A Performance Predictor for Honors Students Based On Elman Neural Network

Moke Xu
Shenyuan Honors College, Beihang University
China
08753@buaa.edu.cn

Xuyang Chen
Shenyuan Honors College, Beihang University
China
freeverc@gmail.com

Wenjun Wu
NLSDE Department of Computer Science and Engineering, Beihang University
China
wwj@nlsde.buaa.edu.cn

Abstract.
With the popularization of higher education, honors education has become an important work of research-oriented universities to cultivate excellent students. In order to evaluate the achievements of honors education and to make a guidance for honor educators, it is necessary to predict the performance of honors students effectively. This paper proposes a data-driven model to make predictions on students' performances based on an adjusted Elman Neural Network (Elman NN). Moreover, to be more significant, we made a comparison between Elman NN and some other methods. The result shows that our model performs much better. The performance predictor may provide a reference for honor educators in the professional choices and enable them to provide appropriate suggestions or motivations for those of the honors students who are at an early stage of learning risk or have a potential of an out-standing talent.

Keywords. Elman Neural Network, Data Mining, Predictive Model， Regression， Classification

1 INTRODUCTION

In recent years, honors programs of higher education have become available among the world famous universities and are widely recognized in American. Among the top 100 universities in the world, which have abundant education resources and small scales, more than 40% of the universities have honors programs providing honors students with challenging courses and high-level scientific research training opportunities. Following the success of their honors programs, top-notch universities in China also adopt honors programs to cultivate elite students.

Our experience of running the honors program in Beihang University for more than a decade reveals that there are two challenges for the student advisors. First of all, it's necessary to find out how to help honors students to choose majors which suit them best. Usually, the students have to choose their majors depending on their willingness and ability after the first school year. It is ideal to give students essential guidance in developing their interest and talent in the most suitable majors for them in the first year. But given the limited manpower of our honors program administers, we need a predictive tool to efficiently assess student ability and predict their future performance. Secondly, every year we have to manually identify the honors students as risk and give them counselling to overcome their academic difficulties and even adjust their negative timing habits in daily life. A powerful predictive model is also very important to help us in fulfilling this responsibility through necessary learning suggestions and teaching interventions.

In this paper, we adopt an Elman Neural Network as a modeling framework to implement the predictive model, which has been widely used in predictive problems in various fields. We redesign weighted context units in the hidden layer of Elman NN to reflect the latent interaction among honors students in the same year. Based on this improvement on the original Elman NN, we establish a predictive model to fit performance of honors students and verify the effectiveness of the model by the actual datasets. Experiments show that our model outperforms some other regular models.

The rest of the paper is organized as follows. Section 2 presents a summary of related work and a brief comparison to our model. Section 3 presents the dataset descriptions and processing details. Section 4 introduces our predictive model and related experimental results. Finally, Section 5 concludes the paper.

2 RELATED WORK

Prediction of student scores is an important research topic in the field of educational data mining. Many researchers have proposed predictive models based on a variety of machine learning techniques. Jie Xu, et al (2017) [1] developed a novel algorithm that enables progressive prediction of students' performance by adapting ensemble learning techniques and utilizing education-specific domain knowledge. It is proved that its prediction results are accurate enough compared to some other methods. Elbadrawy et al. (2016) [2], proposed a predictive model based on regression-based and matrix factorization–based methods to predict student performance. Dekker et al. (2009) [3], presented a case study to evaluate multiple drop-out prediction models.

All these previous efforts only focus on predicting future performance based on student current status and past academic performance without considering behavior features that are not directly related to their course study. They often rely upon Learning Management Systems on campus to collect study records as training datasets for developing their models. Such an approach has inherent limitation because it cannot capture students' daily activities that may have great impact on their study. Especially for the honors students at their first campus year, life style can bring negative influence on their study. To incorporate these factors into our predictive model, we decide to enrich the feature space of our model by introducing student daily activity features including consumption in campus cafeteria, Internet accessing at different time frames and library book-lending transactions. These data are collected from multiple e-campus service systems and assimilated into our training dataset. Given the temporal natural of honors student development and their daily activity data, we choose to adopt a simplified recurrent neural network, which was called Elman Neural Network [5], to build our predictive model.

3 DATA DESCRIPTION AND PROCESSING

3.1 Dataset Description

For honors students in Honors College, the design of honors project follows the principle of a solid foundation and gradual improvement. In the first year, students will learn basic subjects as a basis and preparation for further professional learning. In the following years, students will be major-oriented educated and learn more professional courses. As the knowledge basis of the first academic year is very important, we hope to predict the performance of the first school year in the first semester.

In this paper, we will establish a data-driven predictive model based on the data of students in grade 2015 and grade 2016, including their initial grades, learning and daily behaviors in the first semester, and we'll predict the performance of their core subjects and comprehensive scores. The input dataset contains 501 vectors, of which 205 are from students in Grade 2015 and 296 from students in Grade 2016. Every vector contains a 54-dimensional input vector and a 9-dimensional output vector.

After entering the University, many students will indulge in computer games resulting in reduced learning time. As students' internet access is a major factor affecting their academic performance, we collected students' internet accessing details including total length of Internet time, active periods, traffic, etc. And we organize Internet time, traffic data by month (X25-X54) and active periods by 6-hour periods (X21-X24). The college entrance examination scores represent the students' initial knowledge level and learning ability (X3-X7). The First midterm examination in college comes two months after enrolment, which indicates students' adaptability to university studies to some content. Moreover, we assume that students' monthly consumption, book-borrowing numbers and birth dates will also influence their final results. The initial CEE data, book-borrowing data, consumption and internet-accessing data can be collected easily through multiple e-campus service systems.

The 9-dimensional output data includes a 3-dimensional part of the comprehensive performance and a 6-dimensional part of performances in core courses. The consolidated performance part includes consolidated performance, the average grade of main courses and credit scores. For honors students, the consolidated performance is related to the latter two parameters by the Eq (1) as follows:

$$Y_1 = 0.6 \times \frac{Y_2}{Y_{2max}} + 0.4 \times \frac{Y_3}{Y_{3max}} \tag{1}$$

Y_2max means the maximum value of average performance of main courses for all students in the same grade. *Y_3max* means the maximum value of credit scores for all students in the same grade. The equation indicates the importance of core courses for honors students. The core subject grades section contains 6 elements, corresponding to their performances of the six core courses. These subjects are set especially for honors students in honors project, thus they can measure students' mathematical ability, experimental ability, programming ability, language ability properly, which are representative enough in measuring students' ability distributions.

The details of the input data and output data are shown in table 1. Xi means input data and Yi means output data.

Table 1. List of input data (Xi) and output data (Yi)

Symbol	Meaning	Symbol	Meaning
X1	Birth year	X25-X34	Internet-accessing time
X2	Birth month	X35-X44	Internet-downloading traffic
X3	Total score of CEE	X45-X54	Internet-uploading traffic
X4	Chinese perf. in CEE	Y1	Consolidated perf.
X5	Math perf. in CEE	Y2	Average grade of main courses
X6	Science perf. in CEE	Y3	Credit scores
X7	English perf. in CEE	Y4	Mathematics perf.
X8	Math perf. in FMEC	Y5	Basic Physics perf.
X9	Programming perf. in FMEC	Y6	General Chemistry perf.
X10	Number of books borrowed	Y7	Basic Life Sciences
X11-X20	Monthly consumption	Y8	Advanced Programming perf.
X21-X24	Internet total traffic by 6-hour periods	Y9	College English perf.

CEE: College Entrance Examination (perf. Means performance)
FMEC: The First Midterm Examination in College

3.2 Normalization

To reduce the amount of calculation and speed up the model training process, it is significant to normalize the input data before training. Min-Max Normalization, also known as dispersion standardization, is a linear transformation of the raw data so that the resulting values map between [0 - 1]. It is an effective normalization method.

The normalization equation is as follows. x^* is the normalized value and x the initial value. Max and min means the maximum value and minimum value in all items.

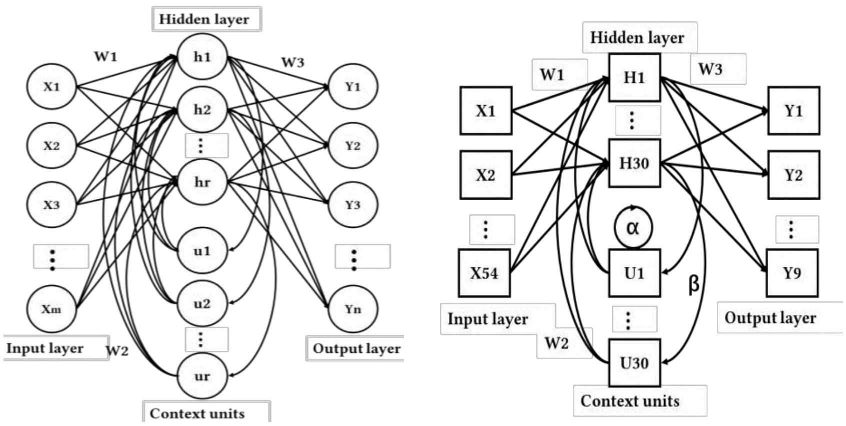
$$x^* = \frac{x-\min}{\max-\min} \tag{2}$$

4 MODEL AND RESULTS

4.1 Principles of Elman NN and Our Adjustment

An Elman neural network is a three-layer network with the addition of a set of "context units" used to remember the output value of the hidden layer units, which can be considered as a delay operator. The hidden layer is connected to these context units fixed with a weight of one initially. At each training step, the input will propagate over the feed-forward part during which a learning rule is applied. The back-connections part is fixed and save a copy of the values of the hidden units in the context units. The saved values will propagate over the connections before the learning rule is applied. That means the network can take into account the internal relationship among input data.

Fig. 1. The structure of basic Elman Neural Network and our model.



For honors students, the performance of every student might be influenced not only by his or her initial scores and daily behaviors, but also by the behaviors of other students. That's why we choose an Elman neural network which can take the inter-influence factors into consideration. The training sequences of the input data in all training epochs are set randomly, making it more reasonable to take mutual influence into account. Suppose that there are m nodes for the input layer, n nodes for the output layer, and r nodes for the hidden layer. Thus there will be r context units. In our model, m=54, n=9, r=30. The structure of the initial Elman NN and our adjusted model are shown in Fig. 1(right).

In traditional Elman neural network, the weight from context units is to the input layers are set as ones. But in fact, the recurrent part will not play an important role as the parameters from the input layer for the hidden layer. The weight shall be adjusted to correspond well to the application in predicting performance. Also, to make it more significant, we assume that the influence from previous two steps should not be neglected. The equations are shown as follows:

$$\mathrm{H(k)} = \mathrm{f}(\,W^1\mathrm{X(k)} + W^2\mathrm{U(k)}) \tag{3}$$

$$\mathrm{U(k)} = \alpha\,\mathrm{H(k-1)} + \beta\,\mathrm{U(k-1)}) \tag{4}$$

$$\mathrm{Y(k)} = W^3\mathrm{H(k)} \tag{5}$$

X(k) is the input value from the input layer. H(k)is the output value of the hidden layer. Y(k)is the output value of the output layer. f(x) is the activation function. It is always set as the sigmoid function. W_1 is the weight matrix between the input layer and the hidden layer W_2 is the weight matrix between the context units and the hidden layer. W_3 is the weight matrix between the hidden layer and the output layer. α is the feedback gain parameter for self-connection. β is the feedback gain parameter for the previous self-connection part.

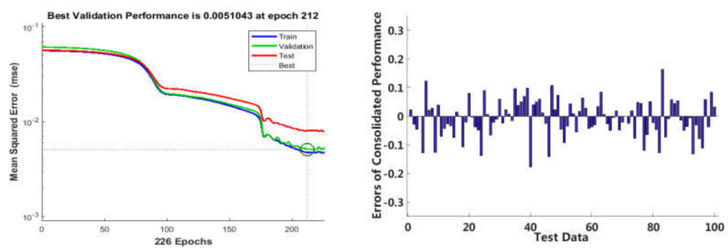
In our model, α and β should be set to a small value. Based on enough experiments, we found that the model performs well when we set α=β=0.05. When the values changes in a small range (0.02-0.2), the final results won't change a lot. That means the model is stable in such parameters.

4.2 Training and Tests

In training part, we set 401 train and validation data and 100 test data. As the predictive model is a regression problem actually, we set loss function (R(y,y*)) as mean square error (MSE) function, which always performs well in regression problems. The equation is as follows.

$$R(y, \hat{y}) = E_y(y - \hat{y})^2 \tag{6}$$

Fig. 2. The dynamic magnetization of training and the errors of consolidated performance



The process of model training and the results are recorded and shown in Fig. 3. The training function is chosen as gradient descent with momentum and adaptive learning rate backpropagation, which works well in Elman neural network according to a lot of experiments. We set learning rate as 0.05. There are 1000 training steps in each epoch. After 212 epochs, the model reached a stable point.

The output value are decimals ranging from 0 to 1 (representing scores ranging from 0 to 100). We calculated the errors of test data and present the errors of consolidated performance above. According the Figure 4, most errors of predicted results are no more than 0.1, which means our results are credible enough.

4.3 Comparisons

To be more significant, we made a comparison among Elman NN , BPNN (Back-propagation neural network), the most frequently used neural network model, and linear model. To ensure that the compared network is in the same size and scale, the BPNN is also arranged by three layers, including a 54-node input layer, a 30-node hidden layer, and a 9-node output layer. The training methods are all set similarly.

To evaluate the two methods properly, we calculated the confidence rate of both outputs based on the confidence interval of 10%.

$$\text{confidence rate (i)} = \frac{m_i}{n} \ , (i = 1,2, \dots ,9) \tag{7}$$

In this formula, n means number of items in test data, m_i means number of credible items in test data. A tested item is treated as a credible one if:

$$\frac{|Y^*{}_i - Y_i|}{Y_i} < 10\% \ , \ (i = 1,2, \dots ,9) \tag{8}$$

Y* means the predicted result and Y_i means the actual results, which is also the labeled value. The confidence rate in output data of both two methods is shown in Table 2. Obviously, the credible rate of our model is better. That means it is credible enough to predict student performance based on our model. Also, the prediction about average grade of core courses is the most accurate, which is also the most valuable parameter in measuring student learning ability.

Table 2. The comparison among our model, the BPNN model and linear model

Symbol	Meaning	CR of ENN	CR of BPNN	CR of LM
Y1	Consolidated performance	89%	86%	69%
Y2	Average grade of Core Courses	91%	88%	81%
Y3	Credit scores	82%	79%	73%
Y4	Mathematics performance	86%	84%	62%
Y5	Basic Physics performance	79%	72%	58%
Y6	General Chemistry performance	83%	85%	71%
Y7	Basic Life Sciences	88%	79%	65%
Y8	Advanced Programming perform	74%	66%	54%
Y9	College English performance	85%	69%	68%
Average Value		**84%**	**79%**	**67%**

The prediction confidence rate of advanced language programming performance is obviously lower than the others, for the uncertainty of the course. On the whole, most output items can be predicted accurately and we can trust the results at a low risk of making mistakes.

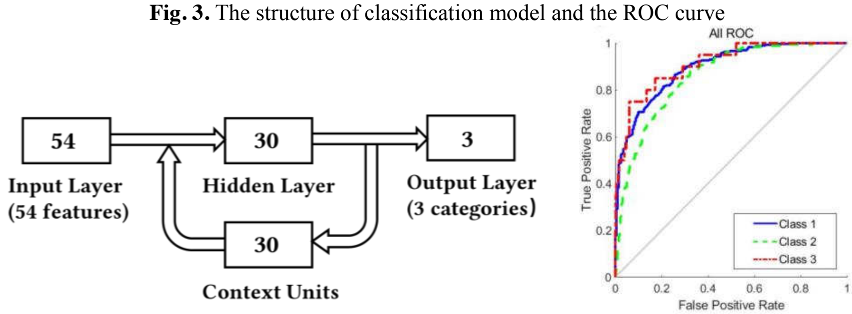
4.4 Classifications

In the classification model, we divide the honors students into three categories according to their consolidated performance, which respectively represent excellent, good and general level. The number of students in each category and the results are presented in Table 3. The structure of the model and the receiver operating characteristic curve (ROC curve) of test data are shown in Fig. 3.

Table 3. The number of students in each category and classification results

Symbol	Meaning	The whole number	The test data number	Correct Items	Correct Rate
C1	85-100	299	64	56	81.4%
C2	70-85	182	32	20	75.8%
C3	<70	20	4	3	75%

Fig. 3. The structure of classification model and the ROC curve



The correct rate of the student category prediction (shown in Table 4) is 77.4%. According to the ROC curve, the classification accuracy of the three categories is at the same level. It is obvious that we can also get good results through the classification model based on Elman NN.

4.5 Application of Models

As the prediction model can predict students' grades and categories at an early stage, risk students can be identified half a year in advance. In application, counselors will combine the results of classification and regression models. Firstly, they will collect data required and input the data into the model, which is an automated process wasting less time. After that, they can easily identify risk students based on the classification results. To know more details, they can consult the regression model to know the ability distribution details of those risk students according to the 9-dimension outputs. Finally, they shall offer some necessary suggestions.

In the early stage, counselors were unable to get sufficient information about students' learning status. Thus it is difficult to assess students' performance manually. But the models offer predictions based on students' data that are easy to get. The results with accuracy of 77.4% are valuable enough for honors educators to assess the learning level of every student. It is a convenient early-stage performance predicting tool on campus.

5 CONCLUSIONS

In summary, we have performed the process of the establishment of our predictive model and presented the results of prediction of students' final performance and ability distribution based on data of 501 honors students. By adjusting the values of feedback gain parameters for self-connection in Elman neural network and training the network reasonably, the predictive model works better compared to BPNN and linear regression method. According to our experiments, the consolidated performance and average grade of core courses in output value can be predicted most accurately. It is convenient for honors program counselors to predict students' performance and provide appropriate suggestions or motivations for different student categories in performance.

REFERENCES

[1] Xu, J., Han, Y., Marcu, D., van der Schaar M.: Progressive Prediction of Student Performance in College Programs. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 (2017).

[2] Elbadrawy A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H.: Predicting Student Performance Using Personalized Analytics. Computer, 49(4), 61-69, (2016).

[3] G. W. Dekker M. Pechenizkiy J. M. Vleeshouwers: Predicting students drop out: A case study. Proc. Int. Conf. Educ. Data Mining pp. 41-50 (2009).

[4] AI-Radaideh, Q, AI-Shawakfa, E., AI-Najjar, M: Mining student data using decision trees. International Arab Conference on Information Technology, Yarmouk University, Jordan, (2006).

[5] Elman, J. L. Finding Structure in Time. Cognitive Science, 14: 179–211. doi:10.1207/s15516709cog1402_1.(1990).

[6] Liu H., Tian H., Liang X., Li Y.: Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. https://doi.org/10.1016/j.apenergy.2015.08.014 (2015).

[7] Xing W., Guo R., Petakovic, E., Goggins S. Participation-based student final performance prediction model through interpretable Genetic Programming, Integrating learning analytics, educational data mining and theory, (2014).

[8] Ahmed A. B. E. D., Elaraby I. S.: Data Mining: A prediction for Student's Performance Using Classification Method. World Journal of Computer Application and Technology, vol. 2, no. 2, 43 – 47 (2014)

[9] Kukk V.: Course Implementation: Value-Added Mode. In: Uden L., Liberona D., Feldmann B. (eds) Learning Technology for Education in Cloud – The Changing Face of Education. LTEC 2016. Communications in Computer and Information Science, vol 620. Springer, Cham (2016)

[10] F. Okubo, T. Yamashita, A. Shimada, H. Ogata. A neural network approach for students' performance prediction Proceeding LAK '17 Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 598-599, (2017).

[11] R. Alkhasawneh and R. H. Hargraves,: Developing a hybrid model to predict student first year retention in STEM disciplines using machine learning techniques, Journal of STEM Education: Innovations and Research, vol. 15, no. 3, 35-42, (2014).