

# *Research on Students' Score Prediction Based on BP Neural Network*

Yan Cheng<sup>1, a</sup>, Quanyuan Weng<sup>1, b</sup>

<sup>1</sup>College of computer information engineering, Jiangxi Normal University, Nanchang, Jiangxi, 330022, China

<sup>a</sup>chyan88888@jxnu.edu.cn, <sup>b</sup>1548244782@qq.com

**Keywords:** BP neural network; educational data mining; score prediction

**Abstract:** With the rapid development of network technology and computer technology, digital campus construction is more and more popular, education workers use information technology to solve the common problems efficiently in the daily teaching, such as the average score, ranking, excellent rate, pass rate. These simple statistical analysis can no longer meet the needs of today's teaching platform, not reflect the relationship between daily performance and final score and not predict students' score, resulting in a large number of data wasted in education management system. Traditional education data mining method is to use decision tree to analyze student's score, so as to find out the factors that influence student's score and the relationship among factors, but the result of prediction is often unsatisfactory. Based on the past achievements and daily behavior data of students, BP neural network algorithm is used to predict the final grades of students. The experiment shows that the accuracy is very high.

## 1. Introduction

In recent years, it has entered the era of information explosion with the rapid development of information technology, a large amount of data has not been fully utilized, and the hidden knowledge in large amounts of data has not been tapped and utilized, resulting in the waste of data resources. In response to solve these problems, data mining technology has gradually gained attention [1]. Data mining generally refers to finding meaningful patterns and knowledge from a large amount of irregular, noisy data [2]. It has been successfully applied in various fields such as finance, e-commerce, banking, insurance, education, and business. Stock prediction [3], anomaly detection [4], and customer segmentation [5] are all successful applications of data mining.

In the field of education, teaching platforms and online teaching platforms have accumulated a large amount of data, which has prompted the generation of educational data technologies. Education data mining refers to the use of computer science, pedagogy, statistics, and other discipline theories and cross-cutting technologies to study and solve education problem [6].

In general, score prediction is based on the student's past performance. the student's previous score is used to predict the score to be achieved, which will prompt the teacher and the student to use high efficiency methods to improve score [7,8]. Traditional score predictions often overlook students' behavioral information. The campus card records student behavior information, including

the number of students entering the library, the frequency of borrowing books, the rules of entering and leaving the classroom, and the law of eating breakfast. These are related to students' efforts. These behavioral data portray students' behavioral portraits, which also quantify two indicators that psychologically affect student performance: the student's life patterns and the student's effort level [9]. Based on the student's past performance and behavioral data, this article uses BP neural network technology to predict students' final score. Experiments show that the predicted results are in line with expectations.

## 2. BP Algorithm Principle

- (1) Calculate forward propagation to get the activation values of  $(L_2, L_3, \dots, L_{n_l})$ .
- (2) Calculating the residuals  $\delta_i^{n_l}$  of each node in the output layer  $((n_l))$

$$\delta_i^{n_l} = \frac{\partial}{\partial z_i^{(n_l)}} \frac{1}{2} \|y - h_{w,b}(x)\|^2 = -(y_i - a_i^{n_l}) \cdot f'(z_i^{n_l}) \quad (1)$$

Among them,  $z_i^{n_l}$  is the input value of the  $n_l$ th layer node,  $a_i^{n_l}$  is the output value of the  $n_l$ th layer node,  $f(\cdot)$  is the activation function.

- (3)  $l = n_l - 1, n_l - 2, \dots$ , Calculate the  $i$ -th node residual of each layer

$$\delta_i^{(l)} = \left( \sum_{j=1}^{S_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}) \quad (2)$$

- (4) Calculate the required partial derivatives

$$\frac{\partial J(W, b; x, y)}{\partial W_{ij}^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} \cdot \frac{\partial z_i^{(l+1)}}{\partial W_{ij}^{(l)}} = a_j^{(l+1)} \cdot \delta_i^{(l+1)} \quad (3)$$

$$\frac{\partial J(W, b; x, y)}{\partial b_i^{(l)}} = \frac{\partial J(W, b; x, y)}{\partial z_i^{(l+1)}} \cdot \frac{\partial z_i^{(l+1)}}{\partial b_i^{(l)}} = \delta_i^{(l+1)} \quad (4)$$

The derivative of the cost function to the parameter is equal to the excitation value of the  $j$ -th node of the  $l$ -th layer times the residual of the  $i$ -th node of the  $(l+1)$ -th layer.

## 3. Experiment

This experiment uses the neural network function provided by the neural network toolbox in Matlab [10]. The program runs on the Shenzhen Computer and the main configuration is: 8G memory, Inter(R) Core(TM) i7-7700HQ CPU 2.80GHz.

### 3.1 Data sources

The data in this paper is selected from the student score data and the data in campus cards. This article aims at this kind of situation, carries on the improvement to the predicting sample, not only based on the student's previous score, but also includes the student behavior information. The student's final score is related to a variety of factors, not just the previous grades. The student's score is also reflected in the student's daily behavior, such as whether the student eats breakfast or goes to the library to borrow books.

During the university, the major homework of students is in the freshman and sophomore year, so it can be said that the freshman and sophomore is the key period for the college students to learn the basic knowledge and professional knowledge. The basic knowledge and professional knowledge directly affect the other scores when they are not solid enough, especially final score may even affect the graduation and employment of students. Therefore, it is very necessary to study the influence of the results of freshman and sophomores on future graduation. There are many courses for freshmen and sophomores, the major public courses are taken as factors that affect the graduation performance of students. There are 10 major public courses, namely Advanced Mathematics, Linear Algebra, Probability Statistics, Discrete Mathematics, Data Structure, Computer Network, Principles of Computer, English, and C++, Software Engineering, Library frequency and the Frequency of breakfast to predict the students' final score. Table 1 represents students' score and behavior information.

Table 1 Student score

| subject | Advanced Mathematics | Linear Algebra | Probability Statistics | Data Structure | ... | Breakfast Frequency | Final score |
|---------|----------------------|----------------|------------------------|----------------|-----|---------------------|-------------|
| 1       | 85                   | 69             | 70                     | 79             | ... | 5.6                 | 85          |
| 2       | 74                   | 63             | 68                     | 78             | ... | 5.8                 | 87          |
| 3       | 87                   | 74             | 75                     | 83             | ... | 4.4                 | 70          |
| 4       | 62                   | 79             | 65                     | 73             | ... | 3.4                 | 65          |
| 5       | 82                   | 79             | 85                     | 82             | ... | 6.1                 | 85          |
| 6       | 69                   | 72             | 71                     | 83             | ... | 4.5                 | 78          |
| 7       | 76                   | 67             | 75                     | 95             | ... | 4.5                 | 90          |
| 8       | 80                   | 86             | 80                     | 96             | ... | 4.7                 | 92          |
| ...     | ...                  | ...            | ...                    | ...            | ... | ...                 | ...         |

### 3.2 Data processing

Data processing refers to the process of deriving valuable and meaningful data from the disorganized and incomprehensible data.

The data is mapped to the [0,1] intervals by normalized. In this experiment, the min-max normalization method is used. The formula is as follows:

$$x^* = \frac{x - \min}{\max - \min} \quad (5)$$

min represents the minimum value, max represents the maximum value, the reasons for data normalization are mainly due to the following reasons:

(1) The input data units are not the same, some data ranges are small, and some data ranges are very large, which can easily cause the BP neural network to slow convergence and long training time.

(2) The range of the activation function of the output layer of the BP neural network is limited and within a certain range.

Table 2 Normalize student data

| Advanced Mathematics | Linear Algebra | Probability Statistics | Discrete Mathematics | Data Structure | Computer Network | ... | Frequency of breakfast |
|----------------------|----------------|------------------------|----------------------|----------------|------------------|-----|------------------------|
| 0.111111             | 0.25           | 0                      | 0                    | 0              | 0.51612903       | ... | 0.586206897            |
| 0.907407             | 0.59375        | 0.875                  | 0.875                | 0.607143       | 0.64516129       | ... | 0.75862069             |
| 0.518519             | 0.28125        | 0.3125                 | 0.982143             | 0.535714       | 0.61290323       | ... | 0.448275862            |
| 0.185185             | 0.4375         | 0.125                  | 0.75                 | 0.392857       | 0.74193548       | ... | 0.137931034            |
| 0.648148             | 0.125          | 0.4375                 | 0.875                | 0.964286       | 0.90322581       | ... | 0.448275862            |
| 0.611111             | 0.15625        | 0.4688                 | 0.928571             | 0.25           | 0.74193548       | ... | 0                      |
| 0.814815             | 0.1875         | 0.2813                 | 0.517857             | 0.392857       | 0.58064516       | ... | 0.827586207            |
| 0.814815             | 1              | 0.8438                 | 0.428571             | 0.535714       | 0.25806452       | ... | 0.655172414            |
| ...                  | ...            | ...                    | ...                  | ...            | ...              | ... | ...                    |

### 3.3 Prediction model based on BP neural network

This article selects 400 student score and campus card behavior information. Using the continuous parameter adjustment test, the hidden layer is set to 5, this experiment randomly generated 320 random records as a training set and the remaining 80 records as validation sets. The traditional BP algorithm uses a fixed learning rate 0.01. The result of the forecast is as follows:

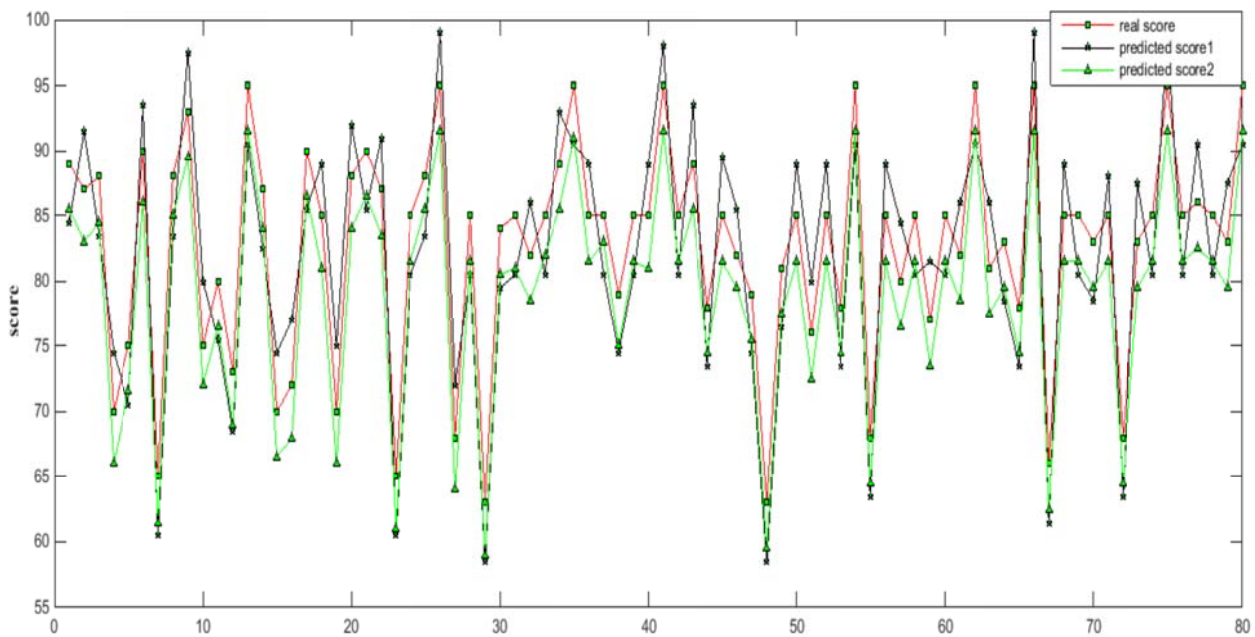


Figure 1 Score prediction

Figure 1 represents the contrasts between the real values and predicted values. The red point represents the real score of the students. The " predicted value1" (black point) represents the predicted value of the score that uses the student's score only as the dependent variable. The " predicted value2" (green point) represents the predicted value of the score that uses the student's score and behavioral information as the dependent variable.

It is shown that the prediction value of student score and behavior information as the dependent variable is closer to the real score. It shows that it is reasonable to take the student's behavior information into consideration of the factors that affect the student's final score, which is in line with the previous expectation.

The relative error can reflect the deviation between the predicted value and the true value. It is a commonly used measure index. The specific formula is as follows:

$$error = |y_i - y|/y \times 100\% \quad (6)$$

Where 'y' represents the true value and 'y<sub>i</sub>' represents the predicted value. This experiment takes the student's graduation grade as the real value and the BP neural network predicted grade as the predicted value.

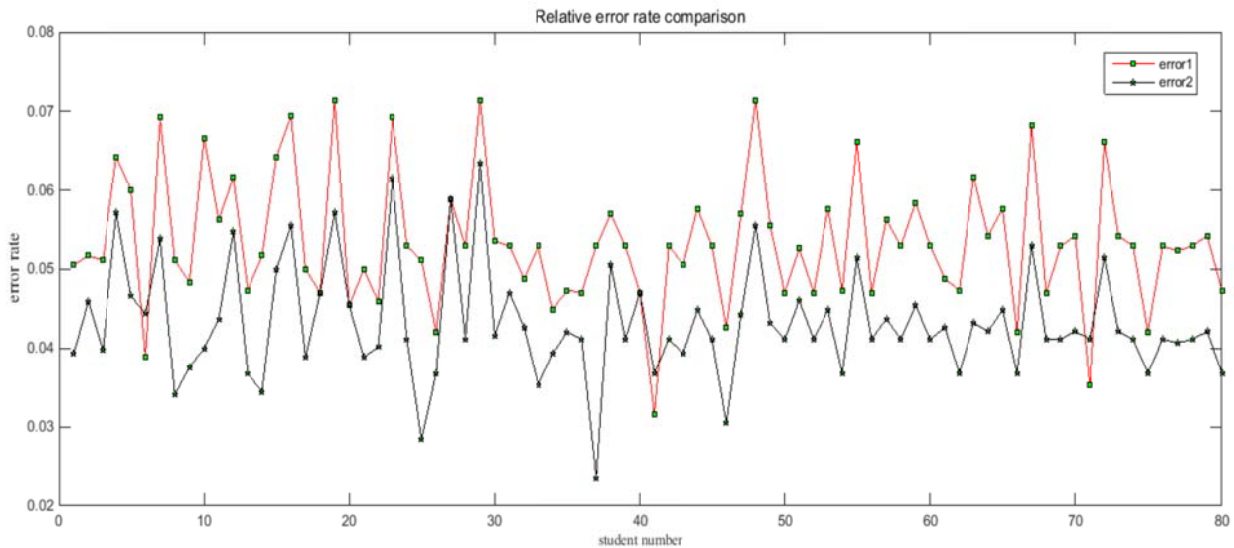


Figure 2 Error rate

Figure 2 represents the comparison of the relative error of the experiment. The red line (error1) represents the relative error that only takes the students' scores into account. The black line (error2) represents the relative error that takes the student's score and behavior information into account. It can be seen from the figure that the black line is mostly below the red line, it shows that the relative error of the forecast relative error which takes the students' score and behavior information into consideration is lower than the relative error that only considers the students' score in most conditions. The experimental results show that the students' behavior has a certain influence on the final score, and the school should not only pay attention to the previous scores of the students, but also pay attention to the students' behavior. It is of great significance to encourage students to go to the library, get up early and keep good habits of study and living.

#### 4. Conclusion

This experiment first introduced the data sources in the data preparation process, and then introduced the process of building a multidimensional data set and the data preprocessing operations. Then, the process of predicting final score is introduced based on the BP neural network prediction model. The prediction performance of the BP neural network model is intuitively reflected in the form of the graph, and the experimental results are evaluated. The results of the experiment show that the prediction results have a certain degree of credibility. It is proved that BP neural network model can be used as a good predictive model, However, it still needs further research and improvement, the next step is to collect more data, use data warehouse technology, carry out mass data analysis, use different data mining methods for mining analysis, find the best mining method to make the result more complete.

## Acknowledgments

This research was supported by National Natural Science Foundation of China (NSFC Grant No.61262080), Jiangxi Province Science and Technology Support Major Project (20151BBE50121), Science and Technology Key Project of Jiangxi Province Education Department (GJJ150299). the Jiangxi provincial science and technology key project (20161BBE50086), All support is gratefully acknowledged, the key projects of Jiangxi Provincial Department of education 20161BBE50086, Development and implementation of quality evaluation software for primary and secondary school teachers in Jiangxi: JXJSZLB15.

## References

- [1] Zhang Haiwei. *Application Research of Data Mining in Student Performance Analysis* [D]. Suzhou University, 2015.
- [2] Witten IH, Frank E. *Data mining: practical machine learning tools and techniques with Java implementations*[M]. ACM, 2002.
- [3] Yu Fufu. *Stock prediction based on neural network* [D]. Harbin Institute of Technology, 2016.
- [4] Miao Yongchun, Cheng Yan. *Outlier detection method and its improvement method in the era of big data*[J]. *Journal of Jiangxi Normal University (Natural Science Edition)*, 2014, 38(05): 454-458+ 495.
- [5] Yang Jian. *Web customer group hierarchical clustering algorithm* [J]. *information technology*, 2015 (15).
- [6] ZHOU Qing, ZHAI Chao, YANG Dan. *A Review of Research Progress in Educational Data Mining*[J]. *Journal of Software*, 2015, 26(11): 3026-3042.
- [7] Chen Qishan, Li Li. *Understanding Assessment and Performance Prediction: Two Different Forms of Meta-Understanding Monitoring* [J]. *Chinese Journal of Psychology*, 2008, 40(9):961-968.
- [8] Cheng Yuxia. *Application research of data mining in prediction of academic performance* [D]. Jiangnan University, 2016.
- [9] Lü Hongwei, Lian Defu, Nie Min, Zhou Tao. *Big Data Leads the Future of Education: Talking about Performance Prediction* [J]. *Big Data*, 2015, 1(4): 118-121. [10] Wen Xin. *MATLAB neural network simulation and application* [M]. Science Press, 2003.