Predicting Higher Education MOOCs Engagement-Level Odds; A Stochastic Approach (0098)

Saman Zehra Rizvi, Sayeed Ghani
Institute of Business Administration, Pakistan

Introduction:

Massive open online courses (MOOCs) are said to be different from previous Open Education Resources (OER) is a way that they almost always provide some kind of assessment, certification or accreditation. With millions of users enrolled in MOOCs, low students retention and certification rate can be considered as a potential threat to the overall phenomenon. [1] Research shows that a clear purpose of participation in MOOCs is to attain some form of certificate, professional diploma or accredited credit hour. Presently, this can be accomplished with or without accessing a substantial amount of course contents available. To maximize the utilization of MOOCs, it is important to understand the association between accessed course contents and certification or retention rate. This study used Markov chain to predict the retention and certification likelihood based on the fraction of course material learners accessed. Since the MOOCs under analysis were higher education MOOCs offered by two prestigious universities; MIT and Harvard, we expect this study to eventually contribute towards useful integration of MOOCs in formal academia.

Related Work:

MOOC is a relatively new research area in e-learning. However, we found a number of studies [2-5] addressing the issue of learners retention and certification. In [2] the researchers advocated that first week performance is a strong indicator of retention. Researchers used assignment performance and social interactions to predict that external incentives help learners in not only to maintain their performance but also to get certification. [3] Used Hidden Markov Models (HMM) to predict learners' dropout based on some features; cumulative percentage of lecture video watched, number of forum threads viewed, number of posts, number of time course progress page was checked and a combination of above. In [4] researchers used behavioral data for early prediction. A correlation was found between student performance and lecture video-watching behavior. Also whether a learner has been corrected using Correct on First Attempt (CFA) or not. [5] Used CFA and clickstream data to predict students' behavior. The researchers proposed two frameworks by using video-watching clickstreams; one based on sequence of events created, and another on sequence of positions visited. [6] Discussed Markov based clustering and social network based modeling approaches to describe different video watching preferences in MOOCs.

Apart from the issue of certification, there are number of studies [7-9] which address the issue of retention and dropout. [7] Predicted the probabilities to identify learners at risk. This paper also suggested when instructor design interventions are needed? Similar issue was explored in [8] which used features such as assessment performance, video skip, assignment skip, etc. [9] classified learners into two categories: first, those who watched videos only and second, who not only viewed videos but attempted the quizzes as well. This study used learners' behavior such as video lecture downloads, taking weekly quizzes, and solving peer assessments, along with some additional features like number of lecture views and video quiz attempts etc. For prediction, two discrete variables were selected; (a) the dropout week when a learner watched less than 10% of the remaining lectures, (b) final grade.

Proposed Approached:

In 2014, HarvardX Research Committee and Office of Digital Learning at MIT issued a report along with a public data set for first HarvardX and MITx courses. It was a significant attempt to enhance and facilitate MOOC related research. Over the academic year from fall 2012 to summer 2013, HarvardX and MITx launched 17 courses on edX, a joint platform for delivering online courses [1]. For this study we used this edX detailed dataset for first two years of MOOCs.

In dataset, attribute *View* represented learners who accessed less than 50% of course chapters, *Explore* who accessed more than 50%, C*ertified* who received certificate. Some learners after completion of one course, opted for another course. During data analysis it was observed that learners were naturally divided into four categories. We used those categories as states in our proposed Markov chain (Table1).

Table 1

Category	**State**	**View**	**Explore**	**Certified**
No participation after enrollment	S1	0	0	0
Accessed less than 50% of the chapters	S2	1	0	0
Accessed more than 50% of the chapters	S3	1	1	0
Accessed more than 50% of the chapters and attained certificate	S4	1	1	1

From the original dataset, we selected two courses offered by MITx. First, *Circuits and Electronics* (6.002x) offered in fall 2012. Second, *Electricity and Magnetism* (8.02x) offered in spring 2013. Both courses had a pre-requisite/post-requisite relationship with each other and are listed as *related courses* on MITx website. It was found that 4179 students were enrolled in both courses. For correct categorization, we removed couple of outliers.

Markov chain and *Chapman–Kolmogorov* equation

Like any other online service, future participation in online learning is dependent upon present experience of the users. Under the above assumption, MOOC participation is essentially a Markovian process. With state space: Si = {S1, S2, S3, S4}, we first calculated the initial distribution for students being in one of these states in one semester. Next, we calculated the transition probabilities for subsequent semester (Figure1).

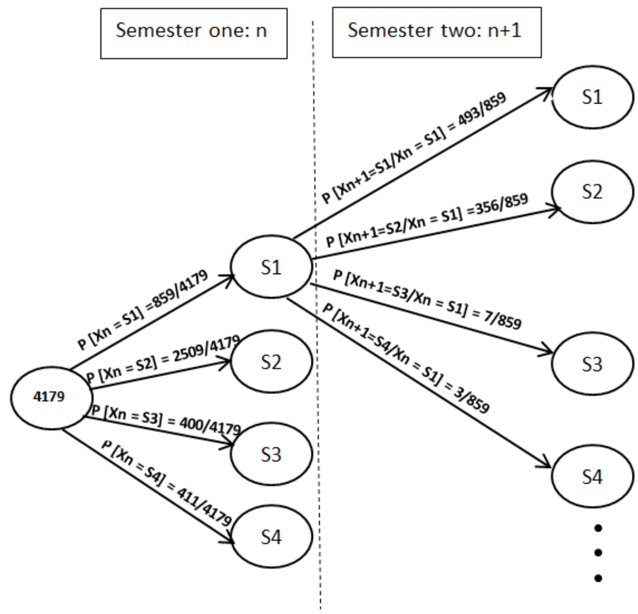

Figure 1

Using the above, we proposed a Markov chain to represent the process (Figure2).

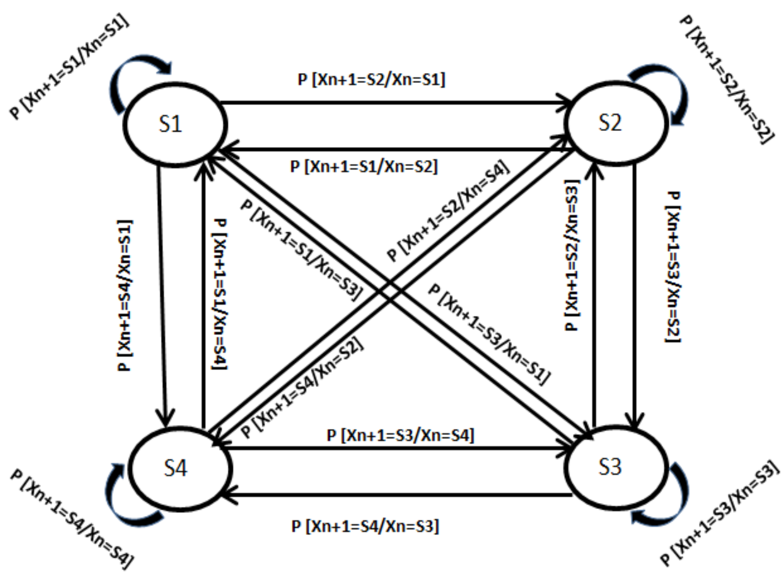

Figure 2

Furthermore, we used one-step transition probabilities to define the *n*-step transition probabilities. Finally, the *Chapman–Kolmogorov equations* provided a method for computing the *n*-step transition probabilities and Markov chain displayed the characteristic of steady state. This method enabled us to predict future participation. For example if a learner started from state S0 (no participation, only enrollment) in first semester? There is almost 55% chance that she will be in state S1 after 4 semesters. This unique approach is useful to foresee the MOOCs engagement odds and learners' transition from one learning state to another learning state. The proposed method can also be used to investigate the effects of interventions on learners' retention in different offering of MOOCs and hence, offer propositions for improvements.

References:

[1] Joseph, AI Maria, and B. Asoke Nath. "Integration of Massive Open Online Education (MOOC) System with in-Classroom Interaction and Assessment and Accreditation: An extensive report from a pilot study." *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2013.

[2] Jiang, Suhang, et al. "Predicting MOOC performance with week 1 behavior."*Educational Data Mining 2014*.

[3] Balakrishnan, Girish, and Derrick Coetzee. "Predicting student retention in massive open online courses using hidden markov models." *Electrical Engineering and Computer Sciences University of California at Berkeley* (2013).

[4] Brinton, Christopher G., and Mung Chiang. "Mooc performance prediction via clickstream data and social learning networks." *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015.

[5] Brinton, Christopher G., et al. "Mining MOOC Clickstreams: On the Relationship Between Learner Video-Watching Behavior and Performance."*ACD Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 2015.

[6] Sinha, Tanmay, et al. "Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions."*arXiv preprint arXiv:1407.7131* (2014).

[7] He, Jiazhen, et al. "Identifying At-Risk Students in Massive Open Online Courses." *AAAI*. 2015.

[8] Halawa, Sherif, Daniel Greene, and John Mitchell. "Dropout prediction in MOOCs using learner activity features." *Experiences and best practices in and around MOOCs* 7 (2014).

[9] Ye, C., & Biswas, G. (2014). Early prediction of student dropout and performance in MOOCs using higher granularity temporal information. *Journal of Learning Analytics, 1*(3), 169–172 (2014)