Hindawi
Mobile Information Systems
Volume 2022, Article ID 7846247, 10 pages
https://doi.org/10.1155/2022/7846247

Research Article

Application of GA-BP Neural Network in Online Education Quality Evaluation in Colleges and Universities

Guodong Sun

Department of Construction Engineering, Shijiazhuang University of Applied Technology, Shijiazhang, Hebei 050081, China

Correspondence should be addressed to Guodong Sun; 2003100403@sjzpt.edu.cn

Received 19 February 2022; Revised 15 March 2022; Accepted 22 March 2022; Published 20 May 2022

Academic Editor: Chia-Huei Wu

The purpose of educational data mining is to find the internal relations and laws hidden in the massive educational data and help students' learning and teachers' teaching and management. The prediction and analysis of student achievement can improve the way of training students and promote the improvement of teaching quality. This paper proposes a student grade prediction algorithm based on a DTGA-BP (decision tree genetic algorithm-back propagation) neural network to predict better and analyze students' grades. Firstly, the algorithm preprocesses the data with a correlation analysis method to generate the initial population. Then, the feature selection of evaluation indexes is carried out through DTGA decision tree, and the number of hidden layer neurons is optimized. Finally, the crossover probability and mutation threshold of the BP neural network are used to optimize the initial weight. The experimental results show that the prediction results of this algorithm are more consistent with the actual results, more scientific and accurate than the traditional methods, and can provide better services for teaching and management.

1. Introduction

With the development and maturity of database technology, the amount of data generated in people's daily activities such as life, study, and work increase exponentially. Data storage has exceeded the ability to use data, so it is of great significance to extract meaningful information from a large amount of data [1]. The development of data mining technology has gradually expanded its application fields and has been successfully applied in business, healthcare, marketing, retail, manufacturing, justice, engineering, and other areas [2]. Data mining technology plays a vital role in stock prediction, anomaly detection, customer group division, and other issues [3]. However, in education, there are few successful cases of research based on data mining technology [4]. With computer technology and network technology development, student information management systems and teaching management systems can be popularized and promoted. The teaching data in the teaching platform is proliferating, but data utilization is still confined to simple query and screening [5]. Students' scores are often dealt with in simple operations such as average score, maximum score, and mean square deviation, and the relationship between scores cannot be found [6]. For example, what factors affect students' performance are the key factors. Are all problems worthy of in-depth exploration, let alone predicting the future development trend of students from the current students' performance and behavior [7]?

To improve teaching quality and formulate reasonable talent training programs, many scholars began to pay attention to the application of data mining technology in the field of education and teaching [8]. Educational data mining refers to the process of data analysis in which data mining technology is used to automatically retrieve many teaching data and sort out relevant information. Thus, adequate potential information in teaching data can solve problems in teaching practice [9]. It provides teaching suggestions and theoretical support for teaching administrators, teachers, students, and other related personnel in education. Prediction of student performance is a vital content in education data mining, and student performance is an essential basis for evaluating teaching quality [10]. With the popularization of information technology, many students' homework scores and test scores have been accumulated in the form of electronic data in the school teaching management system and curriculum operation system [11]. It is worth studying how to mine these data effectively to improve teaching quality.

At present, the prediction and analysis of students' scores and the mining of key influencing factors of scores have attracted the attention of scholars at home and abroad [12]. Regarding student performance prediction, literature [13] selected several typical learning behavior characteristics from many behavioral aspects of MOOC learners and use the desired features to predict whether learners can complete the learning task and obtain the certificate to find out the potential serious learners. In literature [14], eight important attributes are selected by calculating the information gain rate of each attribute among the 18 features that affect students' performance, and the eight essential attributes are used to construct a decision tree to predict students' performance. In terms of the mining of influencing factors of students' scores, literature [15] conducted a study on the scores of 300 students in an Indian university and found that students' scores were greatly influenced by factors such as home address, family annual income, mother's education, living habits, and students' historical scores. Literature [16] proposed that students' sociodemographic characteristics (such as race, gender, and economic status) and academic characteristics (such as school type and school performance) are closely related to their academic performance. Literature [17] proposed that the principal component analysis-radial basis function, principal component analysis (PCA), is used to reduce the data dimension. Although the prediction accuracy is improved, the influencing factors of students' scores are not obtained, reducing the model's comprehensibility. Literature [18] uses the BP neural network to predict students' scores, but the prediction accuracy of the BP neural network is low. Literature [19] puts forward a prediction model for students' grades, which uses the method based on feature selection regression and neural network to predict students' grades. Literature [20] proposed a modified GM (1, 1) model based on the variation of the difficulty coefficient of the test paper. It only takes the difficulty coefficient of the test paper as the standard to predict students' test scores, but the model is not practical enough.

Due to individual differences, students will have significant differences in learning results. The teaching quality can be improved if the students' future grades can be predicted by using their existing grades, and teaching strategies can be appropriately changed, and students can be given hints based on the expected results. This paper presents a student achievement prediction algorithm based on the DTAG-BP neural network. The algorithm simplifies the input by calculating the correlation coefficient. By improving the structure of neural network based on decision tree, the optimal decision tree is generated, and the number of hidden layer nodes and the optimal feature combination are determined. At the same time, a genetic algorithm is introduced to optimize the initial value and threshold of the BP neural network, and the training samples of fitness function are selected to meet the target requirements. The optimized DTAG-BP neural network prediction model can predict students' grades. At the same time, provide online classroom quality feedback for teachers, to better understand students' knowledge and improve teaching efficiency.

2. The Algorithm Is Proposed in This Paper

2.1. Correlation Analysis. Correlation analysis is a statistical method to study the relationship between observed values. At the same time, the correlation direction and degree of dependence among variables are discussed, and the correlation among random variables is analyzed. In this paper, the correlation analysis method is used to calculate the correlation between the grades of introductory courses and the grades of target courses. The influence degree of the grades of introductory courses on the grades of target courses is analyzed.

The Pearson correlation coefficient is generally used to study correlation in correlation analysis. It is defined as shown in

$$\rho = \frac{\sum_{x=1}^{N} (i_x - i)(j_x - j)}{\sqrt{\sum_{x=1}^{N} (i_x - i)^2 \sum_{x=1}^{N} (j_x - j)^2}}. \tag{1}$$

In the formula, let i_x be the result of a basic course of student $x(x = 1, 2, 3 \cdots, N)$. i is the average grade of the foundation course. j_x is the target course grade of $x(x = 1, 2, 3 \cdots, N)$ student. j is the average score of the target course. N is the total number of students collected. The value of ρ ranges from [-1,1]. If the value is greater than 0, it indicates a positive correlation. If the value is less than 0, it indicates a negative correlation. If the value is 0, it indicates no correlation.

2.2. BP Neural Network Model. The back propagation (BP) neural network is one of the most widely used classification models in neural network algorithms. As its application range increases, the BP neural network begins to show some problems that cannot be solved by itself. For example, it is easy to fall into a local minimum, the learning convergence speed is slow, and there is no practical method for selecting input parameters [21]. The design of the BP neural network model mainly includes the input layer, hidden layer, and output layer, and its structure is shown in Figure 1. The gradient descent method is used to learn and correct, and the mean square error is minimized. Then, the weights are adjusted continuously during the training process. The learning process is divided into the following steps.

Step1: input the sample data of the training set and calculate the output values of neurons of each layer from the input layer to the output layer according to the structure of the BP neural network and the determined weights and thresholds

Step2: calculate the error between the predicted output value and the expected output value and constantly adjust the weight from the input layer to the hidden layer and from the hidden layer to the output layer

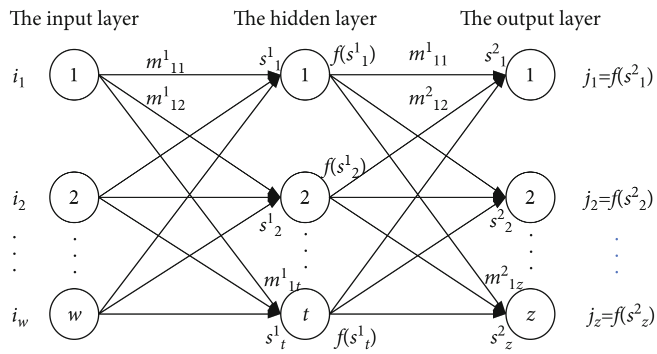

Figure 1: Structure of the BP neural network.

Step3: repeat Step1 and Step2. When the error value within the specified range is reached or the training times are over, the learning is completed

The training process is divided into the following steps:

Step1: initialize the weights. The formula is as follows:

$$M_y(s) = \left(M_{0y}(s), M_{1y}(s),\cdots,M_{ty}(s)\right), y \in [1, w], \quad (2)$$

where t is the number of nodes in the input layer, w is the number of nodes in the hidden layer, s is the number of learning steps, and $M_y(s)$ is the weight matrix after learning s steps

Step2: input training sample set $\{I_u, d_u\}$, where $I_u$is the input vector, $d_u$is the output vector, and u is the current serial number of the training sample

Step3: calculate the actual output value. The formula is as follows:

$$O_y^u = f\left(M_y^S(s)I_u\right) \quad (3)$$

Step4: update the weights of each neuron. The formula is as follows:

$$M_y(n+1) = M_y(n) + \beta\left[d_y^u - o_y^u(n)\right]I_u, \quad (4)$$

where β represents the learning rate, $\beta \in [0, 1]$, and is used to update the speed.

Step5: repeat Step2. When the error value or training times within the specified range is reached, the training is completed

2.3. DTGA-BP Combined Model. The DTGA-BP combined model uses the decision tree algorithm and genetic algorithm to optimize the BP neural network. The decision tree algorithm improved the selection and structure of input parameters of the neural network. The genetic algorithm optimized the initial weights of the neural network with improved selection strategy and crossover mutation operation. The core idea is to use the optimized BP neural network to establish the regional independent innovation capability evaluation and classification model.

By using a decision tree algorithm to improve the structure of the neural network, the complexity of the neural network can be reduced. The training time can be shortened without affecting the classification accuracy, and the operation method is easy to realize. Combined with the optimized genetic algorithm, it overcomes the shortcomings of the traditional BP neural network, such as easy to fall into the local optimum and making the initial weight more reasonable. The DTGA-BP algorithm flow is shown in Figure 2.

2.3.1. Improve the Structure of BP Neural Network Based on Decision Tree. The evaluation model uses a single hidden layer neural network structure. Therefore, the improvement of the neural network structure is mainly to determine the number of hidden layer nodes. The traditional method to determine the number of remote layer nodes cannot effectively solve the complex system of the BP neural network. Inspired by "entropy network" [22], this paper adopts the number of nonleaf nodes of the longest rule chain in the generated decision tree as the number of hidden layer nodes, reducing the model complexity and training time.

Firstly, the C4.5 decision tree algorithm generates a decision tree for judging water quality. The wine dataset downloaded from the UCI machine learning library is used as the dataset. The dataset contains 13 characteristic attributes, such as pH, hardness, and alkalinity, all with continuous values. Then, it is judged according to attribute conditions and classified from root node to leaf node. The decision tree is finally generated by calculating the information gain value, and the pruning strategy is shown in Figure 3.

Figure 3 shows that the number of leaf nodes in the decision tree is 5, and the number of nonleaf nodes in the longest rule chain is 3. Therefore, BP neural networks constructed according to the number of inputs, hidden, and output layer nodes are 3-5-1 and 3-3-1, respectively. The standard normal distribution formula determines the initial weight. After training, 200 test sets were randomly generated and five tests were conducted. The results are shown in Table 1. The number of nodes equals 3 is the method adopted in this paper to determine the number of nodes in the hidden layer. The number of nodes equals 5 is the number of nodes in the hidden layer defined by the "entropy network." It can be seen from the classification results that compared with the traditional "entropy network" method, the method adopted in

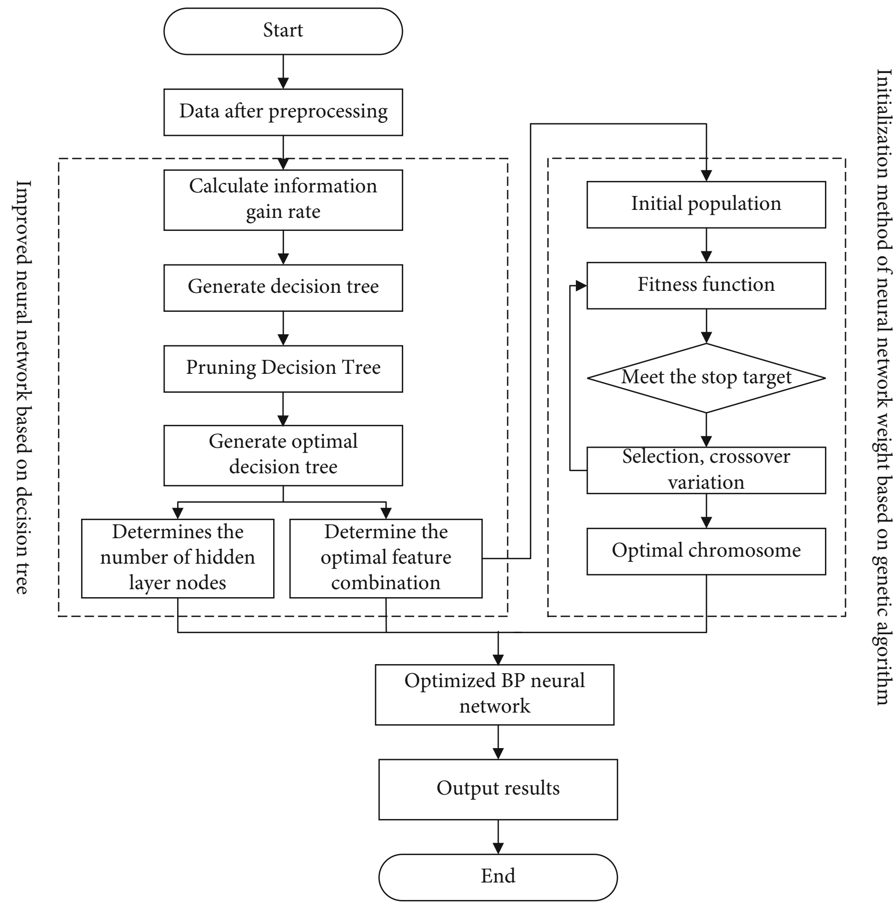

Figure 2: Flow chart of the DTGA-BP algorithm.

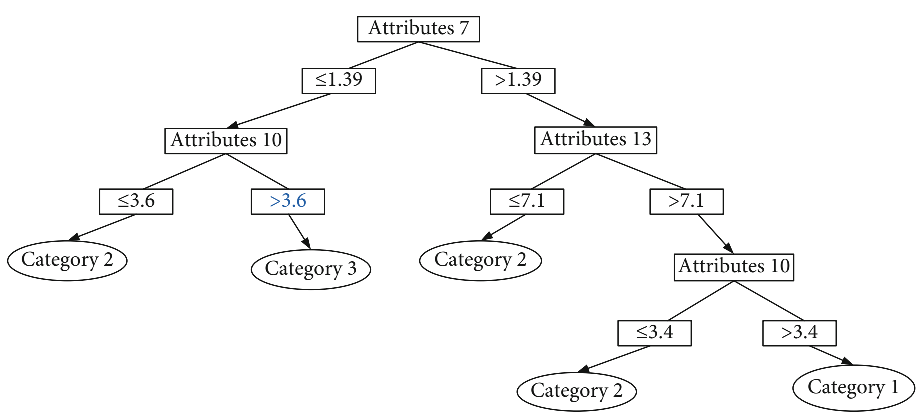

Figure 3: Decision tree constructed by the C4.5 algorithm.

this paper to determine the number of hidden layer nodes can speed up the retrieval speed and reduce the error.

2.3.2. Determine the Initial Weight of BP Neural Network Based on Genetic Algorithm. To solve the local optimization problem caused by improper initial weight adjustment of the BP neural network, genetic algorithm (GA) global search is generally used to solve the problem [23]. But the traditional genetic algorithm still has some issues in optimization, such as low efficiency and slow speed. So, this paper uses a genetic algorithm with an improved selection operator and crossover and mutation operation to determine the initial weight of the neural network. The enhanced genetic algorithm flow chart is shown in Figure 4.

2.3.3. Optimize the Selection Operator. There are many selection strategies in genetic manipulation. An appropriate selection strategy can improve the performance of the genetic algorithm. Therefore, the selection operator should not only prevent the local optimum caused by the

Table 1: Classification results of nodes in different hidden layers.

Experiment number	MSE		Training time (s)	
	Number of nodes = 3	Number of nodes = 5	Number of nodes = 3	Number of nodes = 5
1	0.0059	0.0816	0.0952	0.0541
2	0.006	0.0717	0.0666	0.9775
3	0.0037	0.0032	0.0465	0.0523
4	0.0088	0.0289	0.0788	0.0511
5	0.0045	0.1556	0.4476	0.0588
Average	0.0058	0.0682	0.1469	0.2388

prematurity of the population but also cannot be too divergent and challenging to converge, so it needs to be balanced. Based on this, this paper adopts the improved selection operator method.

Firstly, the fitness value of everyone was calculated according to the fitness function. After sorting, only the former individuals were retained and the latter individuals with small fitness values were eliminated. Then, the former population number of retained individuals was used as the paternal parent, and finally, only the intermediate individuals were genetically manipulated. The selection probabilities of individuals in these populations are calculated by the following formula:

$$\begin{cases} U_t^N = V^N\left(1 - v^N\right)^{t-1}, \\ V^N = \dfrac{v^N}{1 - \left(1 - v^N\right)^{T/2}}, \end{cases} \tag{5}$$

where U is the individual selection probability, v is the optimal selection probability, T is the population number, and T is the individual serial number. $T = 1, 2, 3, \cdots, t/2, N$ is the current iteration number. In this way, the good paternal genes are retained, and the individuals with low fitness are eliminated to balance the population. Finally, the optimal global solution can be obtained. In the early stage of evolution, the individual difference between populations is large, and the corresponding q value is also large. Only in this way can we ensure that the individuals with high fitness are selected with a high probability, to select excellent individuals for the population. However, as the population continues to evolve and the number of the population continues to decrease, the difference between individuals also gradually decreases, and the value of V obtained at this time should also decrease. Based on this analysis, the v value is calculated according to

$$v^N = v_{\max} - \left(v_{\max} - v_{\min}\right) \times \frac{N - 1}{W - 1}, \tag{6}$$

where $v_{\max}$ represents the probability of optimal individual selection, $v_{\min}$ represents the probability of worst individual selection, and W represents the total number of iterations.

According to the selection operator operation in this way, the individuals with the highest fitness are retained to the next generation of population, but the optimal individuals will not be eliminated by genetic operation to ensure global convergence. The fitness values of the worst individuals recorded each time were compared, and those with lower fitness were kept and added into the new population. By combining the optimal preservation strategy with the worst preservation strategy, the selection error can be reduced, the diversity of the population can be maintained, and the optimal solution can be obtained.

2.3.4. Improve Crossover and Mutation Operations. In traditional adaptive genetic algorithms, the random probability of crossover and mutation will be more significant, interfering with an individual's quality in the genetic algorithm. The genetic algorithm will be trapped into local optimization. And when the individual adaptation level of crossover and mutation reaches the maximum value, the probability of crossover and mutation will not exist, and the individual population will be in a state of complete stagnation. To solve this problem, this paper adopts the improved crossover rate $U_c$and mutation rate U_w for genetic operation, and the formula is as follows:

$$U_c = \begin{cases} z_1 \dfrac{\arcsin\left(f_g/f_w\right)}{\pi/2}, & \arcsin\left(\dfrac{f_g}{f_w}\right) < \dfrac{\pi}{6}, \\ z_1\left(1 - \dfrac{\arcsin\left(f_g/f_w\right)}{\pi/2}\right), & \arcsin\left(\dfrac{f_g}{f_w}\right) \geq \dfrac{\pi}{6}, \end{cases} \tag{7}$$

$$U_w = \begin{cases} z_2 \dfrac{\arcsin\left(f_g/f_w\right)}{2}, & \arcsin\left(\dfrac{f_g}{f_w}\right) \geq \dfrac{\pi}{6}, \\ z_2\left(1 - \dfrac{\arcsin\left(f_g/f_w\right)}{\pi/2}\right), & \arcsin\left(\dfrac{f_g}{f_w}\right) < \dfrac{\pi}{6}, \end{cases} \tag{8}$$

where f_g represents the average individual fitness, and f_w represents the maximum individual fitness. arcsin (f_g/f_w) changes rapidly with f_g. Therefore, selecting arcsin $(f_g$

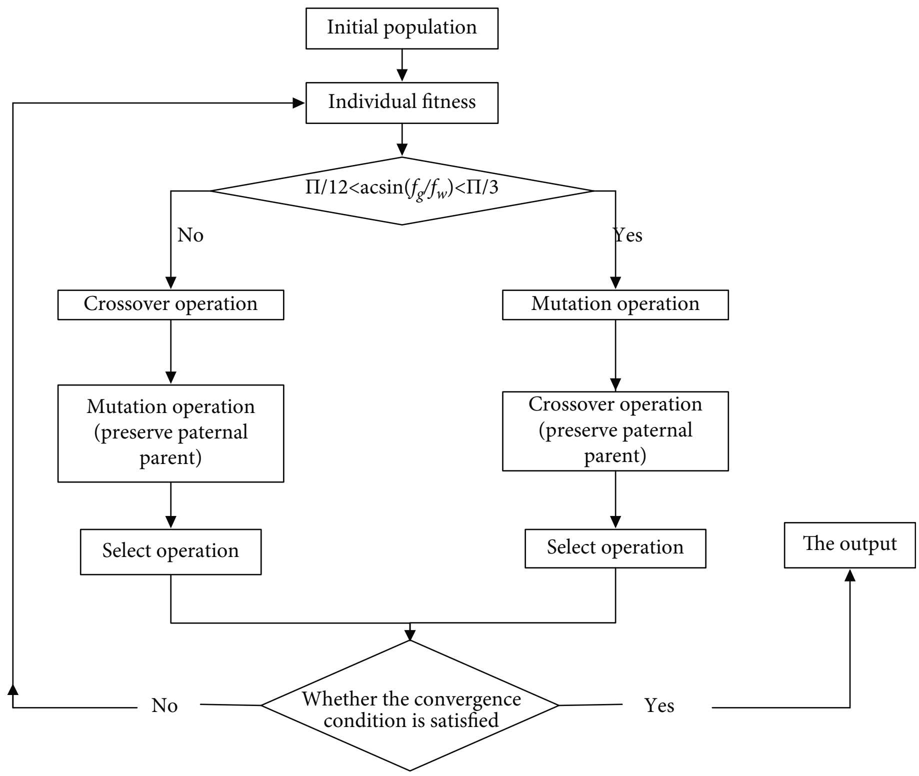

Figure 4: Improved flow chart of the genetic algorithm.

Table 2: Correlation coefficient between introductory courses and target courses.

Entry term	Correlation coefficient
Advanced mathematics all course score	0.52
Advanced mathematics Al course score	0.39
Complex function course score	0.39
Basic A in circuit analysis course score	0.40
University physics lab Al course score	0.36
Linear algebra A grade course score	0.38
Ideological and moral cultivation and law foundation course score	0.33
University physics AI course score	0.28
Circuit basic experimental course score	0.27
College English II course score	0.23
College English 1 course score	0.16
College physical education course score	0.18

$/f_w)$ as the judgment condition can play a good role in judging the degree of dispersion among population fitness. And because $\sin(\pi/6) = 1/2$, $\arcsin(f_g/f_w) \geq \pi/6$ when $\arcsin(f_g/f_w) \geq \pi/6$ means that f_g is close to f_w. Finally, according to the condition $\pi/12 \leq \arcsin(f_g/f_w) \leq \pi/3$, determine whether to perform the crossover operation first. If it does not meet the condition, it carries on the mutation operation first. The traditional operation of genetic mutation is always to cross before mutation, which is easy to lead to the slow generation of good individuals, or even destroy the good individuals. The method adopted in this paper can solve this problem well.

3. Experimental Part

3.1. Data Collection. Data were collected from 99 students majoring in communication engineering in a university. In the first year of college and university second grade 12 courses, higher mathematics AI, respectively, AII higher mathematics, college physics AI, A linear algebra, complex

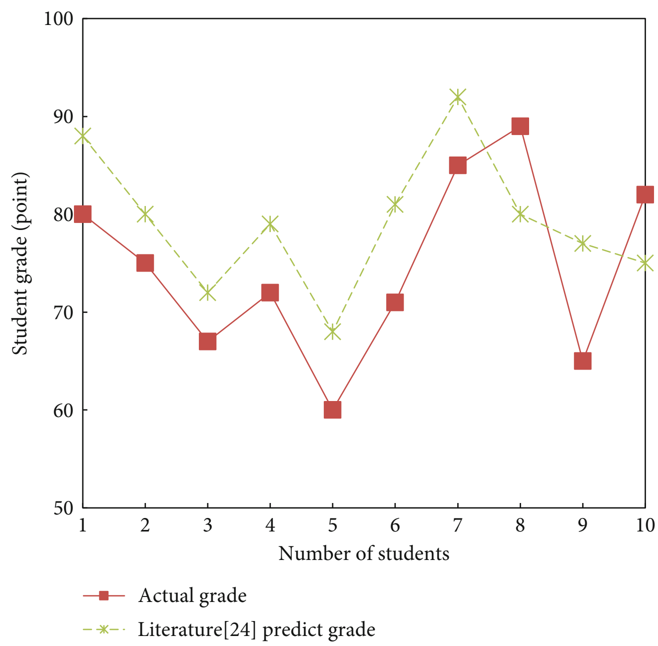

Figure 5: Broken line of the actual predicted value of literature [24]'s model.

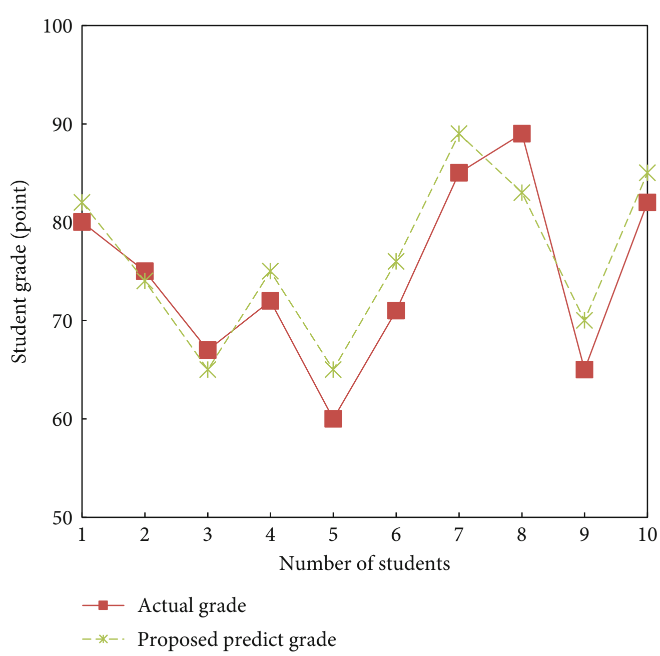

Figure 6: Broken line of the actual predicted value of the DTGA-BP neural network model.

variable function, AI university physics experiment, a basis of circuit analysis, circuit experiment, college English, college English II, thought morals tutelage and legal foundation, and college sports scores are used. At the same time, the final exam scores of the core introductory professional courses, namely, signal and system, were collected. For the initially collected data, do data preprocessing. All the students missing in this exam were deleted, the data of the remaining 80 students are then normalized; the data range was$[0 \sim 100]$to get the final student performance data.

3.2. Correlation Analysis Data Processing. Signal and system were used as a predictive target course. Using formula (1), the results of 12 basic courses of 80 students majoring in communication engineering after data pretreatment can be obtained. The correlation coefficients between the signal and system of target courses are shown in Table 2. It can be seen from Table 2 that the course scores of advanced mathematics AII, advanced mathematics AI, and complex variable function are highly correlated with the target course signal and system scores. However, the correlation between college PE, college English I, college English II, and target course scores is relatively low.

3.3. Experimental Analysis of Student Achievement Prediction. MATLAB R2016a software was selected as the simulation tool, and literature [24]'s model and the model in this paper were used for prediction, respectively. The test set and training set were aet, in which the results of the last 10 students are taken as the test set, and the results of the remaining 70 students are taken as the training set.

3.3.1. Literature [24] Prediction Analysis of Student Achievement. Literature [24]'s model is used to predict students' scores. The initial students' basic course scores are 12, and the input layer neurons of the neural network are set as 12. The signal and the system's performance data are taken as the output, and the neuron of the output layer is set as 1. According to the correlation algorithm, the number of hidden layer neurons can be obtained from 4 to 14, because the number of hidden layer nodes determines the nonlinear mapping ability of the neural network structure. To make the neural network have better prediction performance, the neuron of the hidden layer is set as 14, the number of cycles is set as 3000, and the training error is 0.001. The learning rate determines the weight change in the training process. Here, the learning rate is set as 0.2 [25], and the line chart of the actual value and predicted value is obtained as shown in Figure 5.

The actual and predicted values obtained by literature [24]'s model are analyzed. Student 1 scored 80 on signals and systems and 88 on prediction. Student 2's real grade was 75, and the prediction was 80. Student 3's actual score was 67, with a predicted score of 72. Student 4's actual grade was 72, and the prediction was 79. Student 5's actual grade was 60, and the prediction was 68. Student 6's real grade was 71, and the prediction was 81. Student 7's actual grade was 85, and the prediction was 92. Student 8's real grade was 89, and the prediction was 80. Student 9's real grade was 65, and the prediction was 77. Student 10's actual grade was 82, with a predicted score of 75. Literature [24]'s initial weights and thresholds are randomly set due to the low correlation between some input and output terms. There is a significant error between the predicted result and the real result of each student, so the model still has some shortcomings in the prediction of the student's achievements.

3.3.2. The Algorithm Is Used to Predict Students' Scores. In the performance prediction of the algorithm in this paper, the correlation analysis method is adopted to select advanced mathematics All, advanced mathematics A, complex function, basic circuit analysis A, university physics experiment AI, and linear algebra A whose correlation coefficient is greater than 0.35. The scores of the 6 basic courses are taken as the input items, so the neurons of the input layer and the hidden layer of the neural network are set as 6 and 13, respectively. The performance data of the signal and the system are taken as the output items. The neuron of the output layer is set as *i*. The MATLAB simulation tool is used to realize the codec function and fitness function of the algorithm in this paper. The parameters of the algorithm in this paper and literature [24] are set to the same value. At the same time, initialization and basic parameters of the proposed algorithm were set. The population size of the proposed algorithm generally ranges from 10 to 200. If the population size is too large, the convergence performance of the result is poor. Inbreeding will occur if the population size is too tiny, producing pathological genes. Therefore, the initial population size was set as 100, the maximum genetic algebra was set as 50, and the crossover probability was set as 0.3 ~ 0.9. If too easy to miss the best individual mating probability, the probability is too slight and cannot effectively update population mating, a crossover probability is 0.5, and mutation probability set the range of 0.001~0.2 in order to guarantee the diversity of population does not destroy the existing population patterns; at the same time, a mutation probability is 0.01, a network model in this paper the actual value and predicted value line chart, as shown in Figure 6.

Table 3: Comparison of prediction results of the two models.

Student's actual score/ point	Literature [24] predicted value/ point	DTGA-BP neural network predicted value/point
80	88	82
75	80	74
67	72	65
72	79	75
60	68	65
71	81	76
85	92	89
89	80	83
65	77	70
82	75	85

As can be seen from Figure 6, student 1's signal and system score is 80, and the predicted score is 82. Student 2's real grade was 75, and the prediction was 74. Student 3 had an actual score of 67 and a prediction of 65. Student 4's real grade was 72, and the prediction was 75. Student 5's real grade was 60, with a predicted score of 65. Student 6's real score was 71, with a predicted score of 76. Student 7's actual grade was 85, with a predicted 89. Student 8's real grade was 89, and the prediction was 83. Student 9's actual grade was 65, and the prediction was 70. Student 10's actual grade was 82, and the prediction was 85. The results predicted by the model in this paper are closer to the actual results. By comparing the prediction results of literature [24]'s model with that of the model in this paper, the model in this paper has smaller prediction error. The comparison of the predicted values of the two models is shown in Table 3.

The root mean square error *e* of the two models was calculated according to the predicted and actual values of the two models in Table 3.

$$e = \sqrt{\frac{1}{V}\sum_{x=1}^{V} \left(j_x - o_x\right)^2}, \tag{9}$$

where V is the predicted number of students, which is 10 in this paper. The real grades of middle school students in Table 3 as j_x was obtained. The prediction result of literature [24] and the prediction result of the model in this paper were, respectively, taken as o_x, and the root mean square error of the prediction result of literature [24] model was solved to be 7.9, and the root mean square error of the prediction result of the network model in this paper was 3.1. Through the comparison of root mean square error, the prediction accuracy of the proposed model is higher than that of the literature [24] model. It is more suitable for student achievement prediction.

4. Conclusion

To improve higher teaching quality, student performance prediction has become a research focus in educational data mining and one of the essential objectives of learning analytics. This paper proposes a student's score prediction algorithm based on the DTGA-BP neural network to predict students' scores more accurately. Based on the BP neural network prediction model, the decision tree genetic algorithm is used to optimize the correlation coefficient between the fundamental course score and the target course score. The fundamental course score is removed with a low correlation with the target course score. By combining genetic algorithm with the BP neural network, the optimal initial weights and thresholds of the BP neural network were updated, and a DTGA-BP neural network student performance prediction model was established. By collecting the actual basic course scores of 99 students majoring in communication engineering in a university, the algorithm in this paper is used to predict the basic course scores of these students. Experimental results show that compared with the BP neural network model, the root mean square error of the TDGA-BP neural network model is reduced from 7.9 to 3.1, with higher prediction accuracy, which verifies the effectiveness of the proposed algorithm.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This work is supported by the Shijiazhuang University of Applied Technology.

References

[1] Z. Jensen, E. Kim, S. Kwon et al., "A machine learning approach to zeolite synthesis enabled by automatic literature data extraction," *ACS Central Science*, vol. 5, no. 5, pp. 892–899, 2019.

[2] Y. Lu, "Artificial intelligence: a survey on evolution, models, applications and future trends," *Journal of Management Analytics*, vol. 6, no. 1, pp. 1–29, 2019.

[3] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers & Education*, vol. 143, article 103676, 2020.

[4] X. Wu, B. Feng, and W. Qi, "Design and Implementation of a Novel Student Information Management System[C]//2020 IEEE," in *3rd International Conference on Information Systems and Computer Aided Education (ICISCAE).*, pp. 637–639, Dalian, China, 2020.

[5] P. Smutny and P. Schreiberova, "Chatbots for learning: a review of educational chatbots for the Facebook Messenger," *Computers & Education*, vol. 151, article 103862, 2020.

[6] D. McNeish and M. G. Wolf, "Thinking twice about sum scores," *Behavior Research Methods*, vol. 52, no. 6, pp. 2287–2305, 2020.

[7] Q. K. Fu and G. J. Hwang, "Trends in mobile technology-supported collaborative learning: A systematic review of journal publications from 2007 to 2016," *Computers & Education*, vol. 119, pp. 129–143, 2018.

[8] T. Chen, J. Rong, L. Peng, J. Yang, G. Cong, and J. Fang, "Analysis of social effects on employment promotion policies for college graduates based on data mining for online use review in China during the COVID-19 pandemic," *Multidisciplinary Digital Publishing Institute*, vol. 9, no. 7, p. 846, 2021.

[9] J. König, D. J. Jäger-Biela, and N. Glutsch, "Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany," *European Journal of Teacher Education*, vol. 43, no. 4, pp. 608–622, 2020.

[10] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: a review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, 2019.

[11] E. A. Patall, H. Cooper, and J. C. Robinson, "Parent involvement in homework: a research synthesis," *Review of Educational Research*, vol. 78, no. 4, pp. 1039–1101, 2008.

[12] C. Bouquet and J. Birkinshaw, "Weight versus voice: how foreign subsidiaries gain attention from corporate headquarters," *Academy of Management Journal*, vol. 51, no. 3, pp. 577–601, 2008.

[13] J. Zhuoxuan, Z. Yan, and L. Xiaoming, "Learning behavior analysis and prediction based on MOOC data," *Journal of computer research and development*, vol. 52, no. 3, p. 614, 2015.

[14] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, "Comparative study of attribute selection using gain ratio and correlation based feature selection," *International Journal of Information Technology and Knowledge Management*, vol. 2, no. 2, pp. 271–277, 2010.

[15] T.-T.-H Le, T. Tran, T.-P.-T Trinh et al., "Reading habits, socioeconomic conditions, occupational aspiration and academic achievement in Vietnamese junior high school students," *Sustainability*, vol. 11, no. 18, Article ID 5113, 2019.

[16] S. Helal, J. Li, L. Liu et al., "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Systems*, vol. 161, pp. 134–146, 2018.

[17] S. B. Roh, S. K. Oh, E. K. Park, and W. Z. Choi, "Design of radial basis function neural networks with principal component analysis and linear discriminant analysis for black plastic

identification[C]//2016," in *Joint 8th International Conference on Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems (ISIS).*, pp. 764–768, Sapporo, Japan, 2016.

[18] W. Liu, "An improved back-propagation neural network for the prediction of college students' English performance," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 16, 2019.

[19] X. Li, L. Xie, and H. Wang, "Grade prediction in MOOCs[C]// 2016," in *IEEE Intl Conference on Computational Science and Engineering (CSE) and IEEE Intl Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES)*, pp. 386–392, Paris, France, 2016.

[20] P. B. Lowry and J. Gaskin, "Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use It," *IEEE Transactions on Professional Communication*, vol. 57, no. 2, pp. 123–146, 2014.

[21] Y. Zhang, "Application of improved BP neural network based on e-commerce supply chain network data in the forecast of aquatic product export volume," *Cognitive Systems Research*, vol. 57, pp. 228–235, 2019.

[22] H. He, Y. Zheng, B. A. Bernevig, and N. Regnault, "Entanglement entropy from tensor network states for stabilizer codes," *Physical Review B*, vol. 97, no. 12, article 125102, 2018.

[23] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021.

[24] C. Hu and C. Z. Wang, "Research on student achievement prediction based on BP neural network method," *Advances in Artificial Systems for Medicine and Education IV*, vol. 1315, no. 409, p. 293, 2021.

[25] T. Hua, W. Wang, Z. Xue, S. Ren, Y. Wang, and H. Zhao, "On feature decorrelation in self-supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9598–9608, Montreal, Canada, 2021.