Peer Influence on Attrition in Massive Open Online Courses

Diyi Yang
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
diyiy@cs.cmu.edu

Miaomiao Wen
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
mwen@cs.cmu.edu

Carolyn Rose
Carnegie Mellon University
5000 Forbes, Pittsburgh, PA
cprose@cs.cmu.edu

ABSTRACT

In this work, we investigate the role of relational bonds in keeping students engaged in online courses. Specifically, we quantify the manner in which students who demonstrate similar behavior patterns influence each other's commitment to the course through their interaction with them either explicitly or implicitly. To this end, we design five alternative operationalizations of relationship bonds, which together allow us to infer a scaled measure of relationship between pairs of students. Using this, we construct three variables, namely number of significant bonds, number of significant bonds with people who have dropped out in the previous week, and number of such bonds with people who have dropped in the current week. Using a survival analysis, we are able to measure the prediction strength of these variables with respect to dropout at each time point. Results indicate that higher numbers of significant bonds predicts lower rates of dropout; while loss of significant bonds is associated with higher rates of dropout.

Keywords

Student Dropout, Peer Influence, MOOCs

1. INTRODUCTION

Massive Open Online Courses (MOOCs) such as those run through Coursera[1] have rapidly moved into a prominent place in the media. One notable problem with current MOOCs is the extremely high attrition, which inspires us to investigate what factors might affect student attrition [3, 2]. Prior work has been conducted to explore the connection between participation patterns in the discussion forum and student dropout. However, little attention has been paid specifically to the formation of relationship bonds during participation or how those relationship bonds influence the continuing commitment to the course. In this work, we investigate the connection between relational bonds and commitment to the course, which we refer to as **Peer Influence**. We leverage a statistical analysis technique referred as survival analysis to quantify the extent to which the informal relationships between students influence their dropout. First, we design five alternative operationalizations of relationship bonds based on patterns of communication and common topic focus in posts. We validate these five operationalizations as a single scale that enables us to construct three variables describing important aspects of the experience students have in the MOOC social environment.

[1]https://www.coursera.org/

2. PEER INFLUENCE EXPLORATION

In this section, we describe five separate operationalizations of relationship formation that we use to infer peer bonds.

- *Reply Interaction*: Who replies to whom is an explicit and direct indication of students' intention to socialize with specific other students. We generate a peer candidate set for a student based on the number of replies they have contributed to the posts of each of the other students as a reflection of their connection with them.
- *Co-occurrence Evidence*: Even though students are not talking to others directly, it is possible that they benefit from others' posts when they are exposed to them on the the threads they post to. Furthermore, participating in the same thread might also indicate that students share similar interests. Here, peer candidates are generated by ranking students based on number of common threads they have participated in.
- *Community Connection*: The participation patterns of students can be viewed as a social network graph, and we can use a graph partition method to identify subgraphs where students are located within that representation. Then we count the students within the same subgraph as more closely associated with one another than they are to others outside of the subgraph.
- *Topic Modeling*: Users who share interests usually talk about similar things. Similarity in topic focus can be treated as membership in an implicit interest defined subcommunity. To capture potential relations along this dimension, we use Latent Dirichlet Allocation [1] as a model to select students' peer candidates based on similarity of their topic distribution.
- *Cohort*: Cohort tells when the student has started their participation in the course and could be regarded

Aggregate Variables	Involved Original Variables
Cur	R_{cur}, C_{cur}, O_{cur}, T_{cur}, M_{cur}
$Prev$	R_{prev}, C_{prev}, O_{prev}, T_{prev}, M_{prev}
Num	R_N, C_N, O_N, M_N

Table 1: Variables organized into sets for constructing aggregate measures

as a proxy for their commitment (since students who join later tend to be less committed)[3]. Here, we generate the peer candidates for students based on their registration time.

3. VARIABLES DESIGN

Building on our five defined relationship measures, we formalize what relationship loss means by constructing three separate variables for each bond definition as follows. *Dropout in current week* (Cur), captures how many significant relations of student u dropped out in the current week; *Dropout in previous week* ($Prev$) captures how many significant relations of student u dropped out in the previous week; *Number of friends* (Num) describes how many significant bonds a student u has.

For each operationalization, we construct the three variables described above. Specifically, for reply bonds, we have R_{prev}, R_{cur}, R_N, representing the $Prev$, Cur, Num variables under the category of reply bonds; For co-occurrence bonds, O_{prev}, O_{cur}, O_N are gained; for community connection, we construct C_{prev}, C_{cur}, C_N; for topic modeling, we get T_{prev}, T_{cur} and discard T_{Num} which is the same for all students; for the motivation cohort, M_{prev}, M_{cur}, M_N are extracted. Those 14 variables are organized into three aggregate variables by simply averaging the same types of variables as shown in Table 1.

4. METHOD

The course we use to conduct the experiment is a Python programming course[2]: 'Learn to Program: The Fundamentals'. It has 3590 students who are active in the discussion forum, 24963 posts in total across the eight weeks. After aggregating the 14 variables into Cur, $Prev$, Num as described above, we then use those as input and conduct survival analysis to investigate how the three aspects influence the dropout of students. From the result presented in Table 2, we can observe that, (1) Students are around four times more likely to dropout if the number of their relation loss of close peers Cur are higher than average; (2) a student is 62% more likely to drop out if his/her relation loss $Prev$ is one standard deviation larger than average; (3) Comparably, the number of close peers Num indicates that the more close peers one student has, 74% less likely this student will drop out.

Figure 1 illustrates our result graphically. The middle solid curve shows survival with the number of Cur, $Prev$ and Num all at their mean level. The top curve above this middle one shows the survival when the number of close peers Num is one standard deviation above the mean (High), keeping the Cur and $Prev$ at their mean level. It indicates

[2]https://www.coursera.org/course/programming1

Variable	Hazard Ratio	Std. Err	Z	P>|Z|
Cur	5.05	0.264	35.69	0.000
$Prev$	1.62	0.035	22.09	0.000
Num	0.26	0.014	-26.65	0.000

Table 2: Constructed Variables on Python Course

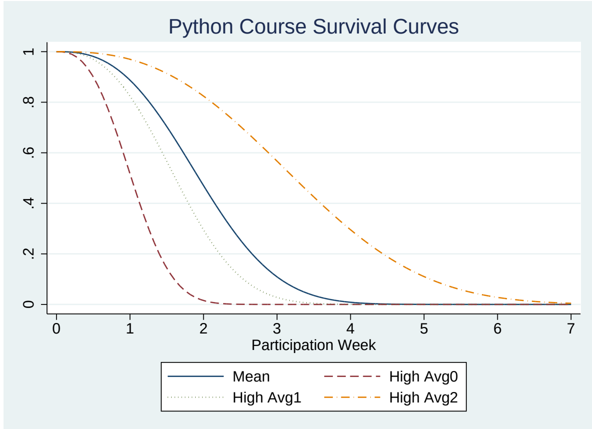

Figure 1: Survival Curves on Python Course

that higher Num is correlated with a longer continuing participation in the course. The bottom two curves show the survival when the dropout number of close peers in previous week or dropout number of such peers in the current week are both one standard deviation above the mean, keeping the other variables at their mean level. This reflects the influences of Cur and $Prev$ again – the more close peers a student loses, the less likely he/she will continue participating in the course forum.

5. CONCLUSION

In this work, we propose to measure peer relations in the MOOC forums and explore how such relations influence student dropout. Reliable operationalizations of relations are constructed as well as variables corresponding to relationship loss. Via modeling of survival analysis , we find strong evidence that relationship loss is an important factor contributing to attrition. These results argue that attention to fostering a positive and supportive social environment could be an important direction for future MOOC development.

Acknowledgement

This work was supported in part by NSF grants IIS-1320064 and OMA-0836012.

6. REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, Mar. 2003.

[2] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé III, and L. Getoor. Modeling learner engagement in moocs using probabilistic soft logic.

[3] D. Yang, T. Sinha, D. Adamson, and C. P. Rose. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Workshop on Data Driven Education, Advances in Neural Information Processing Systems 2013*, 2013.