Early Prediction of Success in MOOC from Video Interaction Features

Boniface Mbouzao[1(✉)], Michel C. Desmarais[1(✉)], and Ian Shrier[2(✉)]

[1] Polytechnique Montreal, Montreal, Canada
{boniface.mbouzao,michel.desmarais}@polymtl.ca
[2] McGill University, Montreal, Canada
ian.shrier@mcgill.ca

Abstract. The popularity of online learning, such as MOOCs (Massive Open Online Courses), continues to increase among students. However, MOOCs dropout remains high. Prediction of student performance that could feed instructors' dashboards and help them adapt their course structure and material, or trigger help and tailor interventions to specific groups of students, is a valuable research objective. Towards that end, this paper focuses on three predictive metrics (student attendance rate: AR, utilization rate: UR, and watching index: WI) of how students interact with MOOC videos in order to predict which group of students will pass or fail the course. Results show that these metrics, taken after the first week and the midpoint, can be highly effective for predicting the students that will pass or fail the course.

Keywords: MOOC · Video watching traces · Attendance rate · Utilization rate · VBL · Watch index

1 Introduction

Can the analytics of student interaction with videos provide effective predictors of their academic performance? With thousands of students registering for online courses every year, it is of great interest to both MOOC developers and instructors to predict, early on in the course, those students that may drop out or who are instead likely to complete the course successfully. Yet, by using data directly from student behaviours with video-based-learning (VBL), we believe we can predict student performance with minimal research bias while helping the student to avoid failing the course. The interactions during the first week of the course are a key reference point in predicting student performance throughout the course [8], should the student remain enrolled in the MOOC. This can inform the MOOC developers, and instructors, of those students who are likely to persist and perform better in the MOOC if the course was tailored to their learning needs rather than focus on large-volume enrollment to the detriment of successful student performance. The aim of this paper is to investigate characteristics of video utilization that can determine if a student will likely fail or pass at an early stage of the MOOC; even as early as after one week of interaction with VBL modules.

2   Related Work

A fair amount of research is devoted to study student video interactions, including specific characteristics of video watching behaviors.

For instance, Giannakos et al. [4] found a relationship between video interactions, repeated viewing, and the level of cognition required for a specific video segments. Li et al. [13], also studied the link between student behavior patterns and learning performance. Their study included features as: videos lectures, lecture slides, shared assignments and open-forum messages. They concluded that video lectures and lecture slide were the most used by students. He et al. [7] measured the student's utilization of video in a MOOC and its relation to academic performance. They proposed indicators based on student interaction data to measure the utilization of video resources such as: attendance and utilization rates plus watch ratio. Our study measures the utilization and attendance rates for performance prediction. Hughes et al. [9], show how the use of data analysis techniques in MOOCs can help to predict at risk of dropout students before it happens. Some studies used learners' behaviour by defining engagement [12], online social networking [3] to predict performance [10] or predict dropouts [6]. Then it is possible to identify student that might dropouts early enough [11] to give them proper help before it happens [14]. The student performance can be influenced by some features of the videos themselves [5] bringing students to play a video more than others [2,17], or spend more time on video than on other online resources [16].

Beyond video interactions, prediction of success and failure was done by Baker et al. [1] based on student online activity using Soomo Learning Environment. Using many prediction models they show that using the logistic regression model of their propose combined model they can identify up to 59.5% of students who will perform poorly in the course way better than chance performance. Recently, Adithya et al. [18] analyzed students logs of three blended courses to predict student performance, (see also Owen et al. [15] with student profile for performance prediction). The results of their research show that it is possible to predict student performance through student level of use of online system.

3   Methodology

The general objective of this study is to use video interaction metrics to assess the likelihood that a student will succeed in a MOOC. We also aim to evaluate the accuracy of these estimates as a function of the MOOC timeline; earlier estimates are deemed to be less accurate but more useful for remedial purposes. We first describe the three metrics involved and then describe how they are used to define groups. The metrics such as AR and UR are inspired from [7].

The attendance rate $AR_{s,c}$ of a student s on a given week c since the beginning of the course, is the number of videos that the student played over to the total number of videos up to that period in time of the course schedule.

The utilization rate $UR_{s,c}$ of a student s on a given week c since the beginning of the course is the proportion of video play time activity of the student over the sum of video lengths for all videos up to week c.

$$AR_{s,c} = \frac{|W_{s,c}|}{|V_c|} \tag{1}$$

$$UR_{s,c} = \frac{\sum_{i=1}^{n} Wt_{s,i}}{\sum_{j=1}^{N} Vt_j} \tag{2}$$

The watch index (WI) is defined as:

$$WI_{s,c} = UR_{s,c} \times AR_{s,c} \tag{3}$$

Where: $UR_{s,c}$ is utilization rate of student s at week c, Vt_j is the duration of video j, N the total number of videos released up to week c, $Wt_{s,i}$ is the total time student s played video i, n is the number of unique videos that student s played up to week c. $AR_{s,c}$ is the attendance rate of student s after week c, V_c is the course's total number of videos released after week c, $W_{s,c}$ is the collection of different videos watched by student s after week c.

The rule for grouping and testing is taking from the classification of a single student compared to the average value of the group of student. The student that has metrics over the average value of the group of student enrolled in the course is likely to pass the course. The algorithm for groupings student is:

IF ($WI_{i,c} < \overline{WI}$ OR (student i did not submit assignment))
THEN {Student i in GROUP I}
ELSE {Student i in GROUP II}
Where $\overline{WI}$ is the average WI over all students.

4   Data Set

Our study uses the traces of an online course at McGill University on the edX platform that had 30,640 students interacting with learning videos. The course has 138 videos (plus one live session). It has 13 weeks duration. Among students who interacted with the videos, 10,424 students are honors and 970 passed the course. Here, the first week of the course had nine (9) videos with a total length of 41 min. We extracted 2,733,169 different student watching events (play, pause, seek forward/backward, stop). Event data is structured such that we can compute the play time of each student for each video.

5   Results

Table 1 shows the results of classifying students into two groups according to the rule defined above. We find that after a week of data, students who passed the course overwhelmingly fall into group II (78%). The trend is even stronger after

week 6 (93%). However, for the students who failed (or dropped) the course, the split is almost even between the two groups. Then, 60% of the students who fail are in Group I, and that ratio grows to 2/3 after the sixth week. We can see that around 4/5 of the students that will succeed the course will be in Group II, whereas around 60% of students who will fail will be in Group I. These results show an improvement compared to previous study using the same kind of metrics to red flag at risk of failing students. Comparing these results to the one that we obtained using the methodology of He et al. [7] (results shown within parenthesis), we see that after the first week there is no majority class separating student who will pass or fail between the two groups. The majority of students who will pass and the majority of student who will fail are all in group II according to He et al. [7] methodology. After six week the results based on their methodology identify majority classes (majority of students who passed are in group II and majority of students who fail are in group I) but there are still 35% of students who passed in group I.

Table 1. Breakdown of success rates and groupings for first and sixth week

	After first week		After six weeks	
	Pass	Fail	Pass	Fail
Group I	22% (15%)	60% (9%)	7% (35%)	67% (93%)
Group II	78% (85%)	40% (91%)	93% (65%)	33% (7%)
	100% (100%)	100% (100%)	100% (100%)	100% (100%)

Values within parenthesis are the results of He et al. [7], reported for comparison purpose.

6 Conclusion

Through quantitative analysis that includes metrics such as Attendance Rate (AR), Utilization Rate (UR), and Watch Index (WI), as defined in this paper, it is possible to identify failure patterns of up to 60% of students who will dropout or fail the course based on the first week student interaction with MOOC videos, based on a total course length of thirteen (13) weeks, and can identify 78% of successful students.

Using the metrics defined, educational institutions can flag students at risk of failing, or dropout, of the MOOC based on the student's interactions with the learning videos in the early stages of the course. Our study shows a better classification compared previous study results of He et al. [7]. Results such as these should help MOOC developers better identify those students that might dropout or fail the course and thus take actions to prevent it.

References

1. Baker, R.S., Lindrum, D., Lindrum, M.J., Perkowski, D.: Analyzing early at-risk factors in higher education e-learning courses. International Educational Data Mining Society (2015)
2. Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., Seaton, D.T.: Studying learning in the worldwide classroom: research into edx's first MOOC. Res. Pract. Assess. **8**, 13–25 (2013)
3. Brinton, C.G., Chiang, M., Jain, S., Lam, H., Liu, Z., Wong, F.M.F.: Learning about social learning in MOOCS: from statistical analysis to generative model. IEEE Trans. Learn. Technol. **7**(4), 346–359 (2014)
4. Giannakos, M.N., Chorianopoulos, K., Chrisochoides, N.: Making sense of video analytics: lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. Int. Rev. Res. Open Distrib. Learn. **16**(1), 260–283 (2015)
5. Guo, P.J., Kim, J., Rubin, R.: How video production affects student engagement: an empirical study of mooc videos. In: Proceedings of the First ACM Conference on Learning@ Scale Conference, pp. 41–50. ACM (2014)
6. Halawa, S., Greene, D., Mitchell, J.: Dropout prediction in MOOCS using learner activity features. Exp. Best Pract. Around MOOCs **7**, 3–12 (2014)
7. He, H., Zheng, Q., Dong, B., Yu, H.: Measuring student's utilization of video resources and its effect on academic performance. In: 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), pp. 196–198. IEEE (2018)
8. Hill, P.: Emerging student patterns in MOOCs: a (revised) graphical view (2013). https://mfeldstein.com/emerging-student-patterns-in-moocs-a-revised-graphical-view/
9. Hughes, G., Dobbins, C.: The utilization of data analysis techniques in predicting student performance in massive open online courses (moocs). Res. Pract. Technol. Enhanced Learn. **10**(1), 10 (2015)
10. Jiang, S., Williams, A., Schenke, K., Warschauer, M., O'dowd, D.: Predicting MOOC performance with week 1 behavior. In: Educational Data Mining 2014 (2014)
11. Kamahara, J., Nagamatsu, T., Kaieda, Y., Ishii, Y.: Behavioral analysis using cumulative playback time for identifying task hardship of instruction video. In: 2010 5th International Conference on Future Information Technology, pp. 1–6. IEEE (2010)
12. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of the Third International Conference on Learning Analytics and Knowledge, pp. 170–179. ACM (2013)
13. Li, L.Y., Tsai, C.C.: Accessing online learning material: quantitative behavior patterns and their effects on motivation and learning performance. Comput. Educ. **114**, 286–297 (2017)
14. Li, N., Kidzinski, L., Jermann, P., Dillenbourg, P.: How do in-video interactions reflect perceived video difficulty? In: Proceedings of the European MOOCs Stakeholder Summit 2015, pp. 112–121. No. EPFL-CONF-207968, PAU Education (2015)
15. Lu, O.H., Huang, A.Y., Huang, J.C., Lin, A.J., Ogata, H., Yang, S.J.: Applying learning analytics for the early prediction of students' academic performance in blended learning. J. Educ. Technol. Soc. **21**(2), 220–232 (2018)

16. Ozan, O., Ozarslan, Y.: Video lecture watching behaviors of learners in online courses. Educ. Media Int. **53**, 27–41 (2016)
17. Seaton, D.T., Bergner, Y., Chuang, I., Mitros, P., Pritchard, D.E.: Who does what in a massive open online course? Commun. ACM **57**(4), 58–65 (2014)
18. Sheshadri, A., Gitinabard, N., Lynch, C.F., Barnes, T., Heckman, S.: Predicting student performance based on online study habits: a study of blended courses. arXiv preprint arXiv:1904.07331 (2019)