A Method of Mining Abnormal Data of College Students' Physical Fitness Test Based on Deep Learning

Liyi Xie[1(✉)] and Hui Liu[2]

[1] Shandong University of Political Science and Law, Jinan 250014, Shandong, China
xieliyi@sdupsl.edu.cn
[2] College of Information Engineering, Fuyang Normal University, Fuyang 236041, China

Abstract. In order to provide effective reference data for the improvement of college students' physique, the depth learning algorithm is used to optimize the design of abnormal data mining method for college students' physique test. Use the hardware equipment to obtain the college students' physique test data samples, according to the designed student physique test anomaly detection criteria, use the deep learning algorithm to extract the physical test data features, and determine whether the current data is the mining target. After the mining target is obtained from the data sample, the association rules of abnormal data mining are generated, and the final abnormal data mining results of college students' physique test are obtained through the steps of missing data interpolation and repeated data filtering. Through the comparison with traditional mining methods, the conclusion is drawn that the accuracy and recall of the optimized design of outlier data mining methods have been significantly improved.

Keywords: Deep Learning · Physical Fitness Test of College Students · Test Abnormal Data · Data Mining

1 Introduction

Constitution test is based on the concept of constitution, and the test content should include four major qualities, namely body shape, body function, physical quality and psychological quality [1]. At present, besides enterprises and institutions, the physical fitness test is also gradually extended to the college environment. Through the physical fitness test of college students, we can timely understand the basic physical conditions of students, and provide basic guarantee for students' learning in colleges and universities. Although the Ministry of Education has repeatedly emphasized the importance of college students' physical health, according to the data of the recent three years' research report on Chinese students' physical fitness and health, the height of college students continues to grow, but other indicators show a downward trend. In order to judge whether the physical health of college students is abnormal, it is necessary to mine the abnormal data of college students' physical fitness test. The abnormal data of college students'

The original version of the chapter has been revised. The first author's affiliation has been corrected. A correction to this chapter can be found at
https://doi.org/10.1007/978-3-031-50571-3_36

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 532, pp. 190–205, 2024.
https://doi.org/10.1007/978-3-031-50571-3_14

physique test mainly refers to the data that is not in the normal range in the students' physical test results. Through the mining and analysis of abnormal data, it provides data reference for the determination of students' health status.

Data mining is the process of extracting potentially useful and credible information and knowledge hidden in a large number of incomplete, noisy, fuzzy, random original data that people do not know in advance. At this stage, more mature data mining methods include: data mining methods based on rough set theory, data mining methods based on genetic algorithms and data mining algorithms based on association rules. Supported by the above different algorithms, they provide reference rules for data mining and obtain data mining results that meet the needs. However, the above traditional data mining methods are applied to the mining of abnormal data of college students' physical fitness test. Because the traditional methods are difficult to judge the abnormal data in the massive data, the data mining results obtained have abnormal data missing, data mining errors and other phenomena. Therefore, the depth learning algorithm is introduced. Deep learning plays a huge role in the field of data processing with the advantages of strong ability to learn data characteristics, wide data coverage, etc., and shows a trend of replacing traditional machine learning methods. The deep learning method can not only automatically optimize the loss function to learn the features of scRNA seq data, but also effectively extract the key features in the data. With the development of college students' physique testing technology, the scRNA seq data is growing exponentially. Due to its powerful computing power, deep learning just meets the needs of computing huge data sets. Deep learning algorithm includes artificial neural network algorithm, classical self coding network algorithm, convolutional neural network algorithm and other branches. In the actual application process, specific algorithms can be selected according to actual needs.

In this paper, the deep learning algorithm is applied to the data mining of college students' physique test anomalies. The abnormal data mining of college students' physique test is realized from the following aspects: setting up the abnormal detection criteria of college students' physique test, obtaining the sample of college students' physique test data, using the deep learning algorithm to extract the characteristics of the physical test data, detecting the abnormal data of college students' physique test, and generating the association rules of abnormal data mining. The experimental results show that this method can improve the mining performance of abnormal data of college students' physique test.

2 Design of Abnormal Data Mining Method for College Students' Physique Test

The process of data mining can be roughly divided into four steps: the first and most critical step is data preprocessing; The second step is to use algorithms for data mining, which mainly uses some machine learning algorithms; The third step is to evaluate the data mining algorithm model and results; Because different data sets and different mining algorithms will produce different results, the fourth step needs to display the results in an appropriate form. Figure 1 shows the basic operation process of the data mining method.

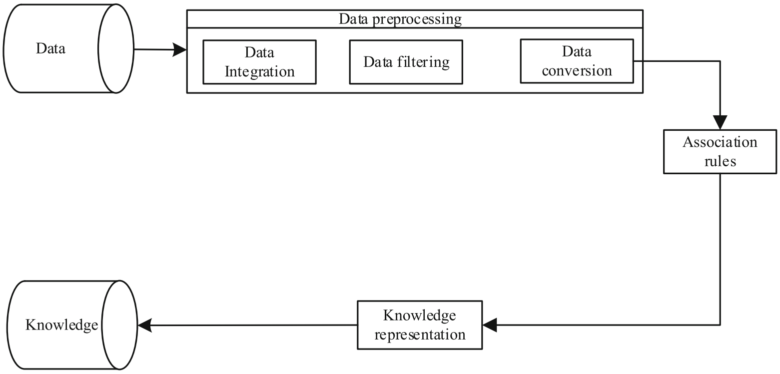


Fig. 1. Data mining flow chart

In the design process of the abnormal data mining method of college students' physique test, we need to use the depth learning algorithm to judge the abnormal conditions in the sample data, and then determine the mining target of the abnormal data, so as to obtain the mining results of the abnormal data of college students' physique test.

2.1 Setting the Detection Criteria for Abnormal Physical Fitness Test of College Students

According to the actual situation of college students and the validity, reliability and objectivity of the test indicators, the height, weight, body composition, vital capacity, cardiopulmonary function, step test, bone density, and 10 m are finally selected × 4 Round trip running, grip strength, standing with eyes closed on one foot, reaction time, sit ups, push ups, vertical jump and sitting forward bending are indicators of physical fitness test [2]. The normal range of some fitness test indexes is shown in Table 1.

Table 1. Data table of normal range of college students' physical fitness test

Physical indicators	Normal range for boys	Normal range of girls
BMI	[17.9, 23.9]	[17.2, 23.9]
Vital capacity	[3100 mL, 5040 mL]	[2000 ml, 3400 mL]
Sitting forward flexion	[3. 7 cm, 24.9 cm]	[6 cm, 25 cm]
Standing long jump	[208 cm, 273 cm]	[151 cm, 207 cm]
Pull up	[10,19]	-
50 m dash	[6.7 s, 9.1 s]	[7.5 s, 10.3 s]
800/1000 m long run	[3.17 min, 4.32 min]	[3.18 min, 4.34 min]
abdominal curl	-	[26, 56]

The test and calculation process of BMI index can be expressed as Formula 1:

$$\beta_{BMI} = \frac{m}{h^2} \tag{1}$$

In Formula 1, m and h represent weight and height respectively. The normal range of other body measurement indicators can be obtained according to the above method. In the actual data mining process, first judge whether the current data is within the normal range set in Table 1. If it is, the current data is the data mining target. Otherwise, the data mining is abandoned.

2.2 Obtaining Physical Fitness Test Data Samples of College Students

In the actual process of college students' physique test, intelligent equipment can be used to obtain the initial data samples of college students' physique test. Taking the data collection of vital capacity test indicators as an example, XGZP6847 gas pressure sensor is used to collect the exhaled gas pressure of the tested person, and some pins of the sensor are connected to the data transmission components. After each module is powered on, the transmission components will conduct ad hoc network communication according to the set components, and the data processing components will also establish communication with the PC through the wireless communication module. When the subject exhales to the sensor input through the trachea, the voltage value output by the sensor will change [3]. As the output of the sensor is analog signal, analog to digital conversion is required, while the data transmission element itself has analog to digital conversion function. Select a register to control the reference voltage and decimation rate of a single conversion channel and store the conversion result. The conversion result is in the form of binary complement, and its high two bits are zero, so it needs to be shifted. The transmission chip of the wireless sensor node is directly connected to the PC through the serial port. From the data displayed on the PC, it can be seen that the sensor has collected the output voltage change caused by the test, which is the result of the vital capacity test data samples of college students. In addition, for the data stored in the college students' physique test management platform, the heuristic crawler tool can be used for data acquisition. The implementation process of the heuristic crawler tool is shown in Fig. 2.

According to Fig. 2, data acquisition is the first step in the process. The data acquisition is completed by the web crawler, and the data set that can be mined is finally obtained. Two rounds of crawling are required. These two rounds of crawling process include all the data in the college student physique test management platform.

2.3 Extracting the Features of Body Measurement Data with Depth Learning Algorithm

Taking the acquired college students' physical fitness test data samples as the processing target, the deep learning algorithm is used to extract the features in the data, which lays the foundation for the determination of data abnormal state. The data mining algorithm for optimal design uses artificial neural networks, that is, mathematical models simulating biological neural networks. Biological neural networks refer to networks composed

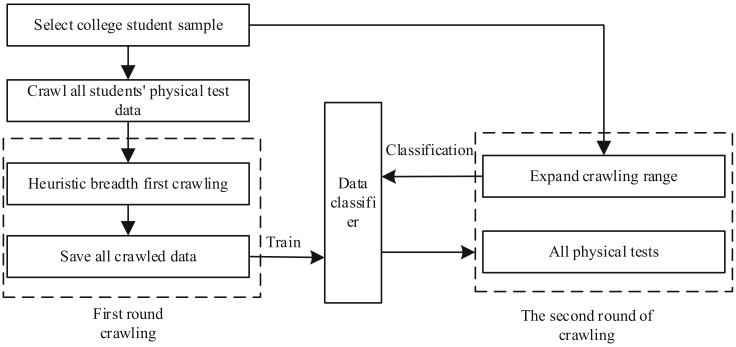


Fig. 2. Flow chart for obtaining physical test data samples of college students

of biological brain neurons, cells, contacts, etc., which are used to generate biological consciousness and help biological thinking and action. Its important component is neurons [4]. Artificial neural network is to use computers and other equipment to simulate the basic functions of biological neurons and learn the ability of human independent thinking. In order to simulate the nonlinear operation of neurons, artificial neurons add some nonlinear activation functions to the weighted sum, so that the depth neural network has nonlinear learning ability. Figure 3 shows the structure diagram of the mathematical model of a single artificial neuron.

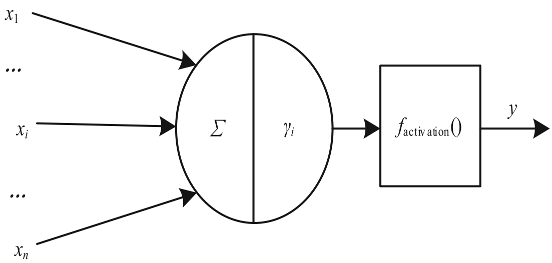


Fig. 3. Structure of neuron model

In Fig. 3, γ_i is the action threshold of the neuron, $f_{\text{activation}}()$ is the activation function of the function, and y is the actual output value of the neuron node. Abstract and summarize with mathematical expressions. There are two states of neurons, active state or inhibitory state, which are represented by ± 1 [5]. The output of any neuron i in the artificial neural network can be expressed as Formula 2:

$$y_i = \operatorname{sgn}\left(\sum_{i=1}^{n_{net}} \varpi_{ji} x_j - \gamma_i\right) \tag{2}$$

In Formula 2, ϖ_{ji} is the weight value between the nodes in the input layer and the hidden layer. The size, positive and negative of the weight value are used to represent the two states of synaptic strength and excitation and inhibition. n_{net} is the number of neurons in the constructed artificial neural network. sgn() is called a step function or symbolic function, and is expressed as Formula 3:

$$\operatorname{sgn}(x) = \begin{cases} -1, x < 0 \\ 1, x \geq 0 \end{cases} \tag{3}$$

In Formula 3, x is the input value of the artificial neural network. The hyperbolic tangent function is used as the activation function of neurons, which is expressed as Formula 4:

$$f_{\text{activation}}(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

In the process of feature extraction of the body measurement data, the forward propagation and backward propagation of the depth learning algorithm are used. The specific propagation principle is shown in Fig. 4.

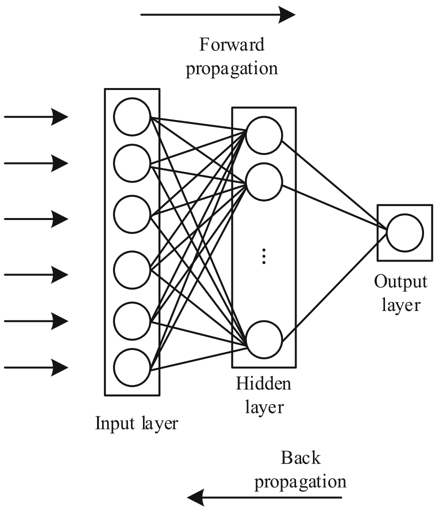


Fig. 4. Schematic diagram of forward and backward propagation of deep learning algorithm

The weights on the connection between the neural network layers form two weight matrices W_1 and W_2. I, G and O are defined as input vectors, hidden layer vectors and output vectors respectively. Then the process of forward propagation can be expressed as Formula 5:

$$\begin{cases} \mathrm{G} = f_{\text{activation}}(W_1 \cdot \mathrm{I}) \\ \mathrm{O} = f_{\text{activation}}(W_2 \cdot \mathrm{G}) \end{cases} \tag{5}$$

Backpropagation aims to update the weight of the connection by returning the network error. Assuming that each training sample is (α, χ), the forward propagation output is u, and each node error is ε, the node error of the output layer is expressed as Formula 6:

$$\varepsilon_{oi} = u_i(1 - u_i)(\chi_i - u_i) \tag{6}$$

The node error of hidden layer is expressed as Formula 7:

$$\varepsilon_{gi} = g_i(1 - g_i)\sum_{i=1}^{n_{net}} \varpi_{ki} \cdot \varepsilon_{oi} \tag{7}$$

No matter how many hidden layers there are, the calculation methods of forward propagation and back propagation are based on the above formula. The update of each weight value is represented by Formula 8:

$$\varpi_{ij} \leftarrow \varpi_{ij} + \gamma \varepsilon_j x_{ji} \tag{8}$$

The emergence of back propagation algorithm solves the problem that the original multi-layer perceptron error cannot be returned, making the deep neural network possible [6]. In the process of actual data feature extraction, set the average value and maximum value of data as feature vectors, input the acquired initial college students' physique test data into the constructed artificial neural network, and the result of feature vector extraction is Formula 9:

$$\begin{cases} \delta_{avg} = \dfrac{\sum\limits_{i=1}^{n_{net}} x_i}{N_c} \\ \delta_{\max} = \max(x) \end{cases} \tag{9}$$

In Formula 9, N_c is the sampling value of college students' physique test data, max() is the maximum solution function, and x_{avg} and $x_{\max}$ are the extraction results of data average and maximum eigenvectors. Finally, formula 9 is used to fuse the extracted feature vector, and the fusion result is expressed as formula 10:

$$\delta_{com} = \sum_{i=1}^{n_t} \delta_i \tag{10}$$

In Formula 10, δ_i is the extracted feature vector, and n_t is the extracted number of feature vectors from abnormal data of college students' physique test. According to the above process, the feature extraction of abnormal data of college students' physical fitness test is completed.

2.4 Detect Abnormal Data of College Students' Physique Test

According to the feature extraction results of college students' physical fitness test data output by the artificial neural network algorithm, we can judge whether the current data is abnormal through similarity measurement. The measurement process of similarity can be expressed as Formula 11:

$$\varphi = \frac{|\delta_{com} \cap \delta_{set}|}{\sqrt{|\delta_{com}||\delta_{set}|}} \tag{11}$$

In Formula 11, δ_{com} and δ_{set} are the extracted comprehensive data features and the set normal range data features respectively. If the calculation result of Formula 11 is higher than the threshold value φ_0, it means that the current college students' physique test data is within the normal range, that is, the current college students' physique test data is normal; otherwise, the current data is considered as abnormal data, which is the target of abnormal data mining.

2.5 Generating Association Rules for Abnormal Data Mining

The main function of association analysis is to discover an association rule, through which the hidden association relationship between different things can be found. Association rules can be expressed as implication expressions in the form of X → Y. Association rules can reflect the interaction between itemset X and itemset Y. Through association analysis, you can find interesting associations or correlations between items in large transactions or relational data sets, and know which transactions frequently occur at the same time, thus helping people make better decisions. Association rules mean that there may be a strong relationship between two items [7]. Association rule mining is a rule based machine learning algorithm, which can find interesting relationships in large databases. Its purpose is to use some indicators to distinguish the strong rules in the database. The commonly used measurement methods of association rules are confidence between association rules and support of association rules. The final strong association rule is based on the minimum support and minimum confidence set by the experiment. That is, when the support of an association rule is greater than the currently set minimum support and its confidence is greater than the currently set minimum confidence, it can be said that this association rule is a strong association rule. Support is the probability of occurrence, that is, the proportion of occurrences of the current item in all item sets. The support is calculated according to Formula 12:

$$\mu_{SUPPORT}(X \to Y) = P(X \cup Y) \tag{12}$$

In Formula 12, $P(X \cup Y)$ represents the probability of simultaneous occurrence of itemset X and itemset Y. Confidence is a conditional probability, that is, the probability of condition Y when condition X occurs. The calculation of confidence is shown in Formula 13:

$$\mu_{Confidence}(X \to Y) = P(Y|X) \tag{13}$$

In Formula 13, $P(Y|X)$ represents the probability that itemset X is included in itemset Y. In the process of generating association rules for actual mining, frequent itemsets are generated first, and support is set to get frequent patterns. That is, the number of times a pattern appears in the established database divided by the number of records of this type of data in the database must not be less than the set support. Otherwise, it is considered as a non frequent pattern, and it is considered that the occurrence of this pattern is accidental and of no value [8]. On the setting of support, because the physique test data is obviously highly consistent with the normal distribution, if the support is set too high, there will be few rules that can be discovered. Therefore, this paper attempts to have multiple support degrees, and then screen the rules found. The confidence level is set to 0.8, that is, if A frequent mode is A, then A is not less than B's probability of occurrence in all frequent modes A is greater than 0.8, then the generated strong rule is considered reliable.

2.6 Realization of Abnormal Data Mining in College Students' Physique Test

In order to achieve accurate mining of abnormal data of college students' physical fitness test, a data mining framework is constructed. The abnormal data mining model is divided into four layers, the first layer is the site layer, where each site corresponds to a site agent, and each site agent corresponds to a site information database; The second layer is the management layer, in which the management agent manages all the agent information and coordinates the work between agents in each layer; The third layer is the cooperative mining layer, which is composed of data preprocessing agent, data mining agent and evaluation agent; The fourth layer is the algorithm layer. The data preprocessing agent can call m preprocessing algorithms, and the data mining agent can call n mining algorithms. According to the characteristics of the application, the communication between the site layer and the management layer adopts the point-to-point message transmission mode of establishing a channel between agents. Based on this communication channel, two-way and peer-to-peer message transmission is carried out. The communication between the management layer and the cooperative mining layer adopts blackboard mode, that is, a group of agents participating in communication share a common area, and information exchange is realized by writing and reading information to the area. Agents can understand and exchange information with each other and process messages through message protocols. In the implementation of the system, the message protocol uses a data mining oriented communication language based on the KQML language definition.

2.6.1 Mining Abnormal Data of College Students' Physical Measurement

With the support of generating association rules for abnormal data mining, the Apriori algorithm is used to mine the abnormal data of college students' physical fitness test. Apriori algorithm is an important algorithm to calculate the frequency of occurrence of both indicators to determine whether there is correlation. In Apriori algorithm, support is obtained based on the frequency of occurrence of indicators in big data. However, in order to solve the important problem of data redundancy, the sort method is often used to disrupt the data set, and the corresponding logic code is used to obtain the training set and the optimal set. The support and confidence are calculated by analyzing the data to further explain the corresponding relationship between some indicators in the data set. In the actual mining process, traverse each indicator in the data set, establish a 1-item set, output all itemids to record the list of each indicator, and judge whether the indicator is in the list by traversing each record and the indicator in each record. If the indicator is not in the list, it will be added to the list. After this process, all indicators will be sorted, the list elements will be mapped to frozenset(), and the list will be returned. The candidate set is further obtained by inputting dataset D, candidate set Ck, and minimum support. The main steps to obtain the candidate set are described as follows: obtain the candidate set Ck by combining the upper frequent item set Lk-1, and filter the candidate set Ck using the minimum support minSupport. After the above process, output the frequent item set Lk of this layer and the support of each item, establish a dictionary to calculate the number of occurrences of each item in the candidate set Ck and in all indicator records, and calculate the number of occurrences by comparing each item in the candidate set with the original indicator records. And traverse each indicator record to traverse each item in the candidate set Ck for comparison. If the items in candidate set Ck appear in the indicator record, the current item is a subset of the current indicator record [9]. If the item in Ck is selected for the first calculation, the number is counted as 1; Otherwise, the total number of records in the dataset and the total number of indicator records are used to calculate the support and record the frequent item set filtered by the minimum support. Generate candidate k-itemsets from upper frequent k-1 itemsets, obtain the number of new candidate set elements k, and output candidate sets. Save the new candidate set and enter the number of frequent itemset records to cycle through. Each project in the more frequent project set is the same as the other projects. Each project is compared to other project elements by using two for loops. Traverse other items in the candidate set except the previous item. Compared with the current item, k-1 elements of the current item in the candidate set. There are k-1 elements of the remaining items in the candidate set. Each time, there is only one item in the remaining items. Convert the tuple into a list for sorting. Merge to generate a $k + 1$ itemset, return the last $k + 1$ itemset, which is the same, and then merge the two items. Find L2 and L3 according to L1 through the while loop, which will create a large list containing a larger itemset until the next larger itemset is empty. The indicator combination length of the candidate set exceeds the maximum indicator record length of the original data set. If the maximum indicator record length of the original data set is 4, the candidate set is at most 4-item set, and k-item candidate sets are generated from frequent $k-1$ item sets. Frequent k-item sets are generated from k-item candidate sets and filtered through minimum support. Update the support dictionary to add new support, add the new frequent k itemset to the list of existing frequent itemsets, and add k plus 1 to generate the next itemset. Scan the entire data set to obtain all the data of one setting that frequently appears as a candidate. $K = 1$, frequent 0 itemsets are empty sets. Enter the apriori function to generate frequent union list L, which supports lists and minimum confidence. The output contains a list of trust rules used to generate association rules. The last item of the rule list that meets the minimum confidence requirement traverses all the items in H. It is used as the confidence calculation of the last item of the association rule, and the set subtraction operation is used. If the confidence level is greater than the minimum confidence level minConf, output the association rule preamble freqSet condeq, and output the association rule condeq. Save the association rules that meet the conditions, save the previous item of the association rules, return the latter part and the rule items whose confidence level meets the minimum confidence requirement, and return the last item that meets the conditions, so as to realize abnormal data mining in all college students' physical test data samples.

2.6.2 Interpolate Missing Data

There may be some missing data in the initially mined abnormal physical measurement data of college students, so the missing data needs to be interpolated. The specific interpolation process can be expressed as Formula 14:

$$x_{\text{defect}} = \frac{x_q - x_h}{2} \tag{14}$$

In Formula 14, x_q and x_h represent the previous data and the next data of missing data respectively. According to the above method, the method of median solution is used to obtain the interpolation results of missing data from college students' physical measurement anomaly mining.

2.6.3 Duplicate Data Identification and Filtering

Use Formula 15 to identify whether there is duplicate data in the college students' body measurement abnormal data mining results.

$$s = \sqrt{\left(x_i - x_j\right)^2} \tag{15}$$

In Formula 15, x_i and x_j represent any two data in the abnormal data mining results respectively, and the calculation result of s represents the coincidence coefficient between the abnormal data mining results [10]. If the value of s is 1, x_i and x_j are considered as duplicate data, and one of the abnormal data needs to be deleted; otherwise, it is considered that there is no duplicate data between them, and no filtering is required.

Finally, the processed abnormal data of college students' physical tests are clustered to obtain the data mining results of college students' physical tests that meet the quality requirements, and finally the data mining results are output in a visual form.

3 Performance Test Experiment Analysis

This data mining performance test experiment is conducted on Intel (R) Core (T1Vl) i3 2.4 GHz and 2G memory Windows? Microsoft Visual C++ 6.0 is used as the data mining platform of this experiment, and MySQL is used as the database management system.

3.1 Prepare Data Samples for Physical Fitness Test of College Students

The experimental data used in this experiment was provided by the physical education institute of all colleges and universities in a city. The initial data included the physical fitness test items of college students in all colleges and universities in the city from 2020 to 2022, specifically including height, weight, vital capacity, step test, grip strength, sitting precursor and standing long jump. Then, the database technology was used to generate the data mining database of this paper. After removing some useless records, the amount of experimental data was 8900. Before the experiment, considering the inconsistency and incompleteness of the experimental data, it is necessary to process the data to make it meet the requirements of mining. The data samples of college students' physical fitness test prepared include normal physical fitness test data and abnormal physical fitness test data, and the data samples are randomly divided into multiple groups. The specific data sample division and settings are shown in Table 2.

Table 2. Sample setting of college students' constitution test data

Experimental group	Normal data volume/MB	Abnormal data volume/MB
1	357.8	307.2
2	465.2	355.9
3	383.9	321.8
4	374.6	289.7
5	452.1	316.4
6	479.6	353.8
7	398.2	279.8
8	407.5	323.7

The setting of abnormal data volume in Table 2 is the mining target value of abnormal data of college students' physical fitness test.

3.2 Input Operation Parameters of Deep Learning Algorithm

Because the optimized design of the abnormal data mining method of college students' physique test uses the deep learning algorithm, it is necessary to set the operation parameters of the algorithm. The network of the artificial neural network algorithm used has six layers. The input layer is the sample of the physique test data prepared. The number of settings of the input layer, output layer and hidden layer is 10, and the network size of the input and output layers is 27×23. The hidden layer network size is 54×46 in 1 step $\times$ 1. The number of neurons in the artificial neural network is 180, and the number of nodes in the output layer is 60. The learning rate of the artificial neural network in the propagation learning process is 0.0001, and the maximum number of iterations is 2000.

3.3 Setting Performance Evaluation Indicators for Abnormal Data Mining

In order to quantitatively evaluate the performance of the method, the accuracy and recall of outlier data mining are set as the quantitative test indicators of mining performance. The numerical results of the accuracy index are shown in Formula 16:

$$\vartheta_{Acc} = \frac{Num_{\text{abnormal}}}{Num_{\text{excavate}}} \times 100\% \tag{16}$$

In Formula 16, Num_{excavate} is the output data volume of abnormal data mining results, and Num_{abnormal} is the volume of body side abnormal data in the output results of mining methods. In addition, the test result of the recall evaluation index can be expressed as Formula 17:

$$\vartheta_{\text{Rec}} = \frac{Num_{\text{abnormal}}}{Num_{\text{set}}} \times 100\% \tag{17}$$

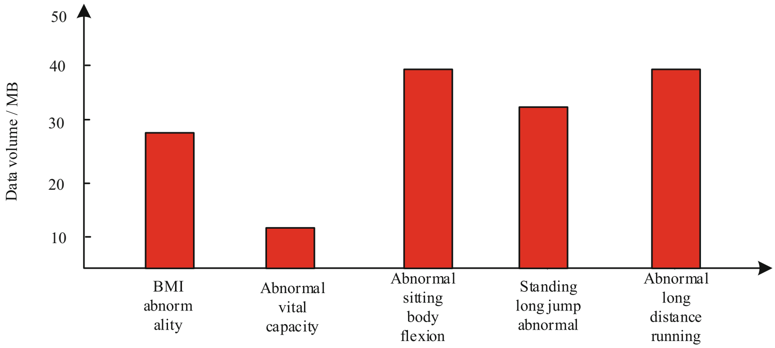
In Formula 17, Num_{set} is the total number of abnormal data samples of college students' physical fitness test. The final test shows that the higher the mining accuracy and recall, the better the mining performance of the corresponding method.

3.4 Performance Test Process and Result Analysis

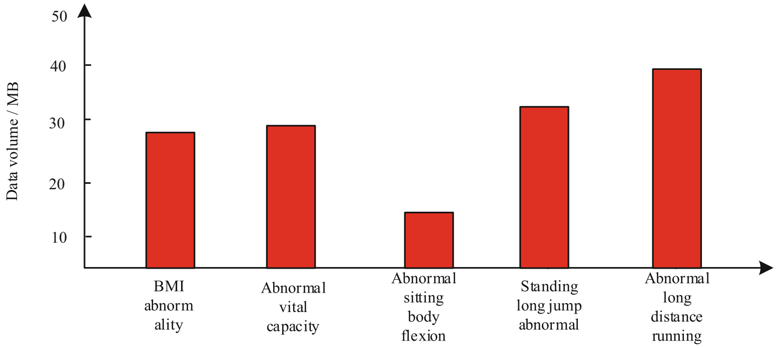
In order to reflect the advantages of optimizing the mining performance of the abnormal data mining method of college students' physical fitness test based on deep learning, the experiment sets the traditional data mining method based on rough set theory and the data mining method based on genetic algorithm as the comparison method of the experiment. Input the prepared college students' physique test data samples into the corresponding mining method to obtain the corresponding mining output results, as shown in Fig. 5.

Based on the statistics of relevant data, through the calculation of Formula 16 and Formula 17, the test comparison results reflecting the mining performance of abnormal data mining methods for college students' physique test are obtained, as shown in Fig. 6.

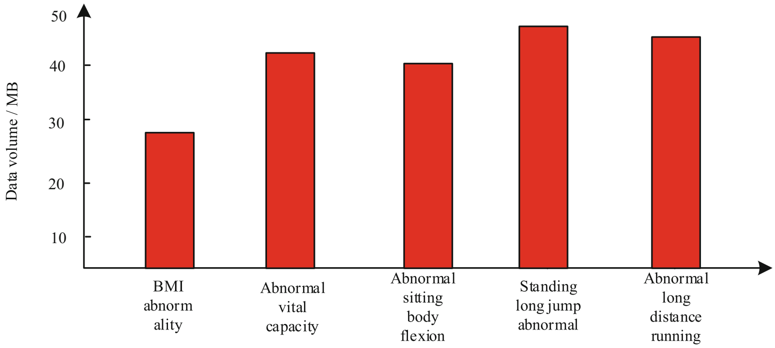
It can be seen intuitively from Fig. 6 that, compared with traditional data mining methods, the accuracy and recall rate of the optimized design of college students' physical fitness test exception data mining methods based on deep learning have been significantly improved, that is, the optimized design mining methods have obvious advantages in performance.



(a) Data mining method based on rough set theory



(b) Data Mining Method Based on Genetic Algorithm



(c) Optimizing the design of data mining methods for students' body measurement anomalies

Fig. 5. Abnormal data mining results of college students' constitution test

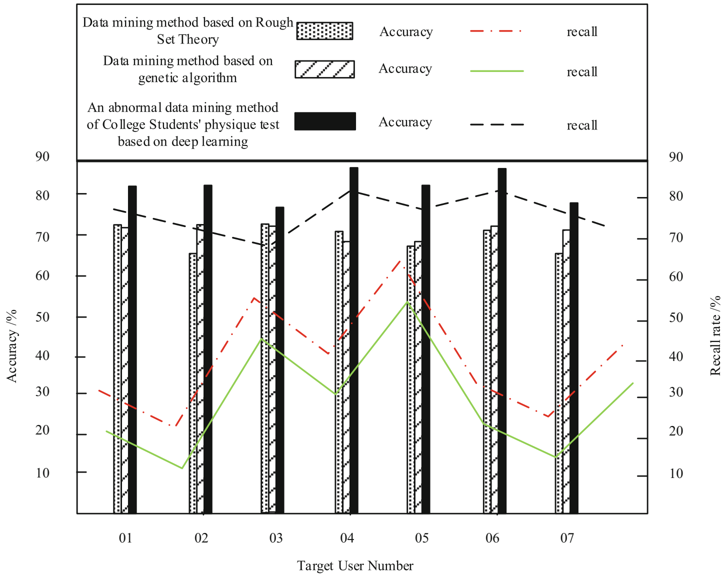


Fig. 6. Comparison results of body testing abnormal data mining performance test

4 Conclusion

In order to accurately analyze the situation of college students' physique test, this paper proposes a data mining method for college students' physique test anomalies based on deep learning. The in-depth learning method is used to collect and mine the data of college students' physique test, so as to accurately find out the students with abnormalities in the process of physique test. The experimental results show that the proposed method provides powerful help for the research of college students' physical health. Through the application of the deep learning algorithm, the optimization design of the data mining method for college students' abnormal physique test is realized, so as to accurately grasp the change trend and take effective measures in time, which is conducive to continuous improvement and improvement of college students' physique. The next research will break through the technical barriers, realize the cross-integration of data mining technology and professional sports data analysis system, broaden its depth and breadth in sports practice, and truly realize the transformation from data to value.

References

1. Lopes-Silva, J.P., Panissa, V.L.G., Julio, U.F., Franchini, E., et al.: Influence of physical fitness on special judo fitness test performance: a multiple linear regression analysis. J. Strength Condition. Res. **35**(6), 1732–1738 (2021)
2. Wang, N., Liu, Y.: A novel most valuable player algorithm for optimization problems. Comput. Simul. **37**(6), 273–282,327 (2020)
3. Lin, S.: Data mining artificial intelligence technology for college English test framework and performance analysis system. J. Intell. Fuzzy Syst. **40**(2), 1–11 (2020)

4. Tomasevic, N., Gvozdenovic, N., Vranes, S.: An overview and comparison of supervised data mining techniques for student exam performance prediction. Comput. Educ. **143**, 103676.1–103676.18 (2020)
5. Yin, X.H.: Construction of student information management system based on data mining and clustering algorithm. Complexity **2021**(2), 1–11 (2021)
6. Liu, S., Li, Y., Fu, W.: Human-centered attention-aware networks for action recognition. Int. J. Intell. Syst. **37**(12), 10968–10987 (2022)
7. Wan, H., Tang, S.: Sentiment analysis of students in ideological and political teaching based on artificial intelligence and data mining. J. Intell. Fuzzy Syst. **3**, 1–10 (2021)
8. Li, J., Lei, H., Tsai, S.B.: Online data migration model and ID3 algorithm in sports competition action data mining application. Wirel. Commun. Mob. Comput. **2021**(7), 1–11 (2021)
9. Yin, Z., Cui, W.: Outlier data mining model for sports data analysis. J. Intell. Fuzzy Syst. **40**(2), 1–10 (2020)
10. Liu, S., Xiyu, Xu., Zhang, Y., Muhammad, K., Weina, Fu.: A reliable sample selection strategy for weakly supervised visual tracking. IEEE Trans. Reliab. **72**(1), 15–26 (2023). https://doi.org/10.1109/TR.2022.3162346