Dropout Prediction in MOOCs using Learner Activity Features

Authors

Sherif Halawa
halawa@stanford.edu
Dept. of Electrical Engineering
Stanford University USA

Daniel Greene
dkgreene@stanford.edu
School of Education
Stanford University USA

John Mitchell
john.mitchell@stanford.edu
Dept. of Computer Science
Stanford University USA

Tags

teacher inquiry into student learning, learning design, learning analytics, orchestration, formative assessment

While MOOCs offer educational data on a new scale, many educators have been alarmed by their high dropout rates. Learners join a course with the motivation to persist for some or the entire course, but various factors, such as attrition or lack of satisfaction, can lead them to disengage or totally drop out. Educational interventions targeting such risk factors can help reduce dropout rates. However, intervention design requires the ability to predict dropouts accurately and early enough to allow for timely intervention delivery. In this paper, we present a dropout predictor that uses student activity features to predict which students have a high risk of dropout. The predictor succeeds in red-flagging 40% - 50% of dropouts while they are still active. An additional 40% - 45% are red-flagged within 14 days of absence from the course.

1. Introduction

Over the past two years, MOOCs have offered educational researchers data on a nearly unprecedented scale. In addition, since MOOCs allow students to join and leave freely, they have enabled new investigations into when and how students voluntarily engage with online course material.

One consequence of the availability of voluntary MOOC data is that researchers can attempt to predict when a student will stop visiting the course based on his or her prior actions. The ability to predict dropout offers both short-term and long-term value. In the short term, predicting dropout helps instructors to identify students that are in need of scaffolding, and to design and deliver interventions to these students. In the longer term, dropout prediction can provide valuable insights into the interactions between course design and student factors. For example, studying the relationship between student working pace and dropout across different courses can provide insight into the features of a course that make it more or less compatible with slow-paced students.

In the short term, the goal of intervention design and delivery defines several bounds on a practically useful dropout prediction model. For the model to be actionable, the instructor needs to know:

- Who is at risk of dropout and who is not: If the model cannot accurately identify high-risk students, then instructors obviously run the risk of sending interventions to the wrong students.
- When the student activity starts exhibiting patterns predictive of dropout: The sooner we can detect dropout risk, the sooner we can intervene. If an intervention is sent too late, it may be less effective.

In order to help instructors to identify high-risk students in a timely manner, we have developed a dropout prediction model that scans student activity for patterns we have found to be strongly predictive of dropout. Once a student starts exhibiting such patterns, the predictor red-flags the student, alerting the instructor or LMS.

This paper is organized as follows: Section 2 provides a brief account of factors from the education literature that we believe affect student persistence in MOOCs. Sections 3 and 4 establish required definitions for dropout and what it means to successfully predict it. The predictor design is discussed in Section 5. Section 6 presents performance results that illustrate the strengths and weaknesses of our prediction model. Conclusions and future work are presented in Section 7.

Persistence Factors and Dropout

In this paper, we only develop our model for students who have joined in the first 10 days of the course and have viewed at least one video. We chose this cutoff because we expect instructors and researchers to develop interventions within the course materials, which would thus only be seen by students with some initial presence.

Given this cutoff point, what factors influence dropout? MOOC dropout is exceptionally heterogeneous (Breslow, Pritchard, DeBoer, Stump, Ho, and Seaton, 2013). Put simply, students have different goals and intentions that interact and change over time, and because of the low cost of entry and exit for MOOCs, the decision to leave can easily be triggered by any number of factors in a student's life. As Lee and Choi (2011) noted, these

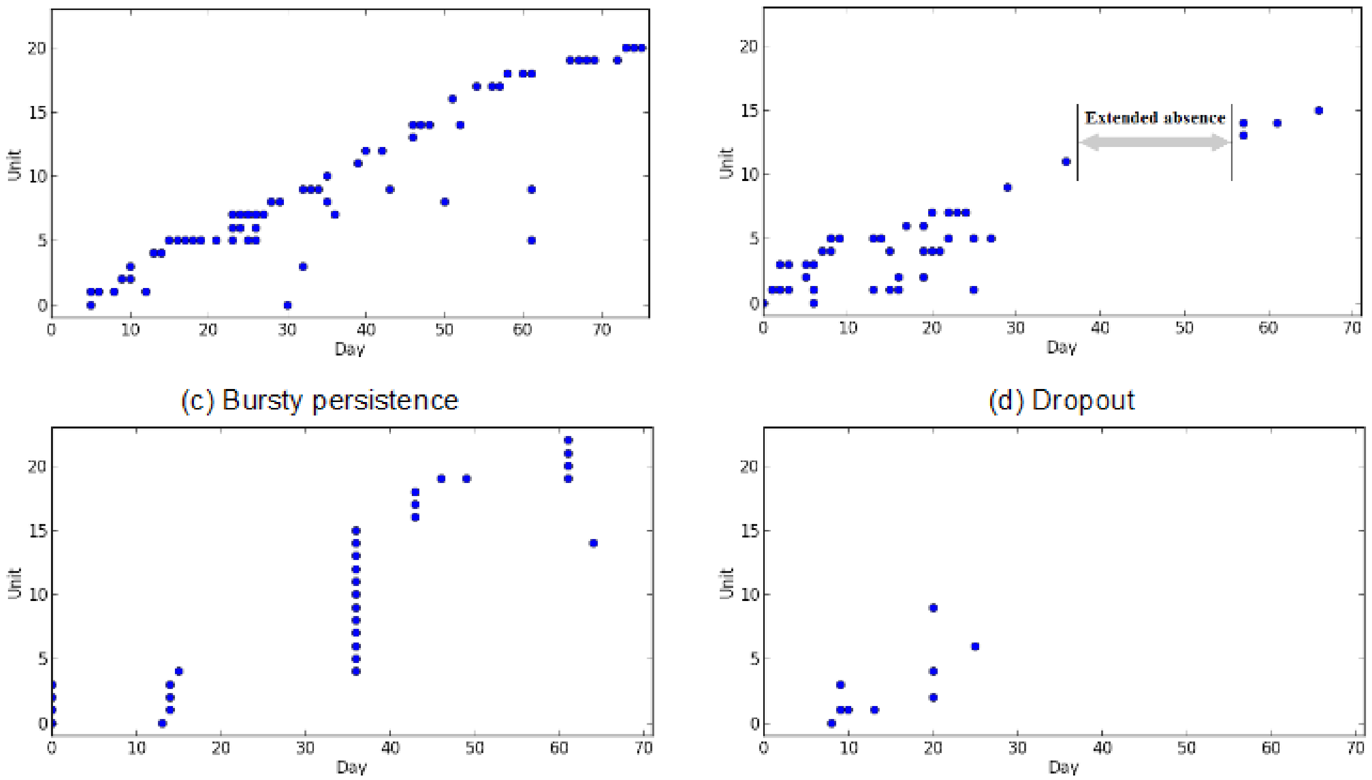


Figure 1. Four common persistence patterns that represent the majority of MOOC students

factors can be roughly divided into internal motivational factors (influencing a student's desire to persist) and external factors like outside life commitments (Rovai (2003)). External factors are practically impossible to intervene upon, and most are also virtually impossible to detect purely through the digital traces of behavior data on a website. They require survey questions such as "Are you taking this course while maintaining a full-time job?" In this paper, we focus entirely upon behavior data that are collected from a learner's interaction with the platform.

Focusing on internal factors, ability is perhaps the most obvious internal predictor of student performance and persistence. Across a wide range of academic settings, low-performing students tend to drop out more frequently than high-performing ones (Hoskins & Van Hooff, 2005). However, the effects of ability on dropout are mediated by self-perceived self-efficacy – the degree to which a student believes that he or she can achieve a particular academic goal. Self-efficacy has been identified as a central construct in motivational models, and self-reported self-efficacy is a strong predictor of academic persistence and performance (Zimmerman, 2000). Students who believe that they can achieve an academic goal are more likely to do so, and students judge their own self-efficacy from their own interpretations of their performance and from social cues (Bandura, 1994). We therefore might expect performance feedback to be an important predictor of dropout.

Students also vary widely in their ability to self-regulate their own learning, a skill set that is particularly important in learning environments like MOOCs with low entrance and exit costs and little external feedback. Researchers have defined taxonomies of self-regulation skills (Zimmerman, 1990), such as time management, self-teaching methods, and metacognitive evaluation of one's own understanding. These skills have been shown to recursively influence learning outcomes, motivation, and further self-regulation (Butler & Winne, 1995).

Other factors affecting dropout include students' level of interest in the material that they are learning. Lack of interest can cause students to dedicate less time to the course, leading them to skip pieces of content, disengage from assessments, or simply proceed through the content at a slow pace. However, pacing and engagement are also affected by external factors. The amount of time a student can allocate the course depends on what other activities the student is involved with in her life (Rovai (2003), Tinto (2006)). It can be challenging to decide whether a drop in persistence is caused by a drop in interest, or by some external factors. In such situations, it is useful to try to elicit more information from the student herself through the use of surveys.

We emphasize that this work is the start of a long process of linking individual factors to student participation, but as a first approximation, we assert that any accurate predictor of student dropout will necessarily be tapping into both internal and external factors.

Defining Dropout

Before discussing our prediction model, we need to present the definition of dropout that we used in this work. We have defined dropout so that it includes any student who meets one of the following two conditions:

1. The student has been absent from the course for a period exceeding 1 month.

2. The student has viewed fewer than 50% of the videos in the course.

Our choice to coin the first condition based on total absence time rather than the last time the student visited the course was the result of a study we undertook to understand what common persistence patterns students follow, and which patterns seem to correlate with drops on certain performance measures. We generated activity graphs for thousands of students from different MOOCs, and were able to identify the 4 common patterns illustrated in Figure 1. Each graph shows which units of content the student visited (viewed any of the unit's videos or attempted any of its assessments) on each day of a course.

Class (a) students visit the course once every few days at most. They usually spend several days on each unit. Class (b) students follow a similarly smooth trajectory, except that there are one or more "extended absences", defined in this work as absences of 10 days or longer after which the student continues from where she stopped previously. The reason for choosing a 10 day threshold is that it separates students who have periodic leaves (e.g.: students who only visit the course on weekends) from students whose persistence changes from continuous to sudden absence and then back to continuous. Class (c) students only visit the course occasionally, and usually sample content from different units each day they visit the course. Selectors (students who view only a selected subset of videos or units), mostly belong to this class. Class (d) students start off

Table 1. Performance comparisons between students of different absence periods for a MOOC

Total absence	Percentage of students*	Median percentage videos viewed	Median percentage assessments taken	Final exam entry rate	Mean final exam score
Less than 2 weeks	37%	77%	62%	66%	71%
2 – 3 weeks	36.4%	62%	60%	64%	68%
3 weeks – 1 month	13.8%	44%	33%	42%	61%
Longer than 1 month	12.8%	21%	17%	13%	46%

* The denominator is the sum of the numbers of students in the 4 groups in this table.

as continuous or bursty visitors, but disappear totally after a certain point before the end of the course.

The analysis revealed that, just like complete dropout after a certain time causes the student to miss a part of the course content, students who have been absent for some time and then return tend to perform worse than class (a) students on many measures, as demonstrated by Table 1. For most of the courses we analyzed, the student's ability to complete videos and assessments as well as the final exam entry rate and performance dropped as the total absence period lengthened.

We consistently observed drops in all of the performance indicators in the table across different courses for students in the third and fourth groups. Our choice was to use the more tolerant threshold of 1 month for our dropout definition.

Dropout Prediction Merit

Our dropout predictor can be implemented as a LMS component that is run periodically (e.g. once every midnight). Every time it is run for a course, the predictor is applied once for each learner

$$pred_l = \begin{cases} 1 & \text{if student } l \text{ is believed to be at risk of dropout} \\ 0 & \text{otherwise} \end{cases}$$

in the course. The predictor analyzes the course activity for learner l and produces the binary output:

The main goal behind dropout prediction is to enable delivery of interventions to red-flagged students (those predicted to be at-risk). This goal must be the basis on which merit is defined.

As with any other predictor, accuracy (the ability of the predictor to accurately predict whether or not a student is going to drop out) is a main criterion. In a course where no treatment of any kind was performed on high-risk students, we have ground truth data (who persisted in the course and who dropped out). Based on the prediction and whether or not the student actually dropped out, four classes of students exist:

1. True negatives (TN): Students who were never red-flagged, and never dropped out
2. False negatives (FN): Students who were never red-flagged, but dropped out
3. False positives (FP): Students who were red-flagged, but never dropped out
4. True positives (TP): Students who were red-flagged, and dropped out

In order to ensure that the sizes of these classes truly reflect the accuracy of the predictor, it is important to ensure that the prediction process has no induced effects on the course or students. Hence, all analysis and discussion must be restricted to courses where no dropout risk information was communicated to the student, and no persistence or performance interventions were implemented.

We can now compute the following traditional quantities:

$$Recall(R) = \frac{|TP|}{|TP| + |FN|} \quad , \quad Specificity(S) = \frac{|TN|}{|TN| + |FP|}$$

Recall measures the predictor's ability to have correctly red-flagged every student who will drop out of the course. Specificity is a measure of the predictor's success in keeping students who will not dropout unflagged. Statistical merit requires the predictor to have high values of R and S.

This, however, is not the only relevant criterion. Practical merit of the predictor also requires that high-risk students be red-flagged early enough to enable timely delivery of interventions. The following prediction rule

Three days before the end of the course, red-flag every student who has been absent for the last four weeks.

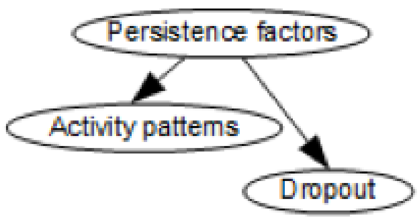
will yield a predictor with excellent specificity and recall but too little practical value because it leaves no time window for intervention.

Predictor Design

Even though activity patterns and dropout decision are two distinct constructs, we believe that influence flows between them, as described by the following claim, which is the main principle underlying our predictor design.

Design Principle

Since a student's activity patterns and dropout probability are both affected by his or her degree of possession of different persistence factors, a flow of influence potentially exists between the two, which may allow the use of activity patterns to predict dropout.



Utilizing student activity to predict dropout might imply that our predictor only operates on a student for as long as she is active in the course. Nonetheless, if some unflagged student goes absent for an alarmingly long period, it is still desired to deliver an intervention. Thus, our "integrated predictor" consists of two components, as illustrated in Figure 2.

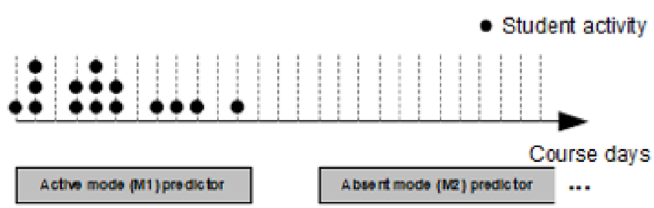


Figure 2. Active mode predictor switched out and absent mode predictor switched in after the student has been absent for an extended number of days.

1. Active mode (M1) predictor: Operates while the student is still active. It analyzes student activity, looking for patterns that suggest lack of motivation or ability. It continues to operate on a student as long as she is performing new activity.

2. Absent mode (M2) predictor: Operates once the student has been absent for a certain time period. It uses the number of days for which the student has been absent to evaluate the probability that the student is heading for a dropout.

Active mode (M1) predictor

This predictor uses the following simple routine to determine whether or not the student should be red-flagged:

1. Compute scores for certain features in the student's activity

2. Make a prediction using each individual feature by comparing its score to a threshold

3. The output prediction is a red flag if any of the individual feature predictors predicts a dropout.

We started off with a large number of candidate features selected based on the persistence factors discussed in Section 2. Candidate features included:

1. Features that suggest a lack of ability, such as low quiz scores or a relatively high rate of seek-back in videos

2. Features that suggest a lack of interest or time, such as: Did the student skip any videos? Does the student re-attempt a quiz if her score on the first attempt was low?

Our goal was to find out which of these features correlate strongly with dropout for the majority of courses. We constructed a course-corpus consisting of 12 courses from different fields of study including mathematics, physics, agriculture, political science, and computer science. We created dozens of variants of our candidate features with different thresholds, aiming to find those that succeed in predicting a substantial number of dropouts with good specificity. Out of all the features and variants, the 4 features listed in Table 2 stood out and were hence selected in the design of the current version of the prediction model.

Note that none of the individual features has a recall that exceeds 0.5. This is acceptable, since students drop out for various reasons. The expectation from a predictive feature is

Feature name	Feature description	S	R
video-skip	Did the student skip any videos in the previous unit? Decision rule: ^{pred} = 1 if yes, 0 otherwise.	0.80	0.31
assn-skip	Did the student skip any assignments? Decision rule: ^{pred} = 1 if yes, 0 otherwise.	0.90	0.27
lag	Is the student lagging by more than 2 weeks? (Some students login to the course every few days, but view too few videos per login. Consequently the student can develop a lag. A lag of 2 weeks, for instance, is when the student is still viewing week 1's videos after week 3 videos have been released.) Decision rule: ^{pred} = 1 if yes, 0 otherwise.	0.86	0.19
assn-performance	Student's average quiz score < 50%? Decision rule: ^{pred} = 1 if yes, 0 otherwise.	0.97	0.007
M1	Combined M1 predictor	**0.77**	**0.48**

Table 2. Median specificity (S) and recall (R) for top ranked features and the combined M1 predictor

to successfully predict a subclass of dropouts without falsely flagging too many persistent students. Recall is of interest for the combined prediction, since a high combined recall suggests that our features have tapped into most of the common dropout reasons. The combined M1 predictor captures almost 50% of the dropouts, falsely flagging almost 1 in every 4 persistent students on the average.

Absent Mode (M2) Predictor

For most students, absences of several days at a time are not uncommon. As the absence lengthens, however, the probability that the student may not continue to persist in the course increases. The job of this predictor is to red-flag a student once he or she has been absent for a certain number of days, called the “absence threshold”.

To determine the optimum threshold, we studied the variation of accuracy with threshold. The threshold was varied from 0 to 3 “course units”, where a course unit is defined as the time period between the release of two units of course content. For most courses, a course unit is 1 week long. The variation of specificity and recall with the threshold is presented in Figure 3.

At very low thresholds, S is very low and R is very high because almost every student has an absence at least as long as the threshold. As the threshold is increased, S improves and R deteriorates. We identified the point at 2 course units (14 days for a typical course) as a convenient threshold, where R and S are both above 0.75, and have selected this value to be the threshold of our M2 predictor.

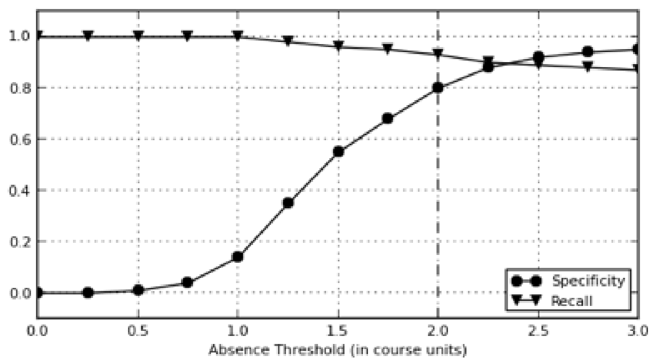


Figure 3. Variation of specificity and recall with the absence threshold

	Individual feature predictors				M1 predictor	M2 predictor	Integrated predictor
	assn-performance	video-skip	assn-skip	lag			
Specificity							
Best	1.0	0.86	0.96	0.96	0.85	0.93	0.68
Median	1.0	0.82	0.92	0.84	0.72	0.80	0.58
Worst	0.96	0.40	0.73	0.47	0.36	0.70	0.29
Recall							
Best	0.008	0.58	0.38	0.43	0.67	0.98	0.99
Median	0.006	0.30	0.25	0.17	0.48	0.93	0.93
Worst	0.00	0.23	0.10	0.14	0.41	0.77	0.91

Table 3. Best, median, and worst specificity and recall for various predictor components

Results

Specificity and Recall

First, we evaluate our predictor's specificity and recall observed over 10 test courses different from the 6 training-set courses. Table 3 shows the best, median, and worst observed recall and specificity figures.

In order to develop an understanding of what the strengths and weaknesses of our predictor are, we need to provide some interpretation of the numbers in Table 3.

The 'assn-performance' (assessment performance) Feature

This feature generally has high precision and specificity. Over 95% of students it flags (students with average assessment scores below 0.5) eventually dropout. However, its recall was observed to be generally very low compared to the other features. For the majority of MOOC quizzes, mean scores are in the range of 70% - 85%. Even though some students occasionally score below 50% on certain quizzes, there are very few students whose average quiz scores are below 50%. This could be attributed to the deliberate easiness for which MOOC assessments are designed, or due to MOOCs' self-selective nature (students who believe that the course will be too difficult refrain from enrolling or refrain from attempting assessments).

The 'video-skip' Feature

This feature was observed to vary in specificity across different courses. Its specificity is high for the majority of courses, as demonstrated by the small difference between the maximum and median, so it is generally a robust feature. Specificity worsens, however, for courses with too many videos per topic. We observed that persistent and dropout students alike tend to start skipping videos when the total duration of videos to watch per week exceeds 2 hours. Some specificity drop occurs in courses where it is not necessary for students to view every video in order to be able to follow future content. In such cases, some students who fell behind in watching some videos skipped them totally and continued viewing other content.

The 'assn-skip' Feature

Similarly, this feature's specificity is generally high, with noticeable drops in courses with heavier assignment workload. The recall of this feature is worse than that of video skip, due to the presence of a group of students who are interested in viewing the videos but not in the assessments.

The 'lag' Feature

This feature was observed to have higher recall in courses with stronger interdependencies between different parts of the content. In such courses, a student who falls back has to view what she has missed before proceeding to the more advanced units. This increases the probability that the student will not be able to continue the course after dropping behind by a certain amount, especially in courses with higher work loads. The peak recall of 0.43 in our study was observed for a probabilistic graphical models course with 2.5 to 3 hours of video per week, and a topic interdependency map that makes it difficult to follow a topic without having mastered the previous topics.

The Active Mode Predictor

This predictor was able to predict between 40% and 50% of dropouts most of the time. Its toughest challenge was courses with high workloads (all students tend to show signals of poor interest at some point in the course if the work load is constantly high, including those who persisted until the end of the course).

The Absent Mode Predictor

This predictor was able to pick up over 90% of dropouts in most of the courses. Lower specificity was observed in courses with lighter workloads, since such courses make it easier for a student to catch up and continue in a course after an extended absence.

The Integrated Predictor

The consistently high values of recall of the integrated predictor are a consequence of the integration of the M2 predictor. Recall of a combined predictor is at least as good as the recall of the best of its components. The biggest weakness in the integrated predictor, however, remains to be specificity, which has to be worse than its worst component. The worst observed specificity (0.3) was for the probabilistic graphical models course, which has a relatively high number of videos and assignments per week, leading the predictor to falsely red-flag many students who skipped some videos and assignments. In future work, we hope to improve the overall specificity by making the features more sensitive to specifics of the course, such as workload. Another strategy is to try to add a second step to filter out false positives. This can take the form of a survey that starts by asking the student about their learning experience in the course to try to confirm whether the student is really at risk. If the presence of risk factors is confirmed, the survey advances the student to the intervention stage.

Distribution of Intervention Window Lengths

The other important figure of merit of the prediction model is the length of the intervention window (the time between the first red-flag the student receives and the last activity the student performed in the course). Figure 4 below shows the distribution of intervention window lengths aggregated over several courses.

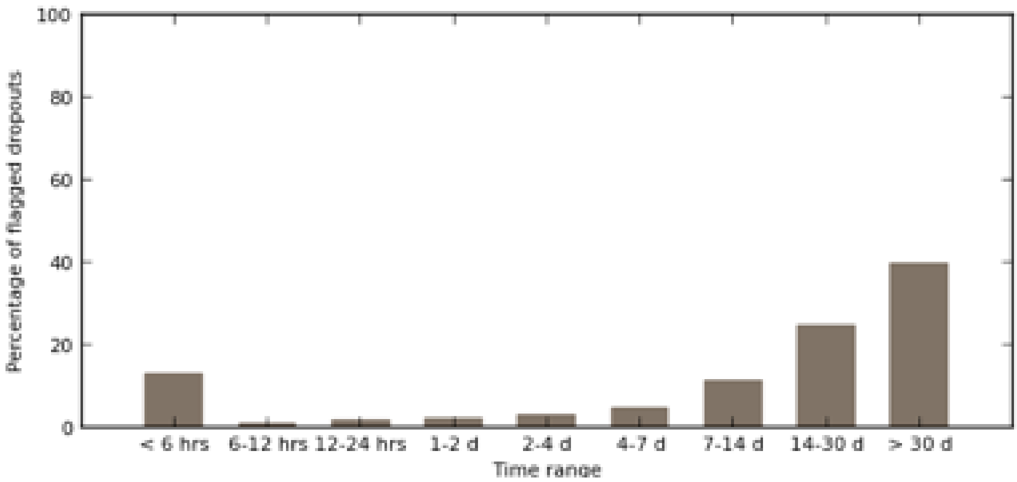


Figure 4. Percentage of flagged dropouts with intervention windows in 9 time ranges

The distribution shows that, for approximately 80% of the flagged dropouts, the student persists in the course for at least 4 days after the red flag is first raised. For well over 60% of the flagged dropouts, the student starts exhibiting activity patterns that raise the red flag more than 2 weeks before the last activity.

Conclusions and Future Work

Predicting student dropout is an important task in intervention design for MOOCs. Our study has shown that complete dropout is only one type of bad persistence patterns. Absence times exceeding 3 weeks are associated with drops on multiple performance metrics.

We have designed a prediction model that scans the student activity for signs of lack of ability or interest that may cause the student to dropout from the course or go absent for dangerously long periods. For most courses, our model predicted between 40% and 50% of dropouts while the student was still active. By red-flagging students who exhibit absences of 14 days or longer, the recall increases to above 90%.

The time window from the first red flag to the last activity shown by the student in the course is a critical figure that affects the effectiveness of the interventions we can deliver. Our analysis reveals that, through our choice of predictive features, we are able to spot risk signals at least 2 weeks before dropout for over 60% of the students. This suggests that it is feasible to design and deliver timely interventions using our prediction model.

As future work, we plan to use multiple strategies to improve the performance and usefulness of our prediction model. First of all, we have answered the question "What are some different activity patterns, inspired by persistence factors, that we can use to predict dropout?" However, we have not answered the question of "Which of the persistence factors do we believe student X lacks?" If our model could be made to distinguish whether a student is at risk due to lack of ability, interest, or both, it would have better implications on intervention design in MOOCs.

Secondly, we believe that other persistence factors exist that have to be studied, including mindset, self-efficacy (Bandura, 1994), goal setting (Locke & Latham (1990), Locke & Latham (2002)), and social belongingness (Walton & Cohen (2007), Walton & Cohen (2011)). Expanding our feature set to measure these factors, as well as using more sophisticated machine learning algorithms to enhance the design and combination of features are two directions that could potentially improve prediction performance and deepen our understanding of what makes a student persist in or leave an online course.

References

Bandura, A. (1994). Self-efficacy. Wiley Online Library. Retrieved December 2013 from http://onlinelibrary.wiley.com/book/10.1002/9780470479216

Breslow, L. B., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. Research & Practice in Assessment (8), 13–25.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. Review of educational research, 65(3), 245–281.

Hoskins, S. L., & Van Hooff, J. C. (2005). Motivation and ability: which students use online learning and what influence does it have on their achievement? British Journal of Educational Technology, 36(2), 177–192.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the Third International Conference on Learning Analytics and Knowledge, 170–179.

Lee, Y., & Choi, J. (2011). A review of online course dropout research: implications for practice and future research. Educational Technology Research and Development, 59(5), 593-618.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. Prentice-Hall, Inc. Retrieved December 2013 from http://psycnet.apa.org.laneproxy.stanford.edu/psycinfo/1990-97846-000

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist, 57(9), 705.

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. The Internet and Higher Education, 6(1), 1–16.

Tinto, V. (2006). Research and practice of student retention: what next? Journal of College Student Retention: Research, Theory and Practice, 8(1), 1–19.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. Journal of personality and social psychology, 92(1), 82.

Walton, G. M., & Cohen, G. L. (2011). A brief social-belonging intervention improves academic and health outcomes of minority students. Science, 331(6023), 1447–1451.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. Educational psychologist, 25(1), 3–17.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. Contemporary educational psychology, 25(1), 82–91.

Edition and production

Name of the publication: eLearning Papers
ISSN: 1887-1542
Publisher: openeducation.eu
Edited by: P.A.U. Education, S.L.
Postal address: c/Muntaner 262, 3r, 08021 Barcelona (Spain)
Phone: +34 933 670 400
Email: editorialteam[at]openeducationeuropa[dot]eu
Internet: www.openeducationeuropa.eu/en/elearning_papers