

Prediction of Final Grade of Advanced Mathematics based on BP Neural Network

Daohai Zhang¹, Shang Zhang²

¹ School of Information Network Security, Xinjiang University of Political Science and Law, Tumushuke Xinjiang, 843900, China

² Office of Academic Affairs, Xinjiang University of Political Science and Law, Tumushuke Xinjiang, 843900, China

Abstract: In recent years, with the continuous expansion of college enrollment, the number of students is increasing year by year, so the scale of classroom teaching is also getting larger and larger, so that it is difficult for teachers to track and understand the learning situation of each student, which affects the quality of teaching to a certain extent. At the same time, every year, a certain number of students in colleges and universities fail the exam, repeat the grade, or even drop out. If these problems cannot be solved in time, they will seriously affect the psychological health and future development of students, and also affect the employment rate of graduates, and thus affect the development of schools. How to predict the situation of students, so that students in the study of trouble, hesitation, timely interference, so that students successfully complete the university study life. Through the analysis of the failed students, it is found that about 80% of the failed students are advanced mathematics. Therefore, this paper will use the machine learning technology based on BP neural network to analyze the scores of senior students in the final exam of advanced mathematics. According to the different scores, the scores will be divided into four categories: excellent, good, medium and failed. Through the establishment of the grade prediction model, the paper predicts the grade of the college students in the final exam of higher mathematics, and then gives the students learning guidance.

Keywords: College Students; BP Neural Network; Advanced Mathematics; Grade Prediction.

1. Background and Significance of the Research

Under the background of the information age, all walks of life have accumulated a huge amount of data, and the amount of data is still growing at an exponential rate. Massive data often implies some valuable knowledge and valuable information, which is a very valuable "soft resource" in various fields and industries. Machine learning technology based on BP neural network can find some laws inherent in data, extract valuable information and knowledge[1], so as to serve to solve problems in various fields, and help managers to make more "scientific" decisions. Based on BP neural network machine learning technology, intelligent analysis and processing of data, so as to use the hidden rules in the data to predict the development trend of things, has become the consensus of today's industry and academia. At present, machine learning technology based on BP neural network has been widely used in many fields such as finance, medical treatment, e-commerce, energy and manufacturing, transportation and transportation.

Big data has an important impact on all walks of life, and education is no exception. Based on the era of big data, education big data has emerged as an emerging discipline. Introducing the concept of big data into education can make the process of education and being educated more clear, accurate and rapid, and even "predict" success or failure in advance. The "big" of educational big data is not only the large quantity, but also emphasizes the great value. Educational big data is a potentially huge wealth that is difficult to estimate, which greatly promotes the research on educational issues [2]. The development of machine learning technology based on BP neural network is realizing the infinite appreciation of this wealth and promoting the development of students in the right direction.

Student achievement prediction, also known as student

academic performance prediction, is one of the important research issues in education data[3]. The purpose is to predict students' future academic performance by using student-related information[4]. This paper obtains the behavior data of previous students through questionnaires, derives the grade information and average grade point of previous students through the educational administration system, and uses the machine learning technology based on BP neural network to establish a grade prediction model to predict the final exam scores of lower grade students. Make academic prediction for students, and then remind students to reflect on time, check the gaps and make progress.

2. Research Status

In recent years, the study of learning prediction based on data mining in foreign countries focuses on three aspects: the selection of prediction indicators, the detection of prediction start time, and the evaluation of prediction model effect; At the level of research data, the existing research data are mostly obtained from the learning management system, and the daily behavior data of students has not been fully utilized. In terms of empirical research, there are few empirical studies and insufficient attention to dynamic data in the selection of learning prediction indicators[5].

Compared with foreign countries, the research on learning prediction in China mainly stays at the theoretical level, mostly focusing on the study of learning prediction mechanism and prediction model design, and the research on application level is lacking. In terms of the prediction mechanism, Chen Qinhua studied the academic prediction mechanism based on the teaching management of the credit system, divided the prediction level and initially designed the working system of the academic prediction mechanism for college students[6]. Wang Zihua et al. designed a prediction system for academic difficulties in view of the increasing number of students with academic difficulties. Through

comprehensive analysis of students' academic conditions, students were evaluated in combination with professional training programs and teaching requirements, and warnings were given to different degrees according to the evaluation results[7]. Shang Weiwei et al. studied the academic prediction mechanism of local college students in view of the poor student source quality in local colleges and universities[8]. He Kui improved the student academic prediction mechanism by establishing effective linkage between the school educational administration department, the student administration department and the parents of students with learning difficulties[9]. In terms of prediction model design, Wu Haifeng et al. built a prediction model based on cluster analysis and association analysis, so as to timely discover and predict students of three types: low, landslide and potential type [10]. From the perspective of counselors, Wan Xinghuo et al. built an innovative management model of college students' academic forecasting based on the kernel principal component analysis method, which combined dynamic qualitative forecasting and quantitative forecasting[11]. Zhang Fusheng et al. proposed a dynamic academic forecasting system reference model with two-way forecasting threshold and implemented a specific help plan for individual forecasting characteristics[12]. Through in-depth mining and research on outlier data, Jin Yifu et al. proposed a prediction information discovery and generation model called LAOMA, which integrates data from three aspects: curriculum, classroom and extracurricular. Based on this model, two kinds of six-level signal systems and feedback mechanisms for academic prediction were established[13]. Based on the improved Apriori algorithm, Du Juan et al. studied the prediction system of college students' grades and found the correlation between different disciplines through the mining and analysis of students' grades[14]. Cui Qiang et al. used BP neural network model to predict students' academic achievement[15]. The study of student learning prediction is mainly divided into curriculum prediction and academic prediction. Curriculum prediction is aimed at a

specific course, and the data involved are mainly the learning process data of the course. Academic prediction can be divided into two categories: The academic prediction studied in the field of education directly divides students into grades at the end of each semester and issues predictions to students at the beginning of the next semester; Academic prediction using statistical theory is to predict whether students can successfully graduate, and to mine the correlation relationship between courses by using correlation analysis, and to predict students' academic completion through early course results.

3. Research Content and Research Methods

3.1. Research Content

This paper proposes a grade prediction model based on BP neural network, which can predict the grade of students at the end of the semester according to their daily learning behavior characteristics. By using online questionnaire to investigate students' behavior information data and students' final grades derived from the educational administration system, the students' final exam course score information and students' daily behavior characteristic data are combined to form a data set of students' daily behavior course score, and the machine learning technology based on BP neural network is used to analyze the data set of students' daily behavior course score. Determine the relationship between students' daily behavior and final exam scores.

3.2. Research Methods

3.2.1. Obtaining the Data Set

Through the combination of the questionnaire survey and the teaching administration system to export the final exam results of students, Table 1 is part of the data set, and the data statistics are based on weekly units. Grades with A, B, C, D to indicate excellent, good, medium, poor four grades.

Table 1. Partial data of the dataset

Student	self-study	Attend lecture	Online learning	After-class tutoring	Complete the exercise	Admission score	Other learning	Grade level
S_1	6h	4h	3h	2h	Y	A	2	A
S_2	3h	3h	1h	0h	N	D	1	C
S_3	3h	3h	0h	0h	Y	B	0	D
...
S_{500}	4h	3h	3h	1h	N	C	0	B

3.2.2. Data Preprocessing

In the process of data learning by BP neural network prediction model, the data of students surveyed by questionnaire is not completely accurate, and students may be subjective when filling in the data. Therefore, the original data is generally incomplete and inconsistent dirty data, which cannot be directly used as the training data set of neural networks. It may cause the training effect of neural network is not very ideal. In order to improve the training effect of the neural network, this paper preprocesses the original data samples before training the neural network. Data preprocessing is very important when using neural network to

predict performance. The following is a detailed introduction to the preprocessing of original data.

(1) Reasons for data normalization

Since each sample has multiple eigenvalues, that is, each grade level has multiple eigenvectors. The number of feature vectors in each grade means that the BP neural network prediction model has the same number of input nodes. Since these feature vectors are quantities with dimensions, and the magnitude of each feature vector is different, which is not conducive to the comparison between different samples of the same feature vector. In order to make the original data sample easy to be processed by the prediction system, it is usually

necessary to eliminate the dimensional interference and analyze the data sample only from the quantitative level of its features, which has the advantage of avoiding the phenomenon of neuron output saturation caused by too large absolute value of the input data sample. After the normalization of the original data sample, each data index reaches the same order of magnitude. Comparative evaluation can be performed.

(2) Data normalization method

The data normalization method used in this paper is Min-max normalization method. Min-max normalization, also known as deviation normalization, is based on a linear transformation of raw data with a very large range of values, so that the resulting values are mapped between [0,1]. For the conversion function, see formula (1).

$$y = \frac{x - MinValue}{MaxValue - MinValue} \tag{1}$$

Formula:

MaxValue-- represents the maximum value of the sample data;

MinValue-- represents the minimum value of the sample data;

x-- represents the original data of the sample;

y-- Indicates normalized data.

Min-max one disadvantage of the normalization method is that whenever new data samples are added, the sum may change, and then the sum must be redefined. However, this shortcoming has no impact on this paper, because the data sets used for BP neural network training in this paper have been determined in advance, and no new data has been added, so it is completely feasible in theory to adopt normalization method to normalize the collected data sets.

(3) Data normalization results

The original dataset used for grade prediction in this paper has 4 grades, each with 7 feature vectors. In the original data set, the data are all dimensional, and the value difference between different feature vectors of the same score level and the same feature vectors of different score level is very large. In order to see the value range of the original data set more directly, the value distribution of the data after normalization is shown in Figure 1.

Before the original data was normalized, the data size distribution was very wide, ranging from 0 to 100, that is, the data was distributed in the interval [0,100]. Such data value range seriously affected the training effect of BP neural network. The score prediction system in this paper is based on BP neural network, if the training effect of neural network is not good, as a result, the accuracy of the prediction results of the performance prediction system is not very high, so the original data set must be normalized before it can be used to train the neural network.

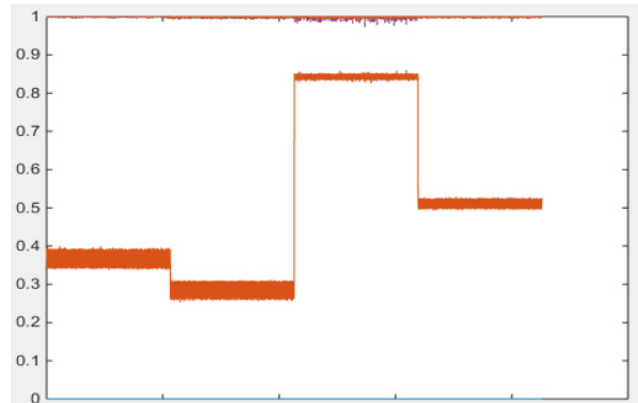


Fig 1. Data value distribution after normalization

3.2.3. BP Neural Network Structure

The classic BP network is a three-layer network structure, namely the input layer, the hidden layer and the output layer. The input layer is responsible for data input and does not process the data, while the hidden layer may be one or more layers, which is responsible for mapping and transforming the data, and the output is responsible for classifying and output. In terms of network structure, BP network is no connection within the layer, and the layer is fully connected. Although multiple hidden layers can reduce the error between the output result and the expected result, the structure of the network is very complex, which increases the running time of the network and is prone to overfitting problems. However, due to the large reduction of parameters, the single hidden layer may lead to low training accuracy, weak network generalization ability and underfitting problems, so specific problems need to be selected for specific network structures.

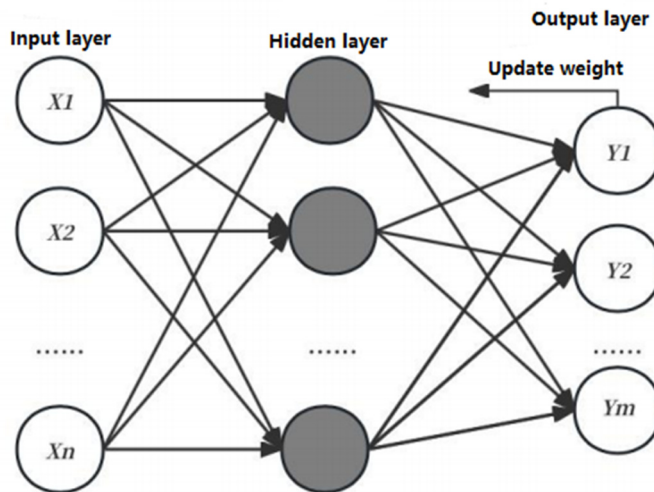


Fig 2. Schematic diagram of three-layer topology of BP neural network

When solving real problems, the number of hidden layer nodes of BP neural network is generally set according to

experience. A three-layer neural network can approximate the function of any function with any accuracy, just like a complicated function model, then any sub-function will be included. Therefore, the three-layer BP neural network can effectively solve most problems in practical applications. Based on the above content, the fault diagnosis model in this paper is also based on three-layer BP neural network, which is composed of an input layer, a hidden layer and an output layer.

The BP neural network with three-layer topology is shown in Figure 2, where x_1, x_2 and x_n are input vectors and Y_1, Y_2 and Y_m are output vectors.

The basis function between neurons at each layer of the network adopts linear basis function, whose expression is shown in equation (2), the activation function $f(\cdot)$ adopts Sigmoid function, whose expression is shown in equation (3).

$$u = \sum_{i=1}^m (\omega_{ji}x_i) - \theta_j \quad (2)$$

$$y_i = f(u) = \frac{1}{1+e^{-u}} \quad (3)$$

Formula:

y_i -- is the output of the I-th neuron in the hidden layer or output layer;

i -- is the number of nodes in the hidden layer or output layer;

m -- is the number of nodes in the input layer or hidden layer;

ω_{ij} -- is the connection weight between the input layer and the hidden layer or the hidden layer and the output layer;

θ_i -- is the threshold of the hidden layer or the output layer;

f -- Activate a function for the hidden or output layer.

Set the output of neurons at the JTH node of layer L as p_j^L , and the threshold of neurons at the JTH node of layer L as θ_j^L .

The expression of p_j^L is shown in equation (4).

$$p_j^L = G(\sum_k \omega_{jk}^L p_k^{L-1} + \theta_j^L) \quad (4)$$

Formula:

k, j -- represents the k or JTH node;

$G(\cdot)$ -- denotes the Sigmoid function;

L -- represents the L-layer neural network, and L is equal to 2 or 3;

p_j^L -- indicates that the output of neurons at the JTH node of layer L is;

p_k^{L-1} -- represents the output of neurons at the k node of layer L-1;

ω_{jk}^L -- represents the connection weight of the KTH input node of layer L-1 to the JTH output node of layer L;

θ_j^L -- Represents the negative value of the threshold corresponding to the JTH output node of layer L.

4. BP Neural Network Learning Algorithm

4.1. Training Process of BP Algorithm

(1) Propagation from the input layer through the hidden layer to the output layer, this process is called forward propagation;

(2) The error between the real output and the expected output of the neural network is regarded as an error signal. If the error exceeds the set threshold, the connection weight and threshold of each layer are adjusted by gradient descent

algorithm from the output layer to the input layer. This process is called error back propagation;

(3) Forward propagation and error backpropagation are repeated for many times, during which the weights and thresholds of the neural network are constantly adjusted. This process is called memory training process;

(4) With the continuous progress of the above process, the global error of the network tends to the minimum value or reaches the termination condition set in advance, which is called the learning convergence process.

The main objective of the BP algorithm is to compute the partial derivatives $\partial E/\partial \omega$ and $\partial E/\partial \theta$ of the cost function E with respect to all network connection weights ω and thresholds θ . The expression of the cost function E is shown in equation (5).

$$E(\omega, \theta) = \frac{1}{2n} \sum_x ||y(x) - p^N(x)||^2 \quad (5)$$

Formula:

n -- is the total number of training samples x;

ω, θ -- is the connection weight and threshold value;

$y(x)$ -- the expected output of training sample x;

N -- is the total number of layers of the network. In this paper, a 3-layer neural network is used, that is, $N=3$;

$p^N(x)$ -- The real output of the training sample in the output layer.

4.2. Performance Prediction Experiment based on BP Neural Network

4.2.1. Parameter Selection

(1) Set the number of nodes in the input layer

Input layer nodes are input interfaces of data samples, and their number depends on the number of features of data samples. The dimensionality of the data set used for neural network training in this paper is reduced by principal component analysis. The data after dimensionality reduction has 5 features, so the number of nodes in the input layer is 5.

(2) Setting the number of nodes in the output layer

When BP neural network is used for gear performance recognition, the output performance type of the neural network is expressed as a combination of 0 or 1, and the number of nodes in the output layer can be determined according to the number of performance types. Since the number of achievement categories in this paper is 4, the output layer nodes of the neural network in this paper are 4.

(3) Setting the number of nodes in the hidden layer

The selection of the number of hidden layer nodes in BP neural network has a great influence on gear grade classification. With fewer hidden layer nodes, the ability of neural network to obtain information is poor, and the error rate of gear performance prediction is high. When the number of hidden layer nodes is large, the neural network is prone to overfitting, and the training time of the network will be increased. However, as far as the current theory is concerned, there is no best way to determine the hidden layer nodes of BP network. In this paper, empirical method and trial and error method are used to select the appropriate number of hidden layer nodes, and the number of hidden layer nodes in this paper is 12.

(4) Other parameter Settings

In BP neural network, the learning factor $\eta = 0.1$ and the critical value of training error MSE is set to 0.001.

4.2.2. Experimental Results and Analysis

The accuracy of the BP God network performance prediction model for the gearbox gear performance prediction is shown in Figure 3. Figure 3 is the confusion matrix. The horizontal axis of the confusion matrix represents the target category, the vertical axis represents the output category of the prediction model, the first four diagonal lines represent the accuracy rate of the prediction results of each grade level, and the bottom diagonal lines represent the total prediction accuracy rate, which is generally used as the measurement standard for the performance of the prediction model. It can be seen from the figure that the accuracy of prediction results of BP God network performance prediction model is 69.3%.

Output Class	1	2	3	4	
1	67 22.3%	1 0.3%	0 0.0%	0 0.0%	98.5% 1.5%
2	8 2.7%	73 24.3%	2 0.7%	10 3.3%	78.5% 21.5%
3	0 0.0%	0 0.0%	7 2.3%	2 0.7%	77.8% 22.2%
4	1 0.3%	6 2.0%	62 20.7%	61 20.3%	46.9% 53.1%
	88.2% 11.8%	91.3% 8.8%	9.9% 90.1%	83.6% 16.4%	69.3% 30.7%
	1	2	3	4	
	Target Class				

Fig 3. Prediction results of BP God network performance prediction model

5. Summary

This paper takes student achievement as the research object, mainly aiming at the historical achievement of senior students to train the prediction model, this paper puts forward the achievement prediction based on BP neural network. Considering that the original data set of students is too large and the range of values is too large, the original data set needs to be pre-processed. This paper proposes to use the Min-max normalization method to pre-process the data, so that the pre-processed data can be used as the input data of the prediction model. The grade prediction model in this paper is based on BP neural network.

Acknowledgments

Xinjiang University of Political Science and Law president's fund+XZZK2021001.

References

- [1] Ertel W. Machine Learning and Data Mining. Springer London, 2011.
- [2] Mostow J, Beck J . Some useful tactics to modify, map and mine data from intelligent tutors[J]. Natural Language Engineering, 2006, 12(Pt2):195-208.
- [3] Hippe J M. A Review of the State of the Art.
- [4] Zhou Qing, Mou Chao, Yang Dan. Review on research progress of educational data mining [J]. Journal of Software, 2015, 26(11):17. (in Chinese)
- [5] Xiao Wei, Ni Chuanbin, Li Rui. Foreign research on learning prediction based on Data mining: Review and prospect [J]. China Distance Education, 2018(2):9. (in Chinese)
- [6] Chen Qinhu. Research on the academic prediction Mechanism of college students under the credit system [J]. Journal of Guangxi Normal University: Philosophy and Social Sciences Edition, 2007(S2):6.(in Chinese)
- [7] Wang Zihua, Zhang Baojing, Shi Yuan, et al. Study on the prediction mechanism of College students' Academic difficulties [J]. Journal of College of Adult Education, Hebei University of Technology, 2009(3):4. (in Chinese)
- [8] Shang Wei, Champion Luo. Study on academic prediction mechanism of local college students [J]. Contemporary Educational Practice and Teaching Research: Electronic Edition, 2016(9):2. (in Chinese)
- [9] He Kui. A study on the construction of academic prediction mechanism for college students under the credit system [J]. Journal of Hubei Adult Education University, 2017, 23(3):4.(in Chinese)
- [10] Wu Haifeng, CHENG Yusheng, HU Xuegang. Design of university achievement prediction model based on data mining technology [C]// 2010.(in Chinese)
- [11] Wan Xinghuo, Zheng Junling, Jin Yongchao. Academic prediction model based on KPCA and its application [J]. Mathematical Theory and Applications, 2013, 33(4):6.(in Chinese)
- [12] Zhang Fusheng, Lu Kaidong, Han Yijia. Research on the Academic Monitoring and Prediction System of College Students Based on Campus Cloud [J]. China Education Informatization: Higher Vocational Education, 2015(5):3.(in Chinese)
- [13] Jin Yifu, Wu Tao, Zhang Zishi, et al. Design and Analysis of Academic Prediction System in Big Data Environment [J]. China Audio-Visual Education, 2016(2):5. (in Chinese)
- [14] Du Juan and Zhai Sheping. Research on College Student Achievement Prediction System Based on improved Apriori Algorithm [J]. Modern Information Technology, 2018, 2(2):3. (in Chinese)
- [15] Cui Qiang, Sun Jiyan. The Application of Big Data in Promoting the Quality of Ideological and Political Education in Colleges and Universities -- Taking Dalian University of Technology Academic Prediction Model as an Example [J]. Journal of College Counselors, 2019, 11(5):5.(in Chinese)