Check for updates

Intelligent Fusion Method for College Students' Psychological Education Score Data Based on Improved Bp Algorithm

Liang Zhang(✉) and Yu Zhao

Changchun University of Finance and Economics, Changchun 130000, China
zliang1510@163.com

Abstract. In order to solve the problem of poor fusion of college students' psychological education score data, an intelligent fusion method for college students' psychological education score data based on improved BP algorithm is proposed. Innovatively adopting genetic algorithms to optimize the weights and thresholds of BP neural networks; Standardize the hierarchical data using matrix transformation and preprocess the fused data using the Grobeis criterion; Extract multi-source heterogeneous level data features, establish an intelligent fusion model for college students' psychological education level data based on improved BP neural network, and achieve intelligent fusion of college students' psychological education level data. The experimental results show that this method has lower latency and higher packet switching rate in data fusion, proving its good fusion performance in data fusion.

Keywords: Improving BP Algorithm · Psychological Education Scores · Data Intelligent Fusion

1 Introduction

As a special group in society, college students are facing various pressures and increasing mental health problems, leading to the emergence of extreme events. Therefore, the mental health status of college students, as well as the psychological intervention models and methods, have become a research hotspot. The use of mental health education content to guide the mental health of college students and identify the existing problems is an effective method to solve mental health problems. With the continuous expansion of enrollment in universities, current university education is transitioning from "elite education" to "mass education". Due to the increasing number of college students facing psychological problems, mental health education for college students is gradually being valued by major universities. Psychological health education for college students is not limited to the study of theoretical knowledge, just like other subjects. Instead, targeted teaching of psychological health knowledge and psychological counseling work are carried out based on the psychological characteristics of this special group of college students. For example, conducting separate psychological counseling or organizing

L. Yun et al. (Eds.): ADHIP 2023, LNICST 547, pp. 64–74, 2024.
https://doi.org/10.1007/978-3-031-50543-0_5

group counseling with classmates with similar psychological problems to help college students alleviate psychological stress, prevent mental illness, adapt to the environment well, handle interpersonal relationships well, help them solve problems in personality development and emotional regulation, and fully tap their psychological potential. With the modernization of national education, the external environment and internal needs of teachers' teaching and research have undergone changes [1]. In order to solve the above psychological education problems in colleges and universities, the networked and information-based working mode has gradually entered colleges and universities. However, students' mental health is a comprehensive reflection of many factors, and the common psychological test data is relatively single, leading to a certain degree of subjectivity in the test results. Moreover, during the use of the system, most teachers, counselors, and mental health educators are limited to data addition, deletion, and modification, without systematically integrating the data in the system. As a result, important data attributes have not been effectively applied, and active warning mechanisms cannot be established, failing to play a proactive warning role. Therefore, it is necessary to intelligently integrate the psychological education performance data of college students.

The purpose of data fusion technology is to obtain simpler and more accurate judgments through multi-source information, that is, by integrating multiple types of data with information from relevant databases to obtain more accurate data. Reference [2] proposed a data fusion algorithm based on distributed Compressed sensing and hash function. Firstly, the distributed Compressed sensing method is used for sparse observation of sensing data to remove redundant data; Secondly, a one-way hash function is used to obtain the hash value of the observed values of the perception data, and it is filled with unrestricted disguised data into the observation values of the perception data to achieve the purpose of hiding the real perception data; Finally, after extracting the camouflage data from the aggregation node, the hash value of the perception data is obtained again and data fusion is completed. Reference [3] designed a deep learning network (DLN) model and its training method, and further proposed a multi-source data fusion algorithm based on DLN-DS.

When the above methods are applied to the data fusion of college students' psychological education achievements, there are problems such as slow Rate of convergence, low fusion efficiency, etc., and the effect of intelligent fusion is poor. To this end, an improved intelligent fusion method for college students' psychological education score data based on BP algorithm is proposed. Innovatively optimize the weights and thresholds of BP neural network through genetic algorithm to obtain a stable network structure and solve the problem of slow Rate of convergence. Standardize the hierarchical data using matrix transformation and preprocess the fused data using the Grobeis criterion; Extract multi-source heterogeneous level data features, establish a data intelligent fusion model based on improved BP neural network, and complete the intelligent fusion of college students' psychological education level data. The test results demonstrate that this method has the following contributions: it reduces fusion delay, improves packet switching rate and fusion efficiency, and has good fusion performance. The overall structure of the current study is shown in Fig. 1:

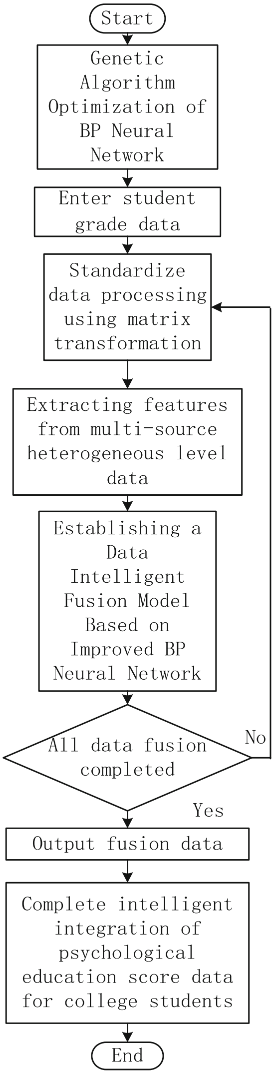

Fig. 1. The overall study structure

2 Improved BP Network Algorithm

Although BP neural network has great advantages in data prediction and image segmentation due to its simple operation and strong practicality, it inevitably has some problems, such as slow convergence speed, poor algorithm stability, and easy falling into local minima. Genetic algorithm performs well in global search. By optimizing the weights and thresholds of the BP neural network through genetic algorithm, a stable network structure can be obtained. It has good improvements and improvements in image segmentation processing, data calculation, and trend prediction, making it a worthwhile optimization solution. The structure of the BP neural network used for data fusion is shown in Fig. 2. Each layer has multiple neural network nodes with different thresholds, and the neural network node is the type of psychological performance data used for fusion.

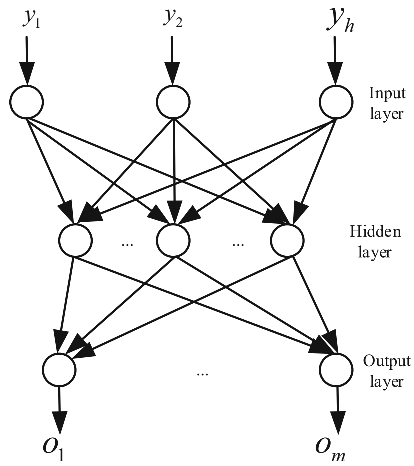

Fig. 2. The BP neural network structure

As shown in Fig. 2, the number of nodes in the output layer, hidden layer and input layer of BP neural network are set as m, n and h successively. The neuronal node connection weights of the input and hidden, hidden and output layers are ϖ_{ij}, ϖ_{jh}. Then the fusion result o_m of the output is:

$$o_m = \sum_{j=1}^{m} g\left(\sum_{i=1}^{h} \varpi_{ij}\varpi_{jh}y_i - \sigma_j\sigma_m\right) \tag{1}$$

In the formula, the input data is y_i; the hidden level threshold of the neural network node j and the output node threshold are σ_j and σ_m; the excitation function of the hidden level node is $g()$.

In order to fully leverage the advantages of genetic algorithm and BP neural network, this article improves and optimizes BP neural network through genetic algorithm, complementing each other's advantages. Genetic algorithm is a biological intelligence optimization algorithm. This solution set is the most ideal solution for the problem to be solved [4, 5]. The steps for solving practical problems using genetic algorithms are as follows:

(1) Initialize. Randomly select an initial population P (0) from N samples through genetic algorithm parameter encoding, forming a set of feasible solutions.
(2) Individual evaluation. The initial population P (0) is introduced into the objective function, and then the fitness of various populations in the sample is calculated.
(3) End condition judgment. Given an initial condition, determine the end condition of the algorithm, and if the condition is met, directly skip to step (8).

(4) Select an operation. Adopt the rule of survival of the fittest and survival of the fittest for the target group, and select a large number of excellent individuals.
(5) Mutation operation. Perform mutation operations on target individuals through mutation probability.
(6) Cross operation. Perform crossover operations on target individuals through crossover probability.
(7) After mutation and crossover random operations, if the next population P (t + 1) composed of N new individuals is obtained, proceed to step (2), and vice versa, proceed to step (4).
(8) Evolutionary individuals. Get the individual with the highest fitness in the function, which is the most ideal solution of the problem, and end the operation.

In order to fully leverage the advantages of genetic algorithm and BP neural network, this article improves and optimizes BP neural network through genetic algorithm, complementing each other's advantages. The specific steps are as follows:

(1) Initialization: Initialize the entire population, using parameter encoding to generate N individuals, which form a new initial population. Then, calculate the population size, determine the chromosome length and range, and provide the ideal error value.
(2) Calculation of fitness: fitness is calculated as follows:

$$G = \frac{1}{2}\sum_{h=1}^{M}(o_h - \hat{o}_h)^2 \tag{2}$$

In the formula, o_h and $\hat{o}_h$ are the basic fusion results and the actual fusion results of BP neural network. Because the value of the fitness function G must be non-zero and positive, the value of G is large in many states, indicating the high accuracy of the underlying result.

The individual is the feasible solution of each connection weight and threshold. If the size of the connection weight and threshold initial population is N, and the fitness of the i individual is G_i, then the possibility of the i individual being left is q_i:

$$q_i = \frac{G_i}{\sum_{i=1}^{N} G_i} \tag{3}$$

The cross-operator action is mating reconstruction, and in the study of this paper, it is mainly used to reconstruct the BP neural network structure. The cross-over operation method is:

$$\begin{cases} \varpi_{ij} = \beta\varpi_{jh} + \varpi_{ij} - \beta\varpi_{ij} \\ \varpi_{jh} = \beta\varpi_{ij} + \varpi_{jh} - \beta\varpi_{jh} \end{cases} \tag{4}$$

where, β is a random number, and its value interval is 0 to 1.

The mutation operator can fine-tune the gene of a coding string, which can ensure the population diversity of the genetic algorithm and realize the improvement and optimization of BP neural network.

3 Intelligent Pre-processing of College Students' Psychological Education Achievement Data

The matrix conversion is used to standardize the score data, and the score information domain is $X = \{x_1, x_2, \cdots, x_n\}$, where x_n is the object to be classified, and each object is measured by n indicators, where x_{i1} is the score nickname information, x_{i2} is the score password information; x_{in} is the score feature. The raw data matrix of the achievement information is:

$$x_{in} = \begin{Bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{in} \end{Bmatrix} \tag{5}$$

After obtaining the specific original data, it is standardized, and the expression formula is:

$$x_{ij} = \frac{x_{ij} - \overline{x}_j}{s_j} \tag{6}$$

In the formula, $\overline{x}_j$ represents the mean of the standardized score data; s_j Represents the unit dimensions after processing, and all variable units have been effectively removed. The data mean is 0, the standard deviation is 1, and the variable mean for grades is between [0, 1].

Due to certain deviations in the collection process of multi-source psychological education score data, the Grobeis criterion should be used to preprocess the merged data. The detailed process is as follows:

Assuming that $x_1, x_2, \cdots, x_n$ is subject to normal distribution, based on this, the data can be preprocessed according to Grobeis criteria. According to the processing results, the two-dimensional information entropy within the cluster is calculated. When the nodes in the cluster belong to the same attribute, the node data is fused in order, so that all the data are converged to the fusion pool [6–8].

In order to realize uneven clustering, it is necessary to introduce different competition radii at the first node of the cluster to reduce the occurrence times of the "hot zone" problem [9]. Candidate nodes use the existing data to construct a new node and use it as the initial cluster head [10, 11]. After the initial cluster first is determined, broadcast the cluster first message, the unselected cluster first will no longer be dormant, and the cluster first with the lowest communication cost is selected to complete the cluster creation. During the fusion process of intra-cluster and intercluster data, the generating tree with the smallest weight was determined to construct the multi-channel fusion path. The weight calculation formula is:

$$\omega_{ij} = \varepsilon \frac{d_{ij}}{Q_i} + \varphi \frac{d_{ij}}{Q_j} \tag{7}$$

In formula (8), ε and φ respectively represent the adjustment coefficients of two nodes i, j Q_i and Q_j Represent the remaining energy of two nodes i, j separately; d_{ij} Represents the distance between two nodes.

Under the premise of determining the multi-channel fusion path, set the size of the multi-channel spatial window and fuse the central subband data of the two nodes within the window.

In response to the changes in two spatial windows, it is necessary to convert them into fuzzy values in order to obtain local decision results in the data center. When two data centers cannot support each other, they will not be able to merge [12, 13]. In order to distinguish between trustworthy and untrustworthy data, a multi-source data regularization method was adopted to determine the relative importance of fusion order and eliminate untrustworthy data. At the multi-channel fusion level, first fuse the sub band data to be fused at the center of two nodes, and then fuse each sub band data accordingly. Due to the good storage characteristics of the fused central data, it can not only integrate the original data but also effectively improve the storage effect of the fused space.

4 Intelligent Fusion of Performance Data Based on the Improved BP Algorithm

4.1 Extracting the Features of Multi-source Heterogeneous Achievement Data

Based on the above data pre-processing results, the multi-source heterogeneous data set is updated, the sample data size is marked M, the data status is expressed in t, and the mean standardization method is used to standardize the data. The results are as follows:

$$Y_{ij}^{*}(t) = \frac{Y_{ij}(t)}{\overline{Y}_{j}} \tag{8}$$

In the equation, the standardized results of the data are marked as $Y_{ij}^{*}(t)$, the data mean is marked $\overline{Y}_{j}$, and the constant is marked i and j. Based on the above data, the standardized processing results are obtained, and the feature vectors of multi-source heterogeneous data are extracted for the first attribute. The calculation results are as follows:

$$\phi\left(Y_{ij}\right) = \sum\nolimits_{t=1}^{T} Y_{ij}^{*}(t)/T \tag{9}$$

In this equation, the feature vector of multi-source heterogeneous data is expressed in $\phi\left(Y_{ij}\right)$, and the data attribute recording period is expressed in T.

4.2 Intelligent Fusion Model of Score Data

The flowchart of intelligent fusion of college students after optimizing BP neural network is shown in Fig. 3.

As shown in Fig. 3, the steps of fusion are specific as follows: (1) Create a BP neural network, including the determination of its activation function and the number of nodes. (2) Population initialization. The goal of GA optimization is all network parameters, which must be encoded as genetic individuals, thus constructing a mapping between neural network parameters and the space of genetic dimensions.

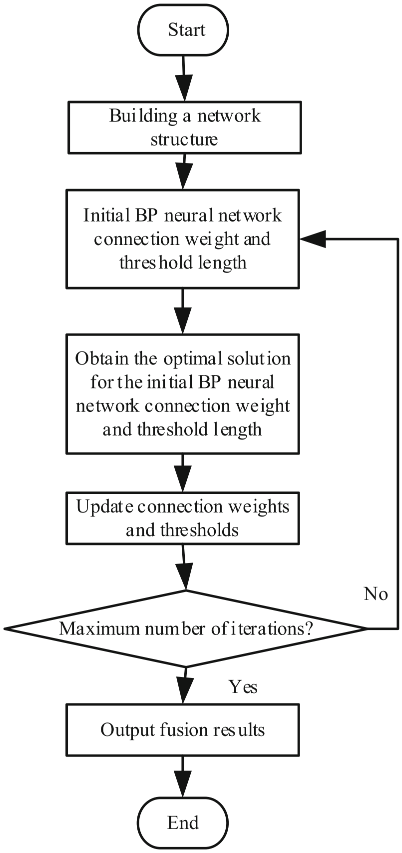

Fig. 3. Flow chart of data fusion

(3) q_i represents the characteristics of each data, and the fitness value of q_i is calculated as follows:

$$\alpha(q_i) = S_{SE} = \sum_{i=1}^{\beta} (\chi_i - \delta_i)^2 \tag{10}$$

In Eq. (9), S_{SE} refers to the sum of squared errors of the neural network model; δ_i Refers to the output value of the i network; χ_i Refers to the expected output of the i network; β refers to the actual number of neurons in the output layer.

(4) Implement genetic population optimization. For q_i, calculate the feature clustering range to represent ε_i; And calculate the number of features to represent ϕ_i.

(5) Implement offset and mutation operations.

(6) For all the population of this generation formed by heredity, calculate the fitness value of all the population, and select the best heredity as the heredity of the next

generation. For other individuals, select them through selection strategies, and combine them with the selected best heredity as the genetic population of the next generation.

(7) Judge the termination conditions: repeat steps (3) to (4). When the termination conditions are met, the optimization is directly ended to obtain the heredity of the optimal fitness.

(8) Implement optimal individual decoding and assign values to the BP neural network to achieve coarse accuracy updates of network parameters;

(9) Implement high-precision training optimization of thresholds and weights using the Levenberg Marquardt algorithm;

(10) After the maximum number of iterations or reaching the target error, end the training and complete the establishment of the optimal model; Otherwise, return to the previous step to continue training. Complete the intelligent integration of college students' psychological education score data according to the above steps.

5 Experiments and Analysis

In order to verify the performance of the intelligent fusion method for college students' psychological education score data, the fusion delay time and packet fusion exchange rate were selected as evaluation indicators to measure the ability of the fusion method. Select the performance data of students majoring in mental health education from a certain university as the test sample, with a total of 3000 data volumes. The above grade data was fused using the methods presented in this article, traditional method 1 (Method proposed in reference [2]), and traditional method 2 (Method proposed in reference [3]), and the fusion effect of the method was calculated.

5.1 Fusion Delay Comparison Test

During the data fusion process, there may be delays. When using the proposed method, traditional method 1, and traditional method 2 for data fusion, the delays generated by the above three methods during data fusion were tested, and the results are shown in Fig. 4.

During the data fusion process, there will be delays. The longer the delay time, the worse the fusion performance of the data fusion method, and vice versa.

Analyzing Fig. 3, it can be seen that as the amount of data to be fused increases, the data fusion delays detected by the three methods all show varying degrees of increase. This is because the method in this paper innovatively optimizes the weights and thresholds of BP neural network with a stable network structure, improves the convergence rate and reduces the latency. Among them, the fusion delay generated by the proposed method during data fusion is lower than the test results of traditional method 1 and traditional method 2.

5.2 Fusion Effect Comparison Testing

Based on the above test results, the packet exchange rate was selected as the data fusion performance detection indicator to test the fusion performance of the data under the

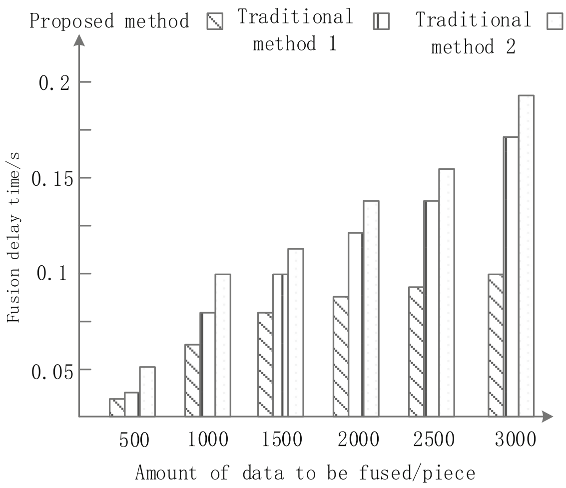

Fig. 4. Results of fusion delay tests for different methods

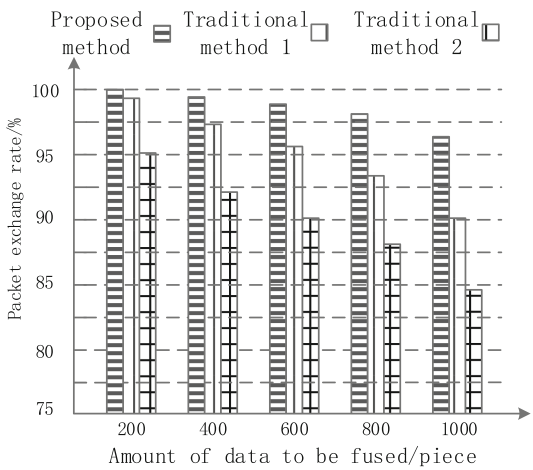

Fig. 5. Test results of packet exchange rate using different methods

proposed method, traditional method 1, and traditional method 2. The results are shown in Fig. 5.

During the data fusion process, the higher the packet exchange rate detected, the better the data fusion effect. The lower the packet exchange rate tested, the worse the data fusion effect.

Analyzing the experimental data in Fig. 5, it can be seen that as the amount of data to be fused increases, the packet exchange rates detected by the three data fusion methods all show varying degrees of decline. However, the packet exchange rate detected by the proposed method during data fusion is the highest among the three methods. From this, it can be seen that the fusion effect of the proposed method in data fusion is higher than that of traditional method 1 and traditional method 2. In summary, the proposed method has low energy consumption, low latency, and high packet exchange rate in data fusion, indicating good fusion performance in data fusion.

6 Conclusion

As the scope of computing technology increases, data fusion algorithms become particularly important. This method innovatively utilizes genetic algorithm to improve the BP neural network, extracts data features based on data preprocessing results, establishes a data fusion model, and completes data fusion through model output. The experimental results show that the data fusion effect of this method is good. However, due to limited conditions, this article did not significantly improve the accuracy of fusion. Future research will improve fusion accuracy while ensuring fusion efficiency.

References

1. Zhang, L., Zuo, X.: Credit of small and medium sized scientific and technological enterprises based on BP neural network evaluation research. MATEC Web Conf. **336**(5), 901–905 (2021)
2. Kou, L., Liu, N., Huang, H., et al.: Data aggregation privacy protection algorithm based on distributed compressive sensing and hash function. Appl. Res. Comput. **6**(1), 239–244 (2020)
3. Li, B., Li, K., Zhong, S., et al.: Research and application of deep learning algorithm based on multi-source data fusion of power grid. Electron. Des. Eng. **29**(10), 116–119, 124 (2021)
4. Stipanovic, I., Bukhsh, Z.A., Reale, C., Gavin, K.: A multiobjective decision-making model for risk-based maintenance scheduling of railway earthworks. Appl. Sci. **11**(3), 965–971 (2021)
5. Baldan, M., Nikanorov, A., Nacke, B.: A novel multi-surrogate multi- objective decision-making optimization algorithm in induction heating. COMPEL **39**(1), 144–157 (2020)
6. Kitaeva, A.V., Kolupaev, M.V., Stepanova, N.V., Zhukovskiy, O.I.: Convergence rate of mean-square error of kernel estimators of non-homogeneous Poisson process intensity function. J. Phys: Conf. Ser. **1680**(1), 012021 (2020)
7. Martín-Hernández, E., Martín, M., Mohammadi, M., Harjunkoski, I.: Modeling and analysis of organic waste management systems in centralized and decentralized supply chains using generalized disjunctive programming. Ind. Eng. Chem. Res. **60**(4), 1719–1745 (2021)
8. Lahmiri, S., Tadj, C., Gargour, C.: Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. IEEE Instrum. Meas. Mag. **24**(2), 24–29 (2021)
9. Singh, Y.P., Gautam, P.: Development of data mining algorithm for giving the loan in banks introduction. High Technol. Lett. **27**(1), 284–295 (2021)
10. Hasheminejad, S., Khorrami, M.: Clustering of bank customers based on lifetime value using data mining methods. Intell. Decis. Technol. **14**(4), 1–9 (2020)
11. Negara, Y., Doni, A.F.: Comparison of data mining algorithm performance on student savings dataset. J. Phys. Conf. Ser. **1569**(2), 022081 (2020)
12. Liu, S., Li, Y., Fu, W.: Human-centered attention-aware networks for action recognition. Int. J. Intell. Syst. **37**(12), 10968–10987 (2022)
13. Kanaan, L., Haydar, J., Samaha, M., Mokdad, A., Fahs, W.: Intelligent bus application for smart city based on LoRa technology and RBF neural network. WSEAS Trans. Syst. Control **15**(1), 725–732 (2020)