

Who Will Be the Next to Drop Out? Anticipating Dropouts in MOOCs with Multi-View Features

FENG JIANG^{1*}, and WENTAO LI²

¹College of City Construction Engineering, Chongqing Radio and TV University, Chongqing 400052, China

²Centre for Artificial Intelligence, School of Software, Faculty of Engineering & Information Technology, University of Technology Sydney, Ultimo, NSW 2007

(Received on September 03, 2016, revised on October 16, 2016)

Abstract: Massive Open Online Courses (MOOCs) have gained rising popularity in recent years. However, MOOCs have faced a challenge of a large number of students dropping out from courses. Most studies predict dropouts based on some general features extracted from historical learning behavior and ignore the diversity of the behaviors. To solve this problem, we first analyze each type of learning behavior independently to get the different behavior patterns between dropout and retention students. We then derive multiple kinds of features from the corresponding types of learning behavior records. After that, we propose three algorithms that make use of these features. The first one trains several detectors based on each types of features. The second utilizes multi-view ensemble learning to anticipate dropouts. The third applies semi-supervised co-training to train the detector. Experimental results justify the rationality of the multi-view features and the proposed approaches achieve better prediction performances.

Keywords: Co-training, Dropout prediction, Massive open online courses, Multi-view ensemble learning.

1. Introduction

Education informationization is made possible with the development of web2.0 and cloud computing. Massive Open Online Courses (MOOCs) are the product of Internet application innovation and collaboration computing in the area of education [1-2]. With the rise of online-learning websites such as edX, Coursera and Udacity, MOOCs have received more and more attention worldwide [1-3].

Since MOOCs are open to everyone, learners in MOOCs have a large difference mainly in their educational background and learning motivation [3-5]. Moreover, the price of dropping courses is quite lower for students, while the dropout rate is very high in MOOCs. Analyzing factors which lead to dropout can help improve the construction of MOOC platform and predict whether students will drop courses. Taking these measures in the early stage to improve retention has important significance to the success of MOOCs. The existing research mostly treated the prediction of student dropout as a classification problem. Yang et al. [6] developed a survival model to reveal significant predictors of dropout. Dernoncourt et al. [12] proposed a prediction model based on over 25 predictive features. Tang et al. proposed a framework that applies big data methods to identify the students who are likely to dropout in MOOC [1]. Xing et al. proposed a temporal modeling approach based on a summed features modeling space. The research anticipates student dropouts only according to some general features extracted from student learning behavior records [2]. They ignore how learning behaviors of students change over time.

*Corresponding author's email: jiangfeng@cqdd.cqu.cn

In this paper, by comparing the number of learning behavior over weeks between dropout students and retention students, we find different behavior patterns between these two kinds of students, thus getting multi-view features based on each kind of behavior. After that, we propose three algorithms based on three kinds of machine learning methods to determine whether a student will drop a course according to the way multi-view features are used.

2. Background

MOOCs are the new product of networking application, and the related research is still in the initial stage. The current research can be divided into three parts [4, 8].

(1) Rosé [9] proposed a survival model to measure the impact of three social factors that make predictions about attrition for students who have participated in the discussion forum. Nesterko et al. [5] formalized the process of evaluating the geographic data of students with regard to enrollment and certificate attainment. Seaton et al. [10] explored how course structure impacts overall video consumption across courses.

(2) O'Reilly [11] described an agenda for developing technology that enables MOOC analytics. The agenda efficiently addresses the detailed, low level, high volume nature of MOOC data and helps exploit the data's capacity to reveal how students behave and how learning takes place. He then developed a platform (MoocViz) [7] to help researchers analyze MOOC data from multiple platforms without the need to share the data.

(3) Dernoncourt et al. [12] predicted dropout for the Fall 2012 offering of 6.002x, which involved the meticulous and crowd-sourced engineering of over 25 predictive features extracted for thousands of students. Yang et al. [6] developed a survival model that measures the influence of factors extracted from discussion forums to predict student dropout.

3. Behavior Analysis and Multi-View Features Extraction

3.1. Types of learning behavior

In this paper, learning behavior data from XuetangX, the Tsinghua University's MOOC platform, is used as an example. It contains 11 courses and about 800 million student behavior records. Learning behaviors divided mainly into six categories (Table 1).

Without the loss of generality, course C is randomly selected from all eleven courses as an example. Learning behavior data of course C includes a total of 2392 users with more than 20 million records of learning behaviors in 5 weeks. We use one week as a unit and then make a statistic of the number of each kind of behavior over five weeks in course C. The number is denoted as $\text{Num}_j(i) = (w_{ij_1}, w_{ij_2}, w_{ij_3}, w_{ij_4}, w_{ij_5})$, where i is the i -th student; j is the j -th kind of behaviors and $j = 1, 2, \dots, 6$; and w_{ij_k} is user i 's number of j -th behavior in the k -th week.

Each student gets $\text{Num}_1(i)$, $\text{Num}_2(i)$, $\text{Num}_3(i)$, $\text{Num}_4(i)$, $\text{Num}_5(i)$, $\text{Num}_6(i)$, and each vector is the number of behavior in every five weeks. In this data set, if a user leaves no records for course C in the website log during the next 10 days, we define him as a dropout student from course C, otherwise he is an anticipant. After figuring out dropout and participant students, we make an average of the number of each behavior between the two types of students. For example, for the first behavior we get $\text{Num}_1(\text{dropout}) = \text{mean}(\text{Num}_1(i))$, where i is dropout student; $\text{Num}_1(\text{participant}) = \text{mean}(\text{Num}_1(i))$, where i is participant student. Then we make comparisons of these six kinds of behaviors to find

the differences in behavior patterns between dropout and participant students.

Table 1: Six kinds of behaviors in data set

Type of learning behavior	Detail
#1 assignments viewing behavior	Working on course assignments.
#2 videos viewing behavior	Watching course videos
#3 other objects accessing behavior	Accessing other course objects except videos and assignments
#4 page closing behavior	Closing the web page
#5 wiki accessing behavior	Accessing the course wiki
#6 course forum accessing behavior	Accessing the course forum

3.2. Behavior analysis and Features extraction

(1) Features extraction from assignments viewing behavior

Assignments viewing behavior reflect that the learner is serious about learning and that this behavior is the manifestation of students' initiative learning. Fig. 1 shows the contrast of the number of assignments viewing behavior between two kinds of students, Num₁(dropout) and Num₁(participant). The number of two types of students' assignments viewing over five weeks is different. Num₁(dropout) decreased gradually as time went on and Num₁(participant) generally showed some fluctuations. This is denoted by Num₁(i).

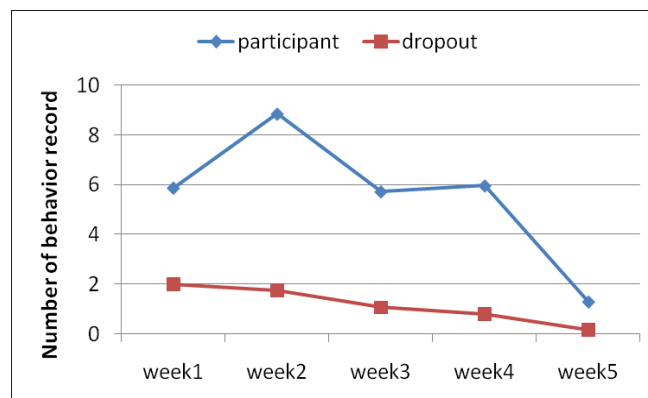


Fig. 1: the assignments viewing behavior

(2) Features extraction from video viewing behavior

Video viewing behavior is included in the whole process of student learning and video resource is the important resource of MOOCs. Thus, watching video behavior can reveal the enthusiasm of a student. Fig. 2 was obtained by the contrast of Num₂(dropout) and Num₂(participant). The number of two types of students' behavior in every week is

different. $Num_2(dropout)$ decreased gradually as the time went on and $Num_2(participant)$ tends to be flat in the third week and the fourth week. This is denoted by $Num_2(i)$.

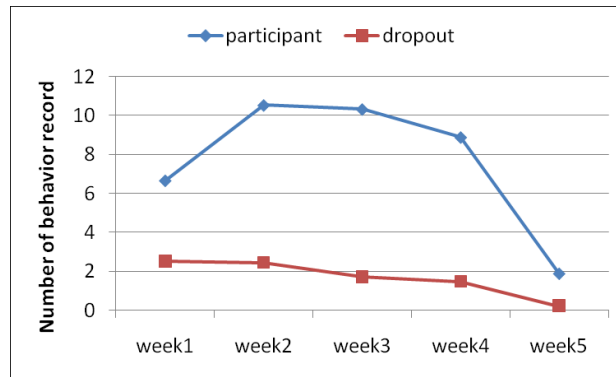


Fig. 2: Contrast of videos viewing behavior

(3) Features extraction from other objects accessing behavior

Other resources are resources besides video and problem, such as teaching plan. Fig. 3 was obtained by the contrast of $Num_3(dropout)$ and $Num_3(participant)$. While other objects accessing behavior and video viewing behavior are different in the specific number, the trend of two kinds of behavior trend is the same between dropout and participant students. This is denoted by $Num_3(i)$.

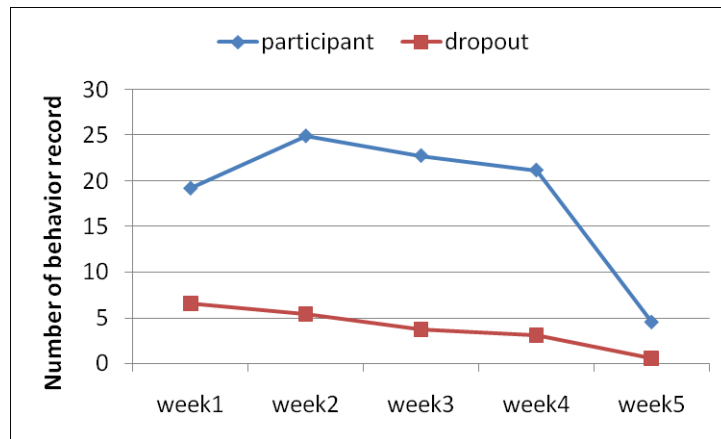


Fig. 3: Contrast of assignments viewing behavior

(4) Features extraction from Contrast of page closing behavior

Closing page represents the end of a student's studying process. Fig. 4 was obtained by the contrast of $Num_4(dropout)$ and $Num_4(participant)$. The trend of this behavior and the former three behaviors is basically the same. $Num_4(participant)$ remains flat in the middle weeks, while $Num_4(dropout)$ continues to decline as time goes on. These features are denoted by $Num_4(i)$.

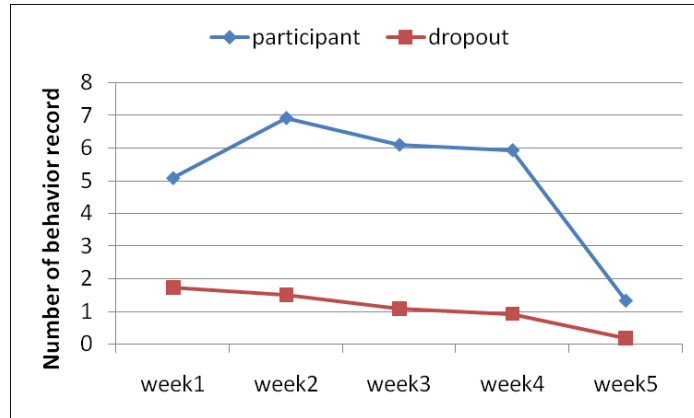


Fig. 4: Contrast of page closing behavior

(5) Features extraction from Contrast of wiki accessing behavior

Students find things about courses through wiki. Figure 5 was obtained by the contrast of $Num_5(dropout)$ and $Num_5(participant)$. Both $Num_5(dropout)$ and $Num_5(participant)$ are declining. These features are denoted by $Num_5(i)$.

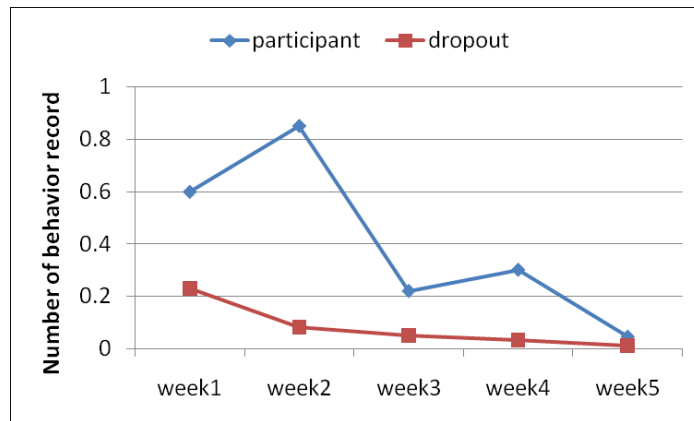


Fig. 5: Contrast of wiki accessing behavior

(6) Features extraction from course forum accessing behavior

Students begin mutual cooperation and learning with each other by discussing problems on forums. Figure 6 was obtained by the contrast of $Num_6(dropout)$ and $Num_6(participant)$. Both $Num_6(dropout)$ and $Num_6(participant)$ are decreasing. These features are denoted by $Num_5(i)$.

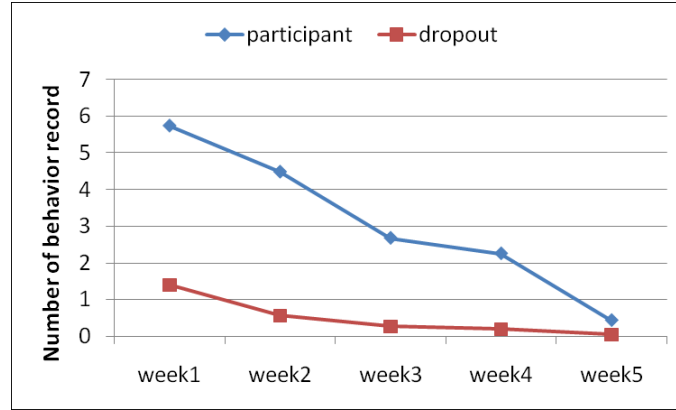


Fig. 6: Contrast of course forum accessing behavior

Through the contrast of six kinds of behavior, we can make the following conclusion:

- 1) The number of each kind of behavior of the dropout students decreases as the time went on, which showed that the dropout students' participation mostly in the beginning.
- 2) The number of each kind of dropout students is far lower than that of the participant students, which shows that the number of students' learning enthusiasm in MOOCs is much lower than the participant users.
- 3) The number of participant students' form four kinds of behavior over every five week looks similar, suggesting that participant students' enthusiasm shows some certain degree of stability.
- 4) The number of participant students' form four kinds of behavior over every five week looks similar with the dropout students, although it is different in the specific number.
- 5) We can make a statistic analysis of the number of six behaviors over five weeks. The number of every kind of behavior over each week can be used as one view of student's features.

4. Methods

From the statistical analysis of the data we have six views of features based on each kind of student learning behavior, which can be used to predict student dropouts. In this part, we propose three algorithms based on different ways to use these multi-view features.

4.1. Dropout detection algorithm based on single-view features (SVF)

The idea of SVF is to train classifiers on each view of features, and then the best classifier on the test data set can be selected as the final detector. Every classifier is based with traditional classification models: Naïve Bayes, Decision Tree and Support Vector Machine (SVM), and the specific steps are shown as in algorithm 1.

Algorithm 1: Student dropout detection algorithm based on single-view features

1. Train classifier in each view of features and corresponding labels. Data samples are divide into training set and testing set.
 2. Six kinds of classifiers based on single view of features are obtained and can be
-

denoted as Classifier k , $k=1, 2, \dots, 6$.

3. The classifier with the best performance on the testing set is selected as the final detector.

4.2. Dropout detection algorithm based on multi-view ensemble learning (MVF)

To analyze the synergistic effect of multiple views of features on prediction, we introduce ensemble learning to make different views of features work together. For the sake of convenience, dual views are integrated as an example, Naïve Bayes model is selected as the classification model and the algorithm is shown below.

Algorithm 2: Dropout detection algorithm based on multi-view ensemble learning

1. Combine two views of features and corresponding label, samples can be divided into training set and testing set.
 2. Train two classifiers under two views in the training set synchronously. Therefore, two Naive Bayes classifier C_j and C_k are produced.
 3. For each sample X_m in testing set, use two classifiers to get the student dropout probability P_{mj} and P_{mk} . When $(P_{mj}+P_{mk}) \geq 1$, student m is predicted as dropout, otherwise student m is predicted as participant.
-

4.3. Dropout detection algorithm based on semi-supervised co-training (SSC)

To adapt to the needs in the real system and reduce labeling effort, the paper uses the third method, student dropouts predict algorithm based on semi-supervised co-training learning (SSC), to make full use of labeled and unlabeled samples, which can be extended to real-life situations. Specific steps are shown as below.

Algorithm 3: Dropouts detection algorithm based on semi-supervised co-training

1. Combine two views of features and corresponding label, samples can be divided into training set and testing set.
 2. While unlabeled sample set S_U is non-empty, it does follow the iteration of S3-S5.
 3. Train classifiers under two views in the S_L respectively, and two Naive Bayes classifier $Classifier_j$ and $Classifier_k$ are produced.
 4. Choose samples randomly in unlabeled samples S_U and for each sample n in the set, use two classifiers to obtain dropout probability P_{nj} and P_{nk} . If P_{nj} and P_{nk} are greater than 0.5, then labeled student n as dropout. If P_{nj} and P_{nk} are less than 0.5, then labeled student n as participant, otherwise we do not label this sample. The labeled sample can then be added to S_L . The process can be denoted by $S_L' = S_L + S'$, $S_U' = S_U - S'$, where S_L' and S_U' is the updated training set.
 5. For each sample X_m in testing set, use two classifiers to get the student dropout probability P_{mj} and P_{mk} . When $(P_{mj} + P_{mk}) \geq 1$, student m is predicted as dropout, otherwise student m is predicted as participant.
-

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this part, we first introduce the experimental setting and evaluation criterion, and then we show the experimental results of three proposed algorithms: SVF, MVF, and SSC.

5.1. Experimental settings and evaluation criterion

The data set is a publicly available data set provide by KDD Cup2015 and XuetangX. We randomly selected a course and produced the corresponding learning behavior records for feature extraction and classifier establishment. In this data set, 2392 user behavior records are included, 546 participant and 1846 dropouts.

The precision (f_p), recall (f_r) and F-measure (F) are used as evaluation criteria. Assuming that N is the number of students who are predicted as dropout students by classifier, N_t is the number of students who are really drop courses, and N_a is the number of dropout students predicted by classifier. The f_p , f_r , and F are calculated by (1)-(3).

$$f_p = \frac{N^a}{N} \quad (1)$$

$$f_c = \frac{N^a}{N^t} \quad (2)$$

$$F = \frac{2f_p f_r}{f_p + f_r} \quad (3)$$

5.2. Experimental results and discussion

1) Results analysis and discussion of SVF

SVF uses classifiers from single view and three kinds of classification models to conduct the classification for each detection: Naive Bayes, Decision Tree (JB48) and SVM. The comparative method consists of two, the first is to use the number of all behaviors on each week as features without the type, and the second comparison method is to merge the six features into a large feature vector.

Table 2: Performance of SVF with Naive Bayes, Decision Tree, and SVM

Kind of features	Naïve Bayes			Decision Tree			SVM		
	f_p	f_r	F	f_p	f_r	F	f_p	f_r	F
Features from view #1	0.814	0.825	0.816	0.838	0.847	0.839	0.834	0.844	0.842
Features from view #2	0.828	0.836	0.816	0.828	0.838	0.828	0.839	0.839	0.839
Features from view #3	0.836	0.844	0.838	0.842	0.851	0.841	0.847	0.829	0.835
Features from view #4	0.824	0.833	0.826	0.843	0.851	0.842	0.839	0.844	0.841
Features from view #5	0.722	0.773	0.707	0.73	0.775	0.703	0.733	0.776	0.704
Features from view #6	0.769	0.792	0.754	0.793	0.81	0.779	0.788	0.808	0.782
Features aggregation	0.837	0.843	0.839	0.837	0.847	0.837	0.6575	0.768	0.678
Features merging	0.831	0.837	0.833	0.826	0.836	0.828	0.637	0.769	0.672

In accordance with the ten-fold cross validation, results are shown in the Table 2. First, features based on the former four views have good performance in predicting student dropouts. Through the discussion of the third part, we found that features based on former four views have a significant difference between participant and dropout students.

Second, the SVM has the best performance of all three models, while the other two models have similar accuracy. Therefore, the proposed features are not very sensitive to the models. Third, compared methods show no significant increase with SVF, while accuracy of SVM classifier on the compared methods model is not good. A possible reason is that more features, which add together, may decrease the distinction effect.

2) Results analysis and discussion of MVF

In accordance with the ten-fold cross validation, the results of MVF are shown in the Table 5 (MVF part). MVF has better effect than SVF, which is ensemble operation even with two kinds of features is better than simple dimension of features. MVF uses Native Bayes model as the basis model which can be extended to other models and multiple views of features.

Table 5: Performance of MVF and SCC

Kind of features	MVF			SCC		
	f_p	f_r	F	f_p	f_r	F
Features from view #1+#2	1	0.915	0.955	0.957	0.873	0.913
Features from view #1+#3	1	0.915	0.955	0.992	0.874	0.929
Features from view #1+#6	0.9815	0.915	0.947	0.954	0.887	0.914
Features from view #2+#3	1	0.915	0.955	0.995	0.875	0.932
Features from view #2+#6	0.9775	0.915	0.946	0.957	0.876	0.912
Features from view #3+#6	0.9815	0.915	0.947	0.951	0.876	0.912
Features aggregation	0.837	0.843	0.839	0.837	0.843	0.839
Features merging	0.831	0.837	0.833	0.831	0.837	0.833

3) Results analysis and discussion of SSC

Ten-fold cross validation is used to conduct the experiment for SSC, in which the training set with labeled samples and unlabeled samples is 1:9. Table 5 (SCC part) shows the results. Though SSC proposed method accuracy is lower than MVF, it still higher than that of SVF. This shows that SSC can be used in the system as a balance of prediction accuracy and cost.

6. Conclusions

In this paper, we first analyze the behavior patterns between dropout students and retention students to get multi-view features. Then, three detection algorithms are proposed based on the way multi-view features are used: SVF, MVF and SSC. SVF uses Naïve Bayes, Decision Tree, and Support Vector Machine to apply each type of feature to get detectors to find the detector with the best performance. MVF makes use of multi-view ensemble learning to use all types of features together. SSC applies semi-supervised co-training due to the lack of labeled data, which can be applied to the real environment.

This paper only presents some simple version of integration. More advanced machine learning models can be used to improve the prediction accuracy. Second, other kinds of

behavior can be used to get more relevant information for user modeling.

Acknowledgements

This work is partially supported by Youth Learner Program of Chongqing Social Science Association (2010QNRW54), The First Youth Backbone Teachers Scheme of Chongqing (Yu Renjiao 2011-65), and Fundamental Research Funds for the Central Universities (106112014CDJZR095502).

References

- [1]. Tang J. K., Xie H., and Wong T.-L. *A Big Data Framework for Early Identification of Dropout Students in MOOC*. in Technology in Education. Technology-Mediated Proactive Learning. Springer. 2015; 127-132.
- [2]. Xing W., Chen X., Stein J., and Marcinkowski M. *Temporal Predication of Dropouts in MOOCs: Reaching the Low Hanging Fruit Through Stacking Generalization*. Computers in Human Behavior. 2016; 58: 119-129.
- [3]. Hew K. F., Cheung W. S. *Students' and Instructors' Use of Massive Open Online Courses (MOOCs): Motivations and challenges*. Educational Research Review. 2014; 12: 45-58.
- [4]. Stein L. A. *Casting a Wider Net*. Science. 2012; 338(6113): 1422-1423.
- [5]. Nesterko S. Q., Dotsenko S., Q. Han, et al. *Evaluating the Geographic Data in Moocs*. In: Neural Information Processing Systems. 2013; 1-7.
- [6]. Yang D., Sinha T., Adamson D., et al. *Turn on, Tune in, Drop out: Anticipating Student Dropouts in Massive Open Online Courses*. Proc. Conf. on NIPS Data-Driven Education Workshop. 2013; 1-8.
- [7]. Tayloy C., Veeramachaneni K., U. M. O'Reilly U M. *Likely to Stop? Predicting Stopout in Massive Open Online Courses*. arXiv preprint arXiv. 2014; 1408(3382): 1-25.
- [8]. Jiang Z. X. *Learning Behavior Analysis and Prediction Based on MOOC Data*. Journal of Computer Research and Development. 2012; 614-628.
- [9]. Rosé C. P., Carolyn R., Penstein D., et al. *Social Factors that Contribute to Attrition in MOOCs*. Proc. Conf. on Learning@ Scale. 2014; 197-198.
- [10]. Seaton D. T., Nesterko S., Mullaney T., et al. *Characterizing Video Use in the Catalogue of MITx MOOCs*. EMOOCs, 2014; 140-146.
- [11]. O'Reilly U. M., Veeramachaneni K. *Technology for Mining the Big Data of MOOCs*. Research & Practice in Assessment. 2014; 9(2): 29-37.
- [12]. Demoncourt F., Taylor C. C, U. M. O'Reill, et al. *MoocViz: A Large Scale, Open Access, Collaborative, Data Analytics Platform for MOOCs*. NIPS Workshop on Data-Driven Education, Lake Tahoe, Nevada. 2013; 1-8.
- [13]. Allione G., and Stein R. M. *Mass Attrition: An Analysis of Drop out from Principles of Microeconomics MOOC*. The Journal of Economic Education. 2016; 47(2): 174-186.
- [14]. Rai L., and Chunrao D. *Influencing Factors of Success and Failure in MOOC and General Analysis of Learner Behavior*. International Journal of Information and Education Technology. 2016; 6(4): 262.
- [15]. Zheng S., Rosson M. B., Shih P. C., and Carroll J. M. *Understanding Student Motivation, Behaviors and Perceptions in MOOCs*. Proc. Conf. on Computer Supported Cooperative Work & Social Computing. 2015; 1882-1895.