IJCSN - International Journal of Computer Science and Network, Volume 8, Issue 3, June 2019
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Dropout Prediction Using Multi-Classifier Prediction and Score Integration Method

[1] Rakhi Viswan; [2] Deepa Rajan S; [3] Dr Radhakrishnan B

[1]CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

[2]CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

[3]CSE Department, BMCE
APJ Abdul Kalam Technological University, Sasthamcotta, Kerala, India

Abstract -A massive open online course is an online course aimed at unlimited participation and open access via the web. In addition to traditional course materials, such as filmed lectures, readings, and problem sets, many MOOCs provide interactive courses with user forums to support community interactions among students, professors, and teaching assistants(TAs), as well as immediate feedback to quick quizzes and assignments. MOOCs are a recent and widely researched development in MOOC represents an ultimate way to deliver educational content in higher education settings by providing high-quality educational material to the students throughout the world. In this paper, system proposes a model to predict the MOOC dropout by considering the events or features of the MOOC course .The proposed approach is designed to estimate the student dropout rate using a neural network ensemble model. The neural network ensemble model is an architecture designed to process the data parallel with two different mechanisms namely MLP, Recurrent Neural Network. The proposed system takes students activities such as problems, videos, seminars etc as input. The Proposed system offers better accuracy which is more than the current systems.

Keywords- *MOOC drop out, MLP, Recurrent neural, Ensemble model*

1. Introduction

A Massive Open Online Course (MOOC) is a large-scale web-based course that is offered to a very large number of participants. In the last few years, MOOCs have seen an increase in popularity due to the emergence of several well-designed online-education websites, such as edX and Coursera, and rising interest from top universities, such as MIT, Stanford and UC Berkeley, to open a variety of their courses to the wider public. The appeal for end consumers lies in the accessibility of high-quality education from any location regardless of personal background. MOOCs typically contain short lecture videos(10-15 minutes), as well as quizzes and homework assignments to assess students' understanding of subject matters.

Massive Open Online Courses (MOOCs) provide a unique opportunity to learners and students. Platforms such as edX and Coursera allow learners around the world to access educational content from a wide variety of universities. Massive Open Online Course (MOOC) has been revolutionizing the way people getting education. However, it also raises up concern that MOOC has very high dropout rate relative to traditional classes. An accurate prediction of dropout becomes very important because it can help MOOC course developers to adaptively tune web designs to students with high dropout rate. Massive Open Online Courses (MOOCs) open a number of educational opportunities for traditional and non-traditional learning. However, the size of classes, which easily reaches into the thousands of students, requires educators and administrators to reconsider traditional approaches to instructor intervention and the manner in which student engagement, motivation, and success is assessed, especially since attrition rates in MOOCs is notoriously high. The uniqueness of MOOCs and the difficulties associated with them has opened new research areas, especially in predicting or explaining completion rates and general student success. Research has mainly focused on predicting success using click-stream data (i.e., student interactions within the MOOC software). Other recent approaches include the use of Natural Language Processing (NLP) tools to gauge students'

affective states, measure the sophistication and organization of students' discourse within a MOOC , and a combination of click-stream and NLP data . MOOCs have become an important component of education research for both instructors and researchers because they have the potential to increase educational accessibility to distance and lifelong learners.

Current research and practices in building prediction models in MOOCs follow three interrelated directions: fixed term dropout prediction model, temporal dropout prediction model, and dropout prediction performance optimization. These methods have serious limitations for deployment in MOOCs overall. Fixed-term dropout prediction model can identify all the at-risk students at once. However, given the large number of dropout students in MOOCs, the instructors cannot provide effective feedback to that many at-risk students at once. The temporal dropout prediction model predicts students at risk of dropping out in the next week. While it tremendously reduces the number of at-risk students the instructor needs to deal with every time, it does not provide any clue for the instructor to offer personalized intervention. The result is that all the students receive the same intervention without any personalization. Much effort on optimizing prediction performance has also been made. However, no studies so far take the advantage of deep learning which has shown great success in so many disciplines. In the next three sections, empirical studies for prediction model building are reviewed in more detail based on the proposed directions.

2. Research Background

VlatkoNikolovski et al[1] uses various data mining technique to predict dropout rate .Provide a predictive model which will identify a subset of students who tend to drop out of the studies after the first year. The classification task aims to identify a pattern among students who tend to drop out.The accuracy of the different classifier algorithms notably depends on the quality of the attributes extracted from the data. The accuracy of the classifiers is tied closely with the quality and sophistication of the data mode.

Jiajun Liang et al [2] using student's learning activities data in a period of time to measure how likely students would drop out in next couple of days.They collect 39 courses data from XuetangX platform, which is based on the open source Edx platform. Using supervised classification approach in the machine learning field, achieve 89% accuracy in dropout prediction task with gradient boosting decision tree model. They describe details in drop out prediction framework, including data extraction from Edx platform, data preprocessing, feature engineering and performance test on several supervised classification models.

Juho Kim1 Philip J. Guo2 et al[3]reports a large-scale analysis of in-video dropout and peaks in viewership and student activity, using second-by-second user interaction data from 862 videos in four Massive Open Online Courses (MOOCs) on edX. We find higher dropout rates in longer videos, re-watching sessions (vs first-time), and tutorials (vs lectures). Peaks in re-watching sessions and play events indicate points of interest and confusion. Results show that tutorials (vs lectures) and re-watching sessions (vs first-time) lead to more frequent and sharper peaks. Based on this observation, we identify five student activity patterns that can explain peaks: starting from the beginning of a new material, returning to missed content,following a tutorial step, replaying a brief segment, and repeating a non-visual explanation. Our analysis has design implications for video authoring, editing, and interface design,providing a richer understanding of video learning on MOOCs.

ErmanYukselturk, SerhatOzekeset al[4]predicts dropouts through data mining approaches in an online program. The subject of the study was selected from a total of 189 students who registered to the online Information Technologies Certificate Program in 2007-2009. The data was collected through online questionnaires (Demographic Survey, Online Technologies Self-Efficacy Scale, Readiness for Online Learning Questionnaire, Locus of Control Scale, and Prior Knowledge Questionnaire). The collected data included 10 variables, which were gender, age, educational level, previous online experience, occupation, self efficacy, readiness, prior knowledge, locus of control, and the dropout status as the class label (dropout/not). In order to classify dropout students, four data mining approaches were applied based on k-Nearest Neighbor (k-NN), Decision Tree (DT), Naive Bayes (NB) and Neural Network (NN). These methods were trained and tested using 10-fold cross validation. The detection sensitivities of 3-NN, DT, NN and NB classifiers were 87%, 79.7%, 76.8% and 73.9% respectively.

Zixun Yang et al[5]built model to predict dropout in Massive Open Online Course(MOOC) platform, hefirst explored different models and optimized parameters selection to reach best performance. With few models of good prediction on validation data, he ensembled them together using XGBoost Classifier to do a second level learning based on the first level training prediction. In addition, I implemented Long Short Term

Memory(LSTM) Recurrent Neural Networks with Keras on 30 days univariate Time series data and reached 0.857.

3. Design & Architecture

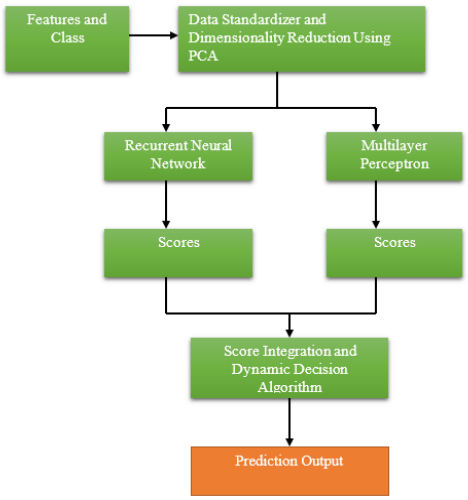

Fig.1.MOOC Proposed Model

The main objective of the proposed system is to find the student dropout from the Massive Open Online Corse (MOOC).Massive Open Online Courses (MOOCs) provide a unique opportunity to learners and students. An accurate prediction of dropout becomes very important because it can help MOOC course developers to adaptively tune web designs to students with high dropout rate. In the proposed model, system extracting the features from the data set, we need to do the feature selection, not only for training speed, but also improve the prediction accuracy. With the selected features, we started to find the dropout rate .This method mainly focus on Network Ensemble model to find the better result. Ensemble learning helps improve machine learning results by combining several different models. Generally Artificial Neural Networks are having high variance, frustrating when developing a new prediction model. To reduce variance, train multiple models instead of single model and combine their predictions. This is known as ensemble learning, which not only reduces variance but also result in better predictions than any single model. The proposed method uses, two neural networks such as Multi Linear Perceptron and Recurrent Neural Network for training. The extracted data from the given set is first standardized and then the input data will be feed into a MLP and Recurrent neural network with hidden layer and output layer. The outputs from both the networks are combined to produce more accurate prediction result.

3.1 Multi Layer Perceptron

A multilayer perceptron (MLP) is a deep, artificial neural network. It is composed of more than one perceptron. They are composed of an input layer to receive the signal, an output layer that makes a decision or prediction about the input, and in between those two, an arbitrary number of hidden layers that are the true computational engine of the MLP. MLPs with one hidden layer are capable of approximating any continuous function. Perceptron is a linear classifier; that is, it is an algorithm that classifies input by separating two categories with a straight line A perceptron a single output based on several real-valued inputs by forming a linear combination using its input weights (and sometimes passing the output through a nonlinear activation function.

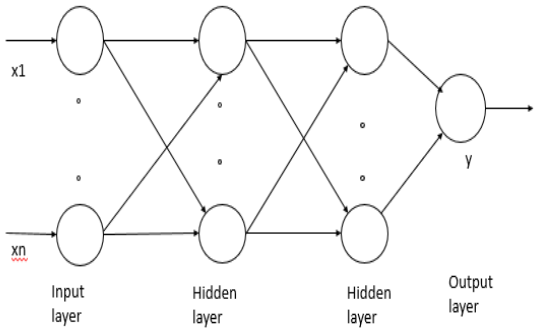

Fig.2.multilayer perceptron

4. Experimental Result

The experimental results of the dropout prediction obtained and discussed. The table 1 illustrates the comparison of the outcomes obtained by using previous methods and those by the present method. student features and events are used for training. For training here use multilayer perceptron and recurrent neural network. Recent results of KDD dataset, by fixing the test data at 20%, the remaining 80% training data. The result shows that the proposed model is better as compared with the previous models and also it outperforms with an accuracy of 96.2%.The proposed model incurs less processing times. The result shows that the proposed model is better as compared with the previous models.

Model	Accuracy
Ensemble model	96.2%
Decision Tree	79%
Naive Bayes	73.9%

5. Conclusion

Predicting dropout student is an important and challenging task for universities,policymakers and educators especially in online learning.Therefore,this study examined whether the use of ensemble model can be helpful in dealing with this problem in an online program.Here,the student activities are selected as input. In this study we use two neural network for training.This project worked on building models to get best of prediction about dropout in MOOC platform.For this different models,tuned the parameters and ensemble them together to reach good accuracy.The result shows that the proposed model is better as compared with the previous models.Also ,this model gives an accurate result.

Reference

[1] Vlatko Nikolovski, Educational Data Mining: Case Study for Predicting Student Dropout in Higher Education:A data mining approach.
[2] Jiajun Liang: Machine Learning Application in MOOCs:Dropout Prediction.
[3] G. Dekker, M. Pechenizkiy, J. Vleeshouwers, "Predicting students drop out: a case study", Educational Data Mining 2009, 2009.
[4] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, R. C. Miller, "Understanding in-video dropouts and interaction peaks in online lecture videos", Proceedings of the first ACM conference on Learning@ scale conference, ACM, pp. 31-40, 2014.
[5] Erman Yukselturk ,Serhat Ozekes : Predicting Dropout Student: An Application Of Data Mining Methods In An Online Education Program .
[6] Zixun Yang: MOOC Dropout Prediction.

Authors Profile

Rakhi Viswan received her B.Tech (CSE) degree from University of Kerala in 2017. She is currently pursuing her Masters in Computer Science & Engineering from KTU.

Deepa Rajan S is working as Assistant Professor in Computer Science and Engineering Department. She has 10 years' experience in teaching. Her research interests focuses on Data Security, Image Processing.

Dr.Radhakrishnan B is working as the Head of CSE Department. He has more than 14 years experience in teaching and has published papers on data mining and image processing. His research interest includes image processing, data mining and image mining.