Broad Learning System for Predicting Student Dropout in Massive Open Online Courses

Shuang Lai
School of Humanities, Economics and Law, Northwestern Polytechnical University
Xi'an, China, 710072
+86 13991366039
andy@nwpu.edu.cn

Yuxin Zhao
School of Humanities, Economics and Law, Northwestern Polytechnical University
Xi'an, China, 710072
+86 15319985665
zxy19990123@163.com

Yuqing Yang
School of Humanities, Economics and Law, Northwestern Polytechnical University
Xi'an, China, 710072
+86 15502929083
emilyus@126.com

ABSTRACT
At a time when the number of Massive Open Online Courses (MOOCs) users, courses and participating universities is increasing rapidly, a short-time training and reliable prediction model for MOOC extremely high dropout rates is needed. This paper proposes a MOOC dropout prediction model which is based on Broad Learning System (BLS) for MOOC dropout prediction. The model first maps the input into the feature node layer, then generates the enhanced node layer according to the feature node layer through activation, and finally performs linear transformation by combining the feature layer and the enhancement layer. The output layer is used for dropout prediction. Experiments are carried out on the dataset provided by KDD CUP 2015, and the experimental results show that the BLS significantly reduces training time and has a better prediction of dropouts than other mainstream research methods.

CCS Concepts
•Computing methodologies→Machine learning→Machine learning approaches→Neural networks •Applied computing→Education→E-learning

Keywords
MOOC, Dropout, Broad-Learning System, Big Data, Machine Learning

1. INTRODUCTION
In the first part of the article, we introduce the MOOCs briefly and talk about the high dropout rate, which is one of the biggest problems MOOCs face now. Then, we collect and summarize the related work done by peers to predict and research this problem.

1.1 Overview of MOOCs and Dropout Rate
MOOCs, launched in 2012 as a new type of "Internet plus Education", now have about 101 million learners. MOOCs not only have the characteristics of "massive", "open" and "online" contained in its name but also have many other traits that distinguish them from traditional school education, including low cost, low risk, and high flexibility.

Although these characteristics indicate the reform approach of the combination of Internet and education, its uniquely open and flexible teaching mode, which is different from traditional education, is difficult to control and supervise learners. So, MOOCs' model leads to some inevitable problems, most notably the high dropout rate of MOOCs.

In 2013, the University of Pennsylvania Graduate School of Education surveyed one million MOOC users. Only 4 percent of the students completed the entire course, and about half of the students took only one class. Ji Shisan, one of the founders of MOOC schools, revealed at Fudan University in 2014 that only about a quarter of the 130,000 MOOC users in China over the past year were able to eventually complete a MOOC course.

At the same time, MOOC teachers and designers are very concerned about how to improve the completing rate of MOOC courses. They hope to predict and intervene in dropout behaviors, analyze dropout reasons, and improve the quality of courses and online teaching methods. Therefore, the analysis and prediction of MOOC-based learner behavior are some of the important research directions in this field.

1.2 Antecedent Research
In order to solve the problem of MOOC prediction, many researchers have proposed different methods in recent years. Because the research data and research purpose of different studies are not completely consistent, the definition of MOOC dropout is not completely consistent among researchers. However, based on various studies, the definition of MOOC dropout can be uniformly defined as "students who have not participated in learning for a long time are regarded as dropouts".

In the following studies, different prediction methods and data are used to predict the dropout of MOOC. [1]Jin Zhiliao et al. proposed a mixture of global and local tensors to represent all available features, and predicted MOOC dropout by capturing the internal relationship between tensors. [2]Mi Fei et al. believed that dropout was a multi-definition problem, and used the methods of RNN and LSTM to predict MOOC dropout under different definitions. [3]Lu Xiaohang et al. constructed a sliding window model that can dynamically predict the dropout of MOOC. In these two following studies[4,5], the researchers used linear SVM and logistic regression to predict the last learning time of students based on feature extraction, and completed the prediction of MOOC dropout. [6]Wang Wei et al. proposed a framework structure combining CNN and RNN to predict MOOC dropout.

ICIET 2020, March 28–30, 2020, Okayama, Japan

ACM ISBN 978-1-4503-7705-8/20/03…$15.00
DOI: https://doi.org/10.1145/3395245.3395252

In the above studies, researchers paid more attention to how to increase the accuracy of prediction of MOOC dropout, but ignored the characteristics of fast data increment of MOOC and strong randomness of data dimension change. Although the sliding window model can dynamically predict MOOC data, training with LSTM still takes a lot of time and is difficult to expand further. Therefore, a model with fast training speed, easy data expansion and convenient model expansion is needed.

1.3 Related Work

Due to the extremely fast development of deep learning system, the current deep learning network needs to optimize a large number of parameters to be optimized, which usually requires a lot of time and machine resources to optimize.

Based on the research of Random Vector Function-Link Neural Network (RVFLNN), [7]Chen Junlong et al. created a Broad Learning System and an Increment of Broad Learning System to solve this problem. [8]Li Wang et al. used the Broad Learning System and the Broad Learning System with enhanced node increment to predict the toxicity of mushroom, which provided support for the Broad Learning System to respond quickly in the face of sparse data and data update, with the characteristics of high accuracy, fast training speed and strong scalability.

Due to the sparsity of MOOC data, large data increment, fast change of data dimension and consistency with the environment suitable for broad learning, this paper will use Broad Learning System and Increment of Broad Learning System to solve the dropout problem of MOOCs.

2. BROAD LEARNING SYSTEM IN MOOCS' DROPOUT

In this part of the article, we provide you a brief introduction to Broad Learning System (BLS). Then the Broad Learning System and Increment Broad Learning System have been shown here. Next, the experiment processes are offered step by step. The experimental results are talked about before we compare Broad Learning System with other methods at the end of this part.

2.1 Brief Introduction for Broad Learning System

The predecessor of the Broad Learning System is the Random Vector Function-Link Neural Network (RVFLNN)[7]. RVFLNN establishes the enhancement node by using the input data and then calculates the weight coefficient matrix between the input and output.

There are some differences in the BLS (see Figure 1). The BLS maps the input data into the sparse feature graph through sparse self-coding, constructs feature nodes, and determines the performance of the learning system based on the sparse feature graph. At the same time, feature nodes are used to generate enhancement nodes to further reduce the correlation between features. Based on the relationship between the sparse feature graph, the enhanced nodes and outputs, the weight coefficient matrix between them is calculated by the ridge regression learning algorithm.

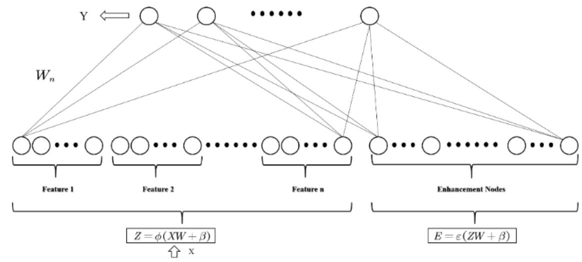

Figure 1. Broad Learning System Structure

2.2 Model in Broad Learning System

In order to predict MOOCs' dropouts, BLS is used to construct the coefficient characteristic graph of collected MOOCs' data. Initially, we take the dataset $X \in R^{N\times M}$, which contains N groups of data, and each group of data has M feature tags, which change into the feature node layer through mapping, also known as feature mapping set Z. The feature mapping formula is

$$Z = \phi(XW + \beta) \tag{1}$$

and W is all the connection weights between the feature map Z and input set X , β represents the offset of each set of mappings.

$$\begin{aligned} Z^n &= [\phi_1(x_1W_1+\beta_1), \phi_2(x_2W_2+\beta_2), ..., \phi_n(x_nW_n+\beta_n)] \\ &= [Z_1, Z_2, ..., Z_n] \end{aligned} \tag{2}$$

Then, we use the established feature node-set Z to generate enhanced node-set E,

$$E = \varepsilon(ZW + \beta) \tag{3}$$

here W is the connection weight and β is the bias value:

$$\begin{aligned} E^t &= [\varepsilon_1(Z_1W_1+\beta_1), \varepsilon_2(Z_2W_2+\beta_2), ..., \varepsilon_n(Z_nW_n+\beta_n)] \\ &= [E_1, E_2, ..., E_n] \end{aligned} \tag{4}$$

In order to avoid the correlation between the input of enhancement nodes, W and β are generated randomly.

The combination of the feature node layer and enhanced node layer can obtain the output set A, so the output can be expressed as $A = [Z^n | E^t]$. So, the output set Y can be obtained, where W represents the output weight.

$$\begin{aligned} Y &= [Z_1, Z_2, ..., Z_n | E_1, E_2, ..., E_t]W \\ &= [Z^n | E^t]W \\ &= AW \end{aligned} \tag{5}$$

$\varepsilon(x)$ and $\phi(X)$ are the activation functions. Because BLS needs to consider the correlation of parameters, the interdependence of parameters and the interference needs to be reduced, we choose Rectified Linear Units (ReLU) function (see Figure 2) as $\varepsilon(x)$ and $\phi(X)$. The ReLU function is used for unilateral suppression and sparsity to find suitable feature node Z and enhanced node E, so as to reduce the training time and alleviate the overfitting problem.

$$\mathrm{Re}LU(x) = \max(0, x)$$

$$\mathrm{Re}LU(x) = \begin{cases} x, \ if\ x > 0 \\ 0, \ if\ x \leq 0 \end{cases} \tag{6}$$

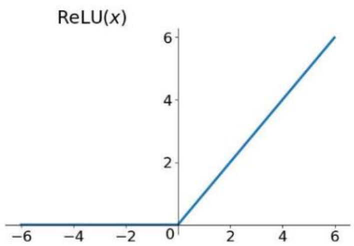

Figure 2. ReLU Function

The output weight W can be solved by the loss function of $Y = AW$,

$$\min(W) = \min \| AW - Y \|^2 + w \| W \|^2 \tag{7}$$

where ω represents the constraint value of the sum of W square weight.

2.3 Model in Increment of Broad Learning System

As the product of the combination of big data and education, MOOCs have an extremely fast updating speed of behavioral information. A large amount of new data will be generated randomly in each cycle, and even new features need to be added to the original data to change the dimension of the whole data. Therefore, it is difficult for BLS to use a fixed model to process MOOCs' data. As every time new data and new features are added, the weights need to be updated and the pseudo-inverse of the previous matrix is calculated. When the amount of data is too large, it is hard to directly calculate the pseudo-inverse of the matrix, and the weight cannot be updated in real-time, which greatly reduces the reliability of BLS. Therefore, when new data and new features are added, it is necessary to consider the increment of input data and features, and continuously update the weight dynamically (see Figure 3).

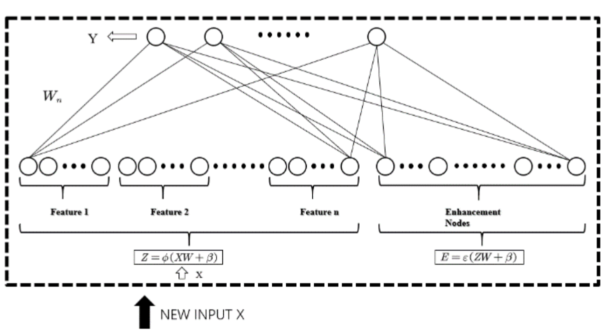

Figure 3. Broad Learning System with increments (increasing input)

When there is new data X_{n+1} input, the input matrix A_{n+1} corresponding to the increased data can be expressed as $A_{n+1} = [Z_{n+1} | E_{n+1}]$. The updated input matrix A can be represented as $A = \begin{bmatrix} A_n \\ A_{n+1} \end{bmatrix}$. Therefore, the updated output W is

$$Y = \begin{bmatrix} A_n \\ A_{n+1} \end{bmatrix} W \tag{8}$$

According to the method of calculating the input weight W , it can be known that

$$W = A^{-1} Y \tag{9}$$

The pseudo-inverse A^{-1} of the input matrix is

$$A^{-1} = [A_n^+ - BD^T, B] \tag{10}$$

where

$$D^T = A_{n+1} A_n^+$$

$$B = \begin{cases} (C^T)^+, \ if\ c \neq 0 \\ A_n^+ D (I + D^T D)^{-1}, \ if\ c = 0 \end{cases} \tag{11}$$

$$C = A_{n+1} - D^T A_n$$

So, we get our new input weight

$$W_{n+1} = W_n + B(Y - A_{n+1} W_n) \tag{12}$$

and the weight value W is updated continuously and dynamically while the input matrix before the input of new data is inversed for many times.

2.4 Experiment

2.4.1 Dataset

Since most MOOC behavior datasets are not open source, we choose to use public datasets. In this article, the MOOC dropout prediction dataset used for training is collected by KDD CUP in 2015. The dataset collects one month learning information of 39 courses which started within half a year.

This dataset is provided by China XueTangX MOOC institutions and includes four training datasets, which are the object.csv, enrollment_train.csv, log_train.csv, true_trian.csv.

The data set structure diagram is as follows:

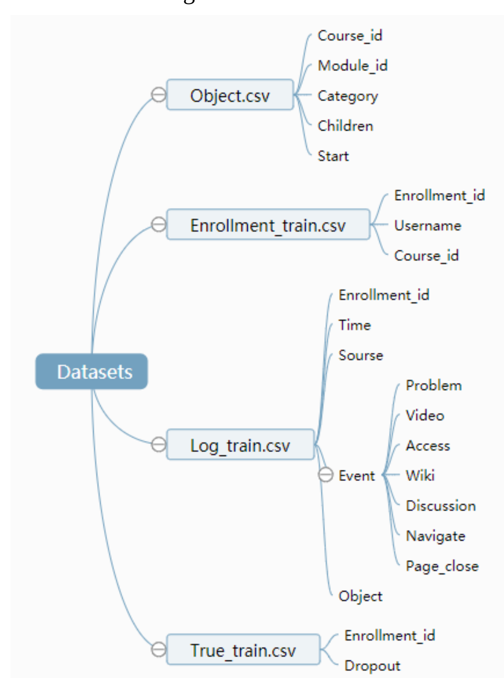

Figure 4. Dataset Structure

After data cleaning and summarizing, 120542 data samples are obtained. In all the experiments, the data are numerically processed, where the indicator {0,1} is used to indicate whether the student skipped a certain course, and the remaining characteristics are similar and expressed by numerical values.

Table 1. Dataset Explanation

Feature	Explanation
Module_id	Module number in a course (different part of each course)
Category	Module type
Children	Each Modules' child module
Time	The time of each event/operation
Source	Event source, whatever severs or browsers
Problem	The number of times you do your homework
Video	The number of times you watch video
Access	The number of times a class object other than video and assignment is read
Wiki	Check Wikipedia of the courses number
Discussion	Number of discussions in BBS
Navigate	Number of times to view other parts of the course
Page_close	Number of closed pages

According to the distribution of students' behavior logs in the training set and test set provided by the data set, students' dropout behavior can be divided into three periods, separately the fall semester, winter vacation and spring semester. Almost all students do not study during the winter break, and between the spring and fall semesters, dropout rates are lower and most students are more active, but many students still dropout. In the process of data cleaning, we clean up the winter holiday data, since the proportion of the total number of dropouts has been stable. At the same time, the data of training set and test set are compared. The data distribution of the two sets is highly similar, and the dropout rate is also relatively similar.

2.4.2 Experiment

In this section, we explore the use of BLS to predict MOOC dropouts. BLS, with its excellent performance and ability to scale and change easily, has been used in some cases to replace deep learning in data mining. Since BLS has good adaptability to sparse data and stronger containment ability to incremental data, BLS is used for sparse data with fast data update speed. These features make it possible to monitor MOOC platforms in real-time and make dropout prediction. In order to further prove the effectiveness of BLS in MOOC dropout prediction, we use BLS to train and test the processed dataset, and then compare the results of BLS with those of several machine learning algorithms[2,4,5,9] tested with the same dataset.

This test runs in a Windows10 (64-bit), Intel i7-7700HP, 2.8ghz CPU, GTX1070(8GB) and 16GB RAM laptop Python3.7 environment.

The schematic diagram of broad learning and training process is as follows:

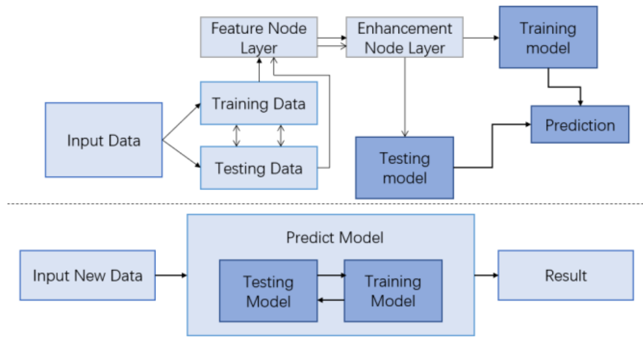

Figure 5. Broad Learning System Training Process

After data pretreatment, 70% (84379) samples from the sample set are randomly selected as the training set and the remaining 30% (36163) samples are selected as the test set. And the classified dataset is as input to the BLS.

Before training the BLS model, the number of feature nodes and the number of enhancement nodes are first manually set for the BLS model. Set 10 groups of mapping features and 10 groups of enhancement nodes. Set 10 nodes for each mapping features and set 100 nodes for each enhancement nodes, which means we set 100 mapping feature and 1,000 enhancement nodes. When the input Z of mapping nodes of 84,379*100 is obtained, the input H of enhancement nodes needs to be obtained. At the same time, in order to prevent the correlation between the input of enhancement nodes, a random weight of 1,000*1,000 is needed for control. The final input $A=[Z|H]$ is 84,379*1,100.

Figure 6. Broad Learning System Training Folds

At the same time, in order to ensure the reliability and accuracy of the experimental results, a 10-fold cross-validation method (see Figure 6) is adopted in the training model to divide the test data into 10 parts, and no repeated sampling is used. Each sample point has only one chance to be included in the training set or test set, and the training set is evenly distributed in turn, and take the average of the accuracy obtained from 10 times of verification as the final accuracy.

The following table shows the comparison of accuracy and training time in the methods of the BLS and Logistic Regression[4], the linear SVM[5], Gradient Boosting[9], the LSTM[10].

The experimental results (see Table 2) can be obtained from the above table. Compared with Logistic Regression, Gradient Boosting, LSTM, and Linear SVM, the BLS is not the most accurate (the accuracy of BLS is 92.73%, second only to 93.66% of LSTM), but the BLS has the fastest training time -- the BLS can complete more than 2 hours of training of LSTM or Linear SVM within 30 seconds. Therefore, it can be concluded that BLS has strong adaptability to the current data set and can make relatively accurate predictions.

Table 2. Comparison for Models

Models	Average Training Time（s）	Average Accuracy (%)
Broad Learning System	28.4	92.73
Logistic Regression[4]	172	89.17
Linear SVM[5]	26260	88.91
Gradient Boosting[9]	549	89.61
LSTM[10]	6104	93.66

While comparing with other algorithms, it is also necessary to consider the fact that the MOOC platform is fast in updating data and can generate a large amount of new data at any time. Therefore, it is also necessary to conduct a data update experiment and use the Increment of BLS to train the ever-expanding training data volume. Since other control models cannot update weights in real-time during incremental training of data, the process of model training and parameter optimization needs to be redone, which takes too long. Therefore, other models are not used for comparison. The figure below is the training result of the Increment of BLS when input data increments.

The experimental results of data increment are as follows. It can be seen that the accuracy of Increment of BLS is significantly improved when conducting data increment test. The training time of Increment of BLS is increased due to data increment, but the training time of it is still much less than that of deep learning method.

Table 3 Input Data Incremental Test Results

Method	Input Data and Incremental Data	Accuracy (%)	Time(s)
BLS	108488	91.887	41
Increment of BLS	108488→206128 97640	93.883	166.2
Increment of BLS	206128→294004 87876	94.218	182.3
Increment of BLS	294004→373093 79089	94.396	199.35
Increment of BLS	373093→444274 71181	94.52	210.82

According to the above experiments, BLS has a relatively outstanding prediction accuracy in the process of using KDD CUP 2015 data to predict MOOC dropout. It consumes a very short training time, and the training cost is lower than other relevant studies.

And, when there is a lot of new data into the training data set because the BLS for the first time after the ridge regression to calculate weight, we only need to use the incremental calculation method, calculation of weight increment and update the weights to complete the updated model. This method can effectively compress the training time, at the same time because of the increasing scale of the test data, the prediction accuracy is improved.

3. CONCLUSION

In this paper, we propose a Broad Learning System for predicting MOOC dropouts. BLS has a fast training speed in MOOC dropout prediction, which can form the dropout prediction of MOOC courses in a short period of time, facilitating course evaluation and other real-time operations. When new data is added, incremental calculation is adopted, which requires a small amount of calculation to update the original weight. It has good adaptability to MOOCs with large data output, multiple data dimensions and fast data changes. By testing the KDD CUP 2015 dataset, this paper shows that the performance of BLS can be comparable to other methods proposed in related studies.

In future, we plan to further apply the BLS to more MOOC data to further evaluate the effectiveness of the model in predicting MOOC dropouts, and we will further investigate the ease with which the BLS can be extended or combined with deep learning features which are not mentioned in this article.

ACKNOWLEDGMENT

This study is supported by the 2019 College Students Entrepreneurship Training Program in Shaanxi Province, China – An Efficient and Intensive Information Service Platform for Universities. We would also like to thank KDD CUP for collating the MOOC datasets provided by XuetangX for their help in this study.

4. REFERENCES

[1] Liao, J. , Tang, J. , & Zhao, X. . (2019). Course drop-out prediction on mooc platform via clustering and tensor completion. Tsinghua Science and Technology, 24(4), 412-422.

[2] Fei, M. , & Yeung, D. Y. . (2015). Temporal Models for Predicting Student Dropout in Massive Open Online Courses. 2015 IEEE International Conference on Data Mining Workshop (ICDMW). IEEE.

[3] Lu Xiaohang, Wang Shengqing, Huang Junjie, Chen Wenguang, & Yan Zengwang. (2017). A sliding window model based moocs dropout rate prediction method. Data analysis and knowledge discovery, 1(4).

[4] Colin Taylor, Kalyan Veeramachaneni, and Una-May O'Reilly. 2014. Likely to stop? Predicting stopout in massive open online courses. arXiv preprint arXiv:1408.3382 (2014).

[5] Marius Kloft, Felix Stiehler, Zhilin Zheng, and Niels Pinkwart. 2014. Predicting MOOC dropout over weeks using machine learning methods. In Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs. 60–65.

[6] Wang, W. , Yu, H. , & Miao, C. . (2017). Deep Model for Dropout Prediction in MOOCs. the 2nd International Conference.

[7] Chen, C. L. P. , & Liu, Z. . (2017). Broad Learning System: an effective and efficient incremental learning system without the need for deep architecture. IEEE Transactions on Neural Networks and Learning Systems, PP(99), 1-15.

[8] Li Wang, Yu Zhuliang. Application of width learning system in mushroom toxicity discrimination [J]. Modern food

science and technology,2019,35(07):267-272+54. (in Chinese with English abstract)

[9] Nagrecha, S. , Dillon, J. Z. , & Chawla, N. V. . (2017). MOOC Dropout Prediction: Lessons Learned from Making Pipelines Interpretable. WWW 2017. International World Wide Web Conferences Steering Committee.

[10] Luo Ye, Li A'li. Analysis of reasons and countermeasures for a high MOOC dropout rate [J]. Theory,2015,(5):139-140. DOI:10.3969/j.issn.1002-2589.2015.05.064.