Social Factors that Contribute to Attrition in MOOCs

Carolyn Penstein Rosé
Language Technologies Institute and Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
cprose@cs.cmu.edu

Ryan Carlson, Diyi Yang & Miaomiao Wen
Language Technologies Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
{rcalrson,diyiy,mwen}@andrew.cmu.edu

Lauren Resnick, Pam Goldman & Jennifer Sherer
Institute for Learning
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15213
{resnick,peg8,jzsherer}@pitt.edu

ABSTRACT
In this paper, we explore student dropout behavior in a Massively Open Online Course (MOOC). We use a survival model to measure the impact of three social factors that make predictions about attrition along the way for students who have participated in the course discussion forum.

Author Keywords
Attrition, Survival Modeling

ACM Classification Keywords
H5.3. **[Information Interfaces and Presentation]:** Group and Organization Interfaces.

INTRODUCTION
Current research on attrition in MOOCs [2] focuses on summative measures rather than focusing on the question of how to create a more socially conducive environment. Understanding better the factors involved in the struggles students encounter along the way can lead to design insights for the next generation of more successful MOOCs. As one preparatory step, we explore how an unsupervised graphical model (a Mixed Membership Stochastic Blockmodel [1]) is able to identify emerging social structure that predicts dropout along the way in one specific MOOC. This model provides one of three social variables we evaluate with respect to predictive power in connection with dropout along the way in one specific Coursera.org MOOC, namely Accountable Talk™: Conversation that Works, launched by the University of Pittsburgh's Institute for Learning in Fall of 2013.

L@S 2014, March 4–5, 2014, Atlanta, Georgia, USA.
ACM 978-1-4503-2669-8/14/03.
http://dx.doi.org/10.1145/2556325.2567879

METHOD

Data and Modeling Techniques
The data used for the analysis presented here was extracted by permission from Coursera.org using a screen scraping protocol and focuses only on participation in the discussion forums. While over 60,000 students signed up for the course, only about 25,000 students accessed the course materials at least once. Of those students, only about 5% of the students ever posted to the discussion forums. At the last scraping, 4,709 posts had been contributed. This analysis focuses on the authors of those posts.

Two types of models were used in our analysis. First, in order to obtain a soft partitioning of the social network of the discussion forums, we used a Mixed Membership Stochastic Blockmodel (MMSB) [1]. The advantage of MMSB over other graph partitioning methods is that it does not force assignment of students solely to one subcommunity. The model can track the way students move between subcommunities during their participation. In our representation of the social network, each week of participation was treated as a disjoint network. This enabled the model to view snapshots of coordinated engagement over time so that it would not be biased to assume consistency of social engagement over time but would be able to find it where it occurred. The model was limited to identify three subcommunities because the amount of data was small. In order to evaluate the impact of social factors on continued participation within the MOOC context, we used a survival model, as in prior work [3,4]. Survival analysis is known to provide less biased estimates than simpler techniques (e.g., standard least squares linear regression) that do not take into account the potentially truncated nature of time-to-event data. In a survival model, a prediction about the likelihood of a failure occurring is made at each time point based on the presence of some set of predictors. The estimated weights on the predictors are referred to as hazard ratios. The hazard ratio of a predictor indicates how the relative likelihood of the failure (e.g., dropout) occurring increases or decreases with an increase or decrease in the associated predictor.

Results

We explored the three types of variables in our model in a stage based way and retained in the final model only those variables that made a significant prediction. Similar to prior work [4], we began with binary cohort variables that indicated which week of the course a student began their active participation. Because the course ran for just over 7 weeks, there were 7 such binary variables. Only the one that indicated that a student began their active participation in the first week of the course made a significant prediction. Table 1 indicates a hazard ratio of .65, which means that students who began their active participation in the first week of the course were 35% less likely to drop out on each time point than the population average.

Table 1: Results of the survival analysis measuring the impact of social factors leading to attrition

Independent Variable	Hazard Ratio	P value
Week 1 Cohort	.65	P < .01
Authority Score	.00	P < .05
Subcommunity 3	9.9	P < .05

The second set of variables we explored were derived from the social network constructed from posts contributed within each week (as opposed to the complete network that existed at the time point). We explored a variety of social network analysis measures, however only Hub and Authority scores yielded a significant prediction, consistent with prior work [4]. Since Hub and Authority scores are highly correlated with one another, we only retained the most predictive one in the model, namely Authority. Table one indicates an almost 0 positive value for the hazard ratio, which indicates a nearly 100% likelihood of dropout on the next time point for students who have an authority score on a week that is a standard deviation larger than average in comparison with students who have an average authority score. Thus, students who serve as authorities in the community appear more committed to the community.

Finally, we explored the indicators that came from the MMSB model. Only one of the three identified subcommunities made a significant prediction about attrition. In particular, students who participated in that subcommunity on a week were nearly ten times more likely to drop out on the next time point that average students. A posthoc analysis verified that the particular community was not unusual with respect to the other variables that predicted dropout as well as not being specific to a period of time, level of intensity of participation, or particular topic related subforum. Thus, the results suggest that the pattern of attrition was related to the engagement of the particular students involved with one another that created the conditions leading to drop out.

CONCLUSION

In this paper we have explored three types of factors, all of which make significant predictions related to dropout along the way in the Accountable Talk MOOC course. The first two factors were already explored in prior work in the context of a different course [4]. This analysis stands as a confirmation of generality of the result. What is new about our analysis is that an indicator that we are able to obtain through unsupervised soft partitioning of the social network of the MOOC makes predictions about dropout, indicating that participation in one particular emergent subcommunity in the MOOC predicts that dropout on the next time point is ten times more likely than for average students not participating in that subcommunity. There could be many interpretations of this finding. Further analysis might reveal some characteristic about norms for participation in that subcommunity that was demotivating for students. In so far as our attempts to identify features that distinguish that subcommunity's behavior from others have not revealed such factors, our current working interpretation is that this result tells us something about the influence students have on one another. As students participate in the MOOC, they begin to form virtual cohorts of students who are moving at a similar pace, are at a similar place in the course, and are engaging with the material in similar ways. If students begin to see others in their cohort leaving, they may find the environment less supportive and engaging and may be more likely to drop out in turn. The results suggest that an important direction for future research in MOOCs is to model the emergent social structure so that we can better understand the influences students have on their emerging cohorts. With greater insight into these social processes, MOOC designers will be in a better position to create affordances that foster a more supportive environment.

ACKNOWLEDGMENTS

This research was supported in part by NSF grant SBE 0836012 and Gates Foundation Grant: The MOOC Research Initiative.

REFERENCES

1. Airoldi, E., Blei, D., Fienberg, S. & Xing, E. P. (2008). Mixed Membership Stochastic Blockmodel, *Journal of Machine Learning Research, 9(Sep):1981--2014, 2008*
2. Jordan, K. (2013). *MOOC Completion Rates: The Data*. Retrieved 23 April 2013.
3. Wang Y, Kraut R, and Levine J. M. (2012) To stay or leave? the relationship of emotional and informational support to commitment in online health support groups. In *Proceedings of Computer Supported Cooperative Work*, p. 833-842.
4. Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, Tune in, Drop out: Anticipating student dropout in Massive Open Online Courses, *NIPS Data-Driven Education Workshop*