</td>
<td>0.002</td>
<td>0.637</td>
<td>0.085</td>
<td>0.001</td>
</tr>
<tr>
<td>Live</td>
<td>3.517</td>
<td>0.813</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peer evaluation</td>
<td>2.124</td>
<td>0.576</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Survival Analysis Results

Table 2 reports the estimates from the survival models for the control and independent variables entered into the survival regression. Effects are reported in terms of the hazard ratio, which is the effect of an explanatory variable on the risk or probability of participants drop out from the course forum. Because the activity variable has been standardized, the hazard rate here is the predicted change in the probability of dropout from the course forum for a unit increase in the predictor variable. Students are 35% times less likely to drop-out if they have one standard deviation more activities.

Figure 1 illustrates our results graphically. The solid curve shows survival with user Activity at its mean level. The curve in the down shows survival when Group Activity is at its mean level in the current period. The above curve shows survival when Group Activity is at its mean level and have at least once Live and Peer evaluation in the current period.

Conclusion

In this paper, we propose to organize live and peer evaluation in MOOCROOM courses. First of all, statistical analysis has a great influence on group learning effect. Secondly, the influence of interactive activity can be obtained through survival analysis. Experimental results show that our proposed method can get lower dropout rates.

In the future, we can have a better in-depth study of the MOOCROOM platform, in the meantime, can also migrate research result to other MOOC platforms.

Acknowledgement

This research was financially supported by the Creative Funds for Wuhan Textile University.

References


