An Approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN)

Arindam Mondal

School of Education Technology
Jadavpur University, India

Joydeep Mukherjee

School of Education Technology
Jadavpur University, India

ABSTRACT

Educational Data Mining able to gain a handsome amount of attention of the researcher of educational technology in recent times. In this paper, Recurrent Neural Network (RNN) is used to predict a student's final result. RNN is a variant of neural network that can handle time series data. The final term class is predicted using the first and second term class along with fifteen others features of a student. This analysis help the teacher to identify the students, who are 'at risk' and based on that he can offer proper remedy to them. In this paper, a comparison based study is also made with Artificial Neural Network and Deep Neural Network with the proposed Recurrent Neural Network.

Keywords

Educational Data Mining, Recurrent Neural Network (RNN), Artificial Neural Network (ANN), Deep Neural Network (DNN)

1. INTRODUCTION

In the field of educational data mining, Academic students' performance is one of the most important topics. It can be a very useful factor for both students and teachers. Data mining concept is basically to extract the meaningful information from the raw data and the Educational Data Mining (EDM) is an emerging discipline which extracts meaningful knowledge in the context of educational field. EDM has contributed to theories of learning investigated by researchers in educational psychology and the learning sciences. The field is closely tied to that of learning analytics.

Many different machine learning approaches are there to predict a student's performance using the student's family details as well as educational background. Particularly, these machine learning method enable to detect the student, who is at the risk i.e. who is likely to fail or drop out from the class. This helps a teacher to provide a remedy to the weak student. It can also help to identify the high performer student in the class.

Decision Tree [1] and Naive Bayes [2] algorithm is highly used in Educational Data Mining. Havan Agrawal [3] stated limitation of these algorithms, the accuracy of the Bayesian classification model reduces when input is provided in a continuous range.

Artificial Neural Network is a powerful tool to make a prediction. Deep learning is the state of art [4] for artificial intelligent research. Currently deep learning concept is highly used in voice/sound recognition [5], Natural Language Processing [6], computer vision [7].

In this paper, we proposed a Recurrent Neural Network (RNN) classifier model to predict the students' performance. RNN is a variant of a neural network. It can obtain the output value based on the past and the current information using the recurrent loop of the network. The proposed RNN model aims to predict a student category through logistic classification analysis. One hidden layer with Relu activation function is implemented. The proposed model is able to predict a student performance with the accuracy of around 85%.

2. RELATED WORK

Kyndt et al.[8] used neural network to predict academic performance of the first year bachelor degree student, based on students' motivation, approaches to learning, working memory capacity and attention.

Ioannis E. Livieris, et al. [9] in their study predicted the performance of students in Mathematics using an Artificial Neural Network (ANN) classifier. They found that the modified spectral Perry trained artificial neural network performs better classification compared to other classifiers in this context.

S. Kotsiantis, et al. [10] investigated in distance learning using machine learning techniques for dropout prediction of students. Their study made an important contribution as it was a pioneer and helped to carve the path for educational data mining.

Moucary, et al. [11] used a hybrid technique on K-Means Clustering and Artificial Neural Network for predicting the performance of students of higher education of new foreign language. They used Neural Network to predict the student's performance and then fitting them in a particular cluster using the K-Means algorithm. This clustering helped the instructors to identify a student's capabilities during their early stages of academics.

Mukta and Usha[12] predicted the academic performance of business school graduates using neural networks and traditional statistical techniques They identified the underlying constructs of a traditional business school curriculum and presented its relevance with the various elements of admission process .

Amrieh, et al. [13] proposed a prediction model for students' performance based on data mining methods. The model was evaluated in three different classifiers; Naïve Bayesian, Artificial Neural Network and Decision tree. The classifier's performance was improved by using random forest, bagging and boosting as an ensemble method. The model achieved up to 22.1% more in accuracy compared when behavioural features were removed and accuracy is increased up to 25.8% after using the ensemble methods.

3. STATE-OF-THE-ART
3.1 Artificial Neural Network

Artificial neural networks (ANNs) are computing systems vaguely inspired by the biological neural networks. An ANN is based on a collection of connected units or nodes called artificial neurons. A signal is transmitted through the connections between artificial neurons from one to another.

ANN is an adaptive system that adjusts its structure based on external or internal information that runs through the network during the learning process. Neural networks can be considered as non-linear numerical data modelling tools. ANN has been effectively applied in the area of prediction, handwritten character recognition.

The architecture of the neural network consists of layers. Layers are groups of neurons that perform similar functions. There are basically three types of layers.

1. The input layer is the layer of neurons that receive input from the user program.

2. The layer of neurons that produce output is the output layer.

3. Hidden layers located between input and output layers. Hidden layer neurons are only connected only to other neurons and never directly interact with the user program.

The hidden layer is optional. The input and output layers are required.

3.2 Recurrent Neural Network

A Recurrent Neural Network (RNN) is a class of artificial neural network where connections between units form a directed graph along a sequence. It is mainly used to handle the time series data. RNN is to make use of sequential information. RNNs are called recurrent because they perform the same task for every element of a sequence. RNN can be described in other way that they have a "memory" which captures information about what has been calculated so far. Figure 1 (a) shows a graphical illustration of a structure of an RNN. Figure 1 (b) shows the unfolding in the time of the computation of an RNN
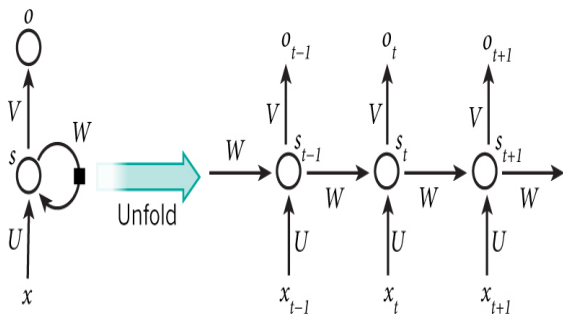

Fig 1(a)   **Fig 1(b)**

Fig 1: Structure of the Recurrent Neural Network

The forward pass of the RNN is almost same as that of an MLP with single hidden layer except that activation arrives at the hidden layer from both the current external input and the hidden layer activations one step back in time.

For the input to hidden units we have :

$$a_h^t = \sum_{i=1}^{I} w_{ih} x_i^t + \sum_{h=1}^{H} w_{\hat{h}h} b_{\hat{h}}^{t-1}$$

$$b_h^t = \theta_h(a_h^t)$$

For the output unit we have :

$$a_k^t = \sum_{h=1}^{H} w_{hk} b_h^t$$

The back propagation of the RNN is just the standard back propagation. The complete sequence of delta terms can be calculated by starting at t = T and recursively applying the below functions, decrementing t at each step.

Note that $\delta_j^{T+1} = 0$, $\forall j$, since no error is received from beyond the end of the sequence

$$\delta_h^t = \theta^j(a_h^t)\left(\sum_{k=1}^{K} \delta_k^t w_{hk} + \sum_{h'=1}^{H} \delta_{h'}^{t+1} w_{hh'}\right)$$

$$\delta_h^t = \frac{\partial O}{\partial a_j^t}$$

4. DATA PROCESSING

The data set is collected from the below mentioned website - https://www.kaggle.com/aljarah/xAPI-Edu-Data to predict the student's performance. It is an educational data set collected from a learning management system. The data set is little bit updated by adding two more features with the existing data set. The data set contains record for total 480 students. The modified data set contains total of 17 features as below.

Table 1. Attribute Details

Attribute	Data Type	Distinct Values
Gender	Text	2
Nationality	Text	14
Place of Birth	Text	14
Stages	Text	3
Grades	Text	12
Section ID	Text	3
Topic	Text	12
Parent Responsible	Text	2
Raised Hand	Numeric	0-100
Visited Resources	Numeric	0-100
Announcements View	Numeric	0-100
Discussion	Numeric	0-100
Parent Answering	Text	2
Parent Satisfaction	Text	2
Student Absence Day	Text	2
Class in 1st Term	Text	3
Class in 2nd Term	Text	3

Using the above features the main goal is to predict the final terminal class of a student. The class of a student can be three types which are based on the marks of the student. Here it is considered that the mark in range 80%-100% is class 'High',

mark in range 60%-80% is class 'Medium' and mark below 60% is 'Low'.

Data processing is required to improve the quality of the dataset. Text data type attributes Gender, Parent Responsible,

Parent Answering, Parent Satisfaction ,Student Absence Days are transformed to binary data '0' and '1'. Other Text data type attributes Nationality, PlaceofBirth, Stage, Grades, SectionID, Topic, Class in 1st Term and Class in 2nd Term are transformed to numerical data type

5. METHODOLOGY

In this paper, a modified vanilla RNN is used to make the prediction of a student's performance. After building the proposed system, a comparative study is made with the Deep neural network (DNN) which is proposed by the Prabu P, et al.[14]. The whole implementation is done in Matlab 2015a.

Once the data pre-processing is completed, the whole data set is divided into two parts. One is training set and another one is the testing set. The data set is divided into the ratio of 3:1(Train/Test). The class variable is one-hot encoded for both train and test dataset. The class variable is represented in the numerical format as below:

Table 2. One-Hot-Encoding

Classes	One-Hot Encoding Format
Low	[1,0,0]
Medium	[0,1,0]
High	[0,0,1]

The proposed neural network contains three layers. The first layer is the input layer and the seventeen features are given as input in the system. The final result of the student is produced as output from the output layer. The hidden layer is located in between the input and the output layer. The hidden layer consists of 40 neurons. One recursive path is there in between the hidden layer neurons to make the neural network recurrent. Random initialization of weights w and bias b to every interconnected layers. This initialization is specifically proposed for hidden nodes with Rectified Linear Unit (ReLU) non linearity. The ReLu activation function is defined as the positive part of its argument. It is represented by

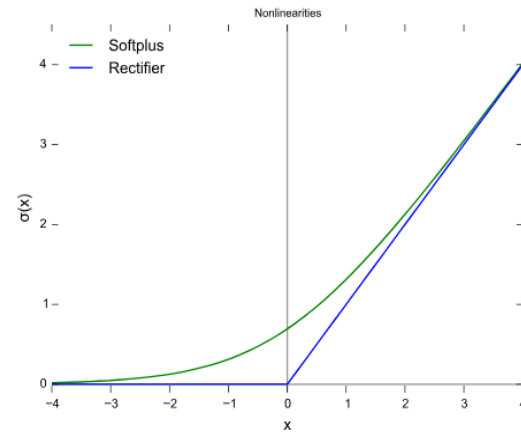$f(x) = x+ = max (0,x)$ *Where x is the input to a neuron.*

Fig 2: Plot of Rectifier Activation Function

The output is then passed to the cost function where output is compared with the actual input. The optimization function updates the weights of layers so that the cost function returns a lesser error value. This model is iterated 200 times in order to get the higher accurate value.

6. RESULT

The proposed method is applied to a dataset of 480 students. 360 data is used for training and the rest 120 data is used for testing. The proposed RNN is able to achieve the accuracy of 85.4% in predicting of a student's final grade. The performance of the training data is determined using the mean square error. After completing the 200 epochs, the mean square error of the training data set is determined as 0.11. The training performance is represented in the below Figure 3.
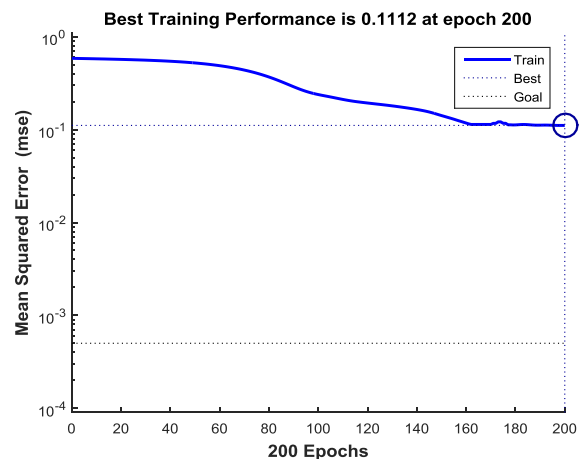

Fig 3: Training Performance

The number of neurons in the hidden layer is taken as 40. The network is trained using the Relu function. The different errors histogram with 20 bins is represented in the Figure 4.
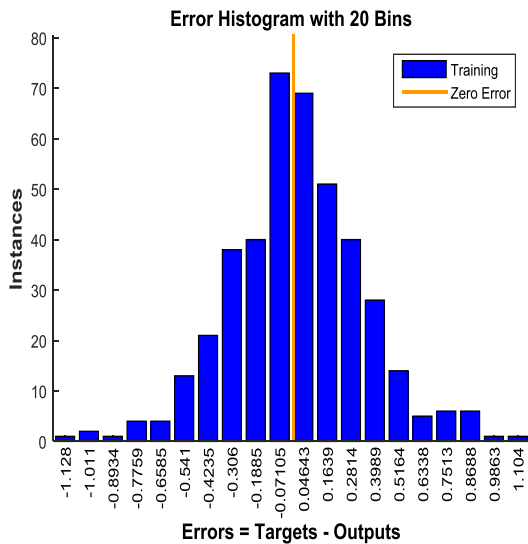
Fig 4: Error Histogram

In the proposed model, the number of epoch is considered as 200. An analysis should be done for defining epoch as high number of epoch leads to over-fitting of model and less epoch leads to under-fitting of model.

A comparison based study is made to Prabu P, et al.[14] proposed model. In their paper, Deep Neural Network (DNN) model was suggested by them. Our proposed model is compared with DNN and ANN using the same data set. The prediction accuracy is calculated and shows the comparisons result in the Figure 5.
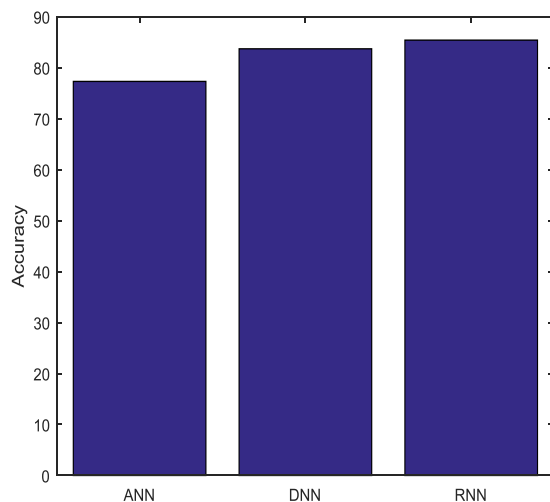

Fig 5: Comparion of Accuracy Among Different Techniques

Table 3. Classification Method Comparison

Classifier	Accuracy (%)
Artificial Neural Network	77.8
Deep Neural Network	83.7
Recurrent Neural Network	85.4

The figure shows that our proposed Recurrent Neural Network method accuracy is 85.4%. The other techniques like Artificial Neural Network and Deep Neural Network values are less than the proposed Recurrent Neural Networks. That is why, it can be concluded that the proposed model works better than the others.

7. CONCLUSION AND FUTURE WORK

In this paper, a method is presented to predict a student's performance using Recurrent Neural Network with Rectified Linear Unit (ReLU) activation function. RNN is suitable for analysis of the time series data. In this paper, student's final term class is predicted using first two term class along with other fifteen features. Here it is assumed that the student's final class is a sequential output of other two term class. The proposed model achieved an accuracy of 85.4%, which is outperformed the other machine learning algorithm. Higher accuracy can be obtained with larger dataset and features. This model can be used to predict student's performance and help to identify the students who have higher chance of failing.

As a part the future work, the model can be tested with larger data set by adding more features. A clustering technique can be used to categorize the students at the very early phase, which will help to deal with the cold start problem. The proposed method can also be used in the intelligent tutorial system (ITS) to predict a student performance and based on that ITS can provide proper feedback to the students.

8. REFERENCES

[1] P. Saini , A.K. Jain, "Prediction using Classification Technique for the Students' Enrollment Process in Higher Educational Institutions", International Journal of Computer Application, Springer, Berlin, Heidelberg, 2013, 84–89.

[2] M. Koutina, K.L. Kermanidis, "Predicting Postgraduate Students' Performance using Machine Learning Techniques", in Artificial Intelligence Applications and Innovations, Springer, Berlin, Heidelberg, 2011, 159–168.

[3] H. Agrawal , H Mavani, "Student Performance Prediction using Machine Learning", in International Journal of Engineering Research and Technology, 2015, 271–280.

[4] Y. LeCun, Y. Bengio and G. Hinton, "Deep Learning", Nature, 521(7553), 436–444.

[5] G.E. Dahl, D. Yu, L. Deng and A. Acero, "Context-Dependent Pre-trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE Transactions on audio, speech and language processing, 20(1), 2012, 30–42.

[6] R. Collobert , J. Weston, "A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning", in proceedings of the 25th International Conference on Machine Learning, 2008, 160–167.

[7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Convolution Architecture for Fast Feature Embedding", in proceedings of the 22nd ACM Intenational Conference on Multimedia, 2014, 675–678.

[8] E. Kyndt, M. Musso, E. Cascallar and F. Dochy,

"Predicting Academic Performance in Higher Education: Role of Cognitive, Learning and Motivation," Earli Conference, 2011.

[9] Livieris, et al, "Predicting Students' Performance using Artifical Neural Networks", 8th PanHellenic Conference with International Participation Information and Communication Technologies, 2012, 312–328.

[10] S. Kotsiantis, et al, "Preventing Student Dropout in Distance Learning Systems using Machine Learning Techniques", Applied Artifical Intelligence, 18(5), 2003, 411–426.

[11] C.E. Moucary, M. Khair and W. Zakhem, "Improving Student's Performance using Data Clustering and Neural Networks in Forgein-Language Based Higher Education", The Research Bulletin of Jordan ACM, 2(3), 2011, 27–34.

[12] P. Mukta, A. Usha, "A Study of Academic Performance of Business School Graduates using Neural Networks and Statistical Techniques", Expert Systems with Applications, vol. 36, 2009, 7865–7872.

[13] E.A. Amrieh, T. Hamtini and I. Aljarah, "Mining Educational Data to Predict Student's Academic Performance using Ensemble Methods", International Journal of Database Theory and Application, 9(8), 2016, 119–136.

[14] Prabu P. , Bendangnuksung, "Students' Performance Prediction using Deep Neural Networks", International Journal of Applied Engineering Research, vol. 13, 2018, 1171–1176.

[15] Wim De Mulder, Steven Bethard, and Marie-Francine Moens, "A survey on the application of recurrent neural networks to statistical language modeling", Computer Speech & Language, 30(1), 2015, 61–98.