A Method of Abnormal Psychological Recognition for Students in Mobile Physical Education Based on Data Mining

Changyuan Chen(✉) and Kun You

Department of Physical Education, Xi'an Shiyou University, Xi'an 710065, China
chenchangyuan895@163.com

Abstract. At present, the abnormal psychological recognition of middle school students is mainly through psychological questionnaire, after data statistical processing, to assess whether there is abnormal psychological students. The recognition accuracy is strongly dependent on the reliability of the questionnaire, which leads to the poor recognition accuracy and stability. In order to solve these problems, the method of abnormal psychological recognition of students in mobile PE teaching based on data mining will be studied. After analyzing the influence of PE teaching on students' psychology, the behavioral characteristics that represent students' psychology are extracted. After constructing the students' psychological view, the students are classified preliminarily. Through constructing mental state mining decision tree, using iForest algorithm to realize abnormal mental recognition for middle school students. The test results of recognition method show that the accuracy of the mental recognition method is stable between 87.28% and 87.95%, and the recognition reliability is higher.

Keywords: Data mining · Physical education curriculum · Mobile teaching · Middle school students' psychology · Anomaly recognition · Decision tree

1 Introduction

All kinds of research and survey data show that the mental health problems of middle school students are becoming more and more serious at the present stage. Among them, sensitivity of interpersonal relationship, abnormal behavior and mental endurance are more common, and some other phenomena are depression, anxiety, weariness of study, suicide and so on. The rapid development of modern society, the intensification of competition, the acceleration of life rhythm, the increasing complexity of interpersonal relationships, the pluralism of values, make people face great pressure, people's mental diseases are also generally increasing, all kinds of mental diseases seriously trouble people, especially adolescents. Adolescence is the key period of students' physical and psychological development, as well as the high incidence of psychological conflicts and emotional and behavioral problems. In the field of psychology, the psychological characteristics of an individual are mainly described by the "state" psychological variables

W. Fu and L. Yun (Eds.): ADHIP 2022, LNICST 469, pp. 286–296, 2023.
https://doi.org/10.1007/978-3-031-28867-8_21

represented by mental health state and the "trait" psychological variables represented by personality. In the past, most of the methods to judge mental health status were self-assessment or other-assessment, but there were obvious deviations in this way, and the reliability of the evaluation results was poor [2].

Mobile teaching of physical education courses is another new attempt of physical education, which can enrich the physical education teaching methods and alleviate the dilemma of traditional physical education mode limited by venues. However, when the PE course adopts the mobile teaching method, it needs a lot of information technology, which can not only help to complete the basic teaching content, but also collect the students' behavior and feedback data. These data generated in the process of mobile teaching can not only be used to evaluate students' learning effect, but also be used to analyze students' psychological state. In related research, some scholars proposed a mental workload estimation method using deep BLSTM-LSTM network and evolutionary algorithm [3]. This approach proposes a deep hybrid model based on bidirectional long short-term memory (BLSTM) and long short-term memory (LSTM) for workload-level classification, which can effectively monitor the mental state of mental activity. Other scholars have proposed a mental health sentiment analysis method based on deep convolutional networks [4]. This method is able to process facial images and account for the temporal evolution of emotions through a novel solution, and utilize standard linear discriminant analysis classifiers to obtain final classification results. The method can aid in the detection, monitoring and diagnosis of human mental health. However, when the above two methods are applied to the monitoring of students' mental health in mobile learning, there are still time and space limitations or untimely detection.

Therefore, how to effectively use a large number of data, through data mining technology to analyze the amount of information mining data, so that simple data can express the deep meaning of the data, to enhance the deep use of mobile teaching data, to strengthen the psychological state of students concern and guidance, is one of the important issues facing. Most of the traditional statistical analysis put forward deterministic analysis, which may not fully explain the extracted data, while data mining focuses on exploring the value information in the collected data. Using data mining technology can analyze the physiological data of students in the mobile teaching of physical education courses, thus improving the accuracy of abnormal psychological recognition. Based on the above analysis, this paper focuses on the mental health of students, and studies the method of abnormal mental recognition in PE mobile teaching based on data mining.

2 Abnormal Psychological Recognition Method for Students in Mobile Physical Education Based on Data Mining

2.1 Analysis of the Influence of Mobile Physical Education on the Psychological Change of Middle School Students

Antonouski presents the mental health model shown in Fig. 1. Through this model, the physical and mental health can be explained, and the necessary prerequisites for maintaining health and the necessary procedures for maintaining health are analyzed [5].

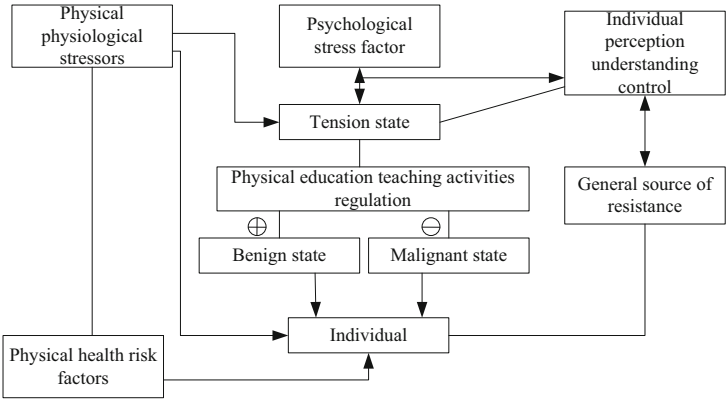

Fig. 1. Antonouski mental health model

From the health model, it can be determined: First of all, targeted physical activity directly contributes to the establishment of protective factors and resistance sources, thus resisting various pathogenic factors. Secondly, in addition to the related factors, physical activity has a direct impact on happiness and tension. The teaching process of physical education has the characteristics of intuition, which makes students have to use all kinds of senses synthetically: perceiving the action image through sight and hearing; perceiving the action essentials through touch and muscle itself, the degree of muscle exertion and the space-time relationship of the action, so as to establish the complete and correct action representation. In this process, the students' ability of perception, observation, image memory and action memory have been developed and improved.

Compared with the traditional physical education, the content of mobile education in physical education is more abundant, which can make students experience satisfaction, joy, tension, excitement, anxiety and so on. The team spirit of PE teaching and the mutual help and study among students can inspire students' social consciousness and enhance their self-esteem, confidence and sense of responsibility. The competitive nature of PE teaching can inspire students' enterprising spirit, inspire their will, and make various emotional experiences more profound. In this relaxing and autonomous activity, students are willing to participate in physical exercise actively and persist for a long time, so that the development of physical and mental health is more coordinated and lasting, and the control and regulation of anxiety, depression and other negative emotions are more significant [6].

Middle school is the key period to develop students' will quality. Mobile P. E. teaching can develop students' will quality such as consciousness, decision and self-control. In the process of PE teaching, students need to participate in various group activities, which provides conditions for cultivating students' social adaptability and promoting the diversified development of personality. Emotional infection, happy teaching, self-confidence training, willpower training, social contact training and psychological inducement are added to PE teaching to improve students' social adaptability, willpower, self-confidence, cooperation, physical and mental adjustment and self-evaluation.

Mobile PE teaching has a positive effect on students' mental health, but once the students' mental health is abnormal, the learning feedback collected by mobile PE teaching platform will appear abnormal. From the above analysis, the data that can reflect students' psychological state in the process of mobile physical education course teaching are students' learning enthusiasm, students' performance in team activities, students' emotion in physical training, students' will quality and students' specific learning situation. The premise of the application of data mining is to create a data warehouse with mass data and all the information. This paper uses the data stored in mobile teaching platform for data mining analysis. After determining the influence of mobile physical education on students' psychology and the data object of data mining analysis, the corresponding behavior data are collected and the psychological view of students is established.

2.2 Student Behavior Data Processing and Psychological View Construction

There are a lot of student data stored in PE mobile teaching platform, but the format of these data is not uniform, and there are dirty data and redundant data in this paper.

Because the data in this article comes from different terminals, data fields are different, and the format of data storage is different. So in this paper, heterogeneous data is stored in a database through data integration technology. A number of tables are associated with the desensitized student number as the main key, and gender, nationality and other characteristics are replaced by one-hot coding. In order to improve the efficiency of mining, it is necessary to simplify the dataset, that is, to delete the unused data according to the task of mining, and to decide which dataset to mine. The data selection work undertaken in this paper includes [7]:

1. School numbers, names and dates of birth are deleted directly, as these attributes are meaningless for mining.
2. Direct deletion of ethnicity, which has little impact on mining results.

The missing data in this paper is generally handled according to different situations. In the case of large enough data, some of the student's data is missing, this paper chooses to delete this data or delete the current student's information. If a student is missing a large number of course grades, delete the student from the student list in the grade module. In the part of noise anomaly, there are two kinds of anomaly data, one is the abnormal data beyond the fact, the other is the real anomaly data. Therefore, some of the data through visualization, according to the specific situation of noise data selection.

When students have psychological problems, their academic performance will be greatly changed, and generally with the straight line decline of academic performance, so the student's academic performance may be an indicator variable of psychological problems. In different mobile sports learning scene or student learning period, the student's psychological view may have mutual influence, and each individual's inner thoughts and habits will affect his performance in different views. Therefore, this article uses the MANE algorithm to fuse the different view performance, attempts to reconstruct each student this intrinsic thought, the custom. MANE algorithm can be formalized as: Given a network G, the multi-view network embedding goal is for each node $i \in U$, learning a low-dimensional vector representation of $d_i \in R^D$. To preserve the diversity of views

as much as possible, for each node $i \in U$ and each view $p \in P$, multi-layer network embedding learns an intermediate representation of $d_i^{(p)} \in R^{D/P}$, which holds only the information within the view. After a series of intermediate processes (the intermediate representation of collaborative updates between views), this intermediate representation also preserves the potential information between views, and the final representation of a node is the splicing of the vectors of that node within each view, namely:

$$d_i = \oplus p \in d_i^{(p)} \tag{1}$$

Students' psychological view used in this paper is directed network, that is, the connection between nodes has directivity, node A connection B and node B connection A have different meaning, so in the network there are degrees and degrees. But each connection in the view p will have an Entry $I_d^{(p)}$ and an Exit $O_d^{(p)}$, so the average of the views is:

$$\overline{k} = \frac{N\left[E^{(p)}\right]}{N(U)} \tag{2}$$

The E represents the set of edges in a view and the U is the set of nodes. Different students have different psychological view, but the psychological view of abnormal students and other students psychological view of the similarity between the greater differences. Through calculating the similarity coefficients between student's psychological views, the students' psychological views with low similarity are separated. Jaccard coefficients are used to measure the similarity between different views. Because of the particularity of mobile PE teaching data, students have different behaviors. After setting up the psychological view of students in the process of learning PE, the decision tree algorithm is used to construct students' mental state decision tree.

2.3 Constructing Mental State Mining Decision Tree for Middle School Students

In the actual educational environment, the majority of students are self-care, and the abnormal students are only a small part. Therefore, there is interference in mining abnormal data. However, for schools and psychological institutions, psychological abnormality students should be more concerned about the object.

Decision tree is an inverted tree structure that can classify data automatically. Every node of decision tree is tested on its attribute by top-down recursion, and two or more branches are generated from different answers to questions on each node, which leads to different results. More common decision tree algorithms are the classic ID3 algorithm, C4.5 algorithm and CART algorithm and so on. This paper adopts C4.5 algorithm to construct decision tree of middle school students' psychological state.

1. The data with the lowest similarity coefficient shall be taken as the data to be divided according to the results of similarity calculation of students' psychological views. Calculate the information gain rate for each split attribute in the set as follows;

Set node N to store data dividing all samples of P. The expected information required for classification of samples in the P is given by the following formula:

$$I(P) = -\sum_{i=1}^{n} v_i \lg v_i \tag{3}$$

v_i is the probability that any sample in P belongs to C_i, and n is the sample serial number. $I(P)$ is also called entropy of P. Assume that the samples in the P are divided by attribute K, which has m values $\{k_1, k_2, \cdots, k_m\}$. If the value of the attribute K is discrete, then the attribute K can divide the P into a subset of m $\{P_1, P_2, \cdots, P_m\}$, where the value of the sample in the P_j is k_j on the attribute K. These subsets correspond to branches that grow from node N. The expected information required for classification of samples of P by attribute K can be derived from the following formula:

$$I(P)_K = \frac{|P_j|}{|P|} \times I\sqrt{|P_j|} \tag{4}$$

where $\frac{|P_j|}{|P|}$ is the weight of a subset of values of k_j on the attribute K. $I(P)_K$ is based on the desired information needed to classify the samples of P by attribute K.

The rationale behind C4.5 and ID3 is the same, except that C4.5 replaces the information gain as the attribute selection measure (splitting rule) in order to compensate for ID3's tendency to use information gain to select attributes with higher values. The information gain rate is defined as follows:

$$GR(K) = \frac{G(K)}{S(K)} \tag{5}$$

The entropy reduction is $G(K)$; split information is used in the above expression to normalize the information gain.

$$S(K) = -\frac{|P_j|}{|P|} \times \log_2\left(\frac{|P_j|}{|P|}\right) \tag{6}$$

2. By comparing the information gain rate of each split attribute, it is determined that the split attribute with the maximum information gain rate is the root node of the decision tree, and that the attribute has several values, and the data set is split into several subsets, and if there is only one value of the attribute, the split is ended;
3. Recursively perform steps 1 and 2 on each subset of the split data set.

Repeat the above steps to classify the sub datasets of each of the outgoing branches to lead to branching again. With the increase and extension of tree branches, the data set is divided into smaller subsets recursively, and finally, the decision tree of students' mental state is generated.

2.4 Realization of Abnormal Psychological Recognition for Middle School Students

According to the classification attributes, a tree-like structure is established, and a path from the root node to the leaf node forms a classification rule. Accordingly, the whole decision tree forms a set of disjunctive rules, which can be easily converted into IF-THEN classification rules, according to which it is easier to classify, identify and predict new data. After the decision tree of students' psychological state is formed, the trees are pruned. This paper chooses PEP algorithm to prune the decision tree of students' mental state.

PEP pruning algorithm is to overcome the shortcomings of REP algorithm that pruning dataset is not needed, but based on the false estimation of training dataset pruning algorithm. However, it also has some drawbacks, that is, it will lead to the large error of the estimation error rate of the algorithm. Therefore, the continuous correction in statistics is introduced to make up for this deficiency. That is, each leaf node is assumed to automatically misclassify the 1/2 instances it covers, and a constant is added to the subtree's training error. In calculating the standard error rate, the continuous correction follows the binomial distribution.

T represents the original tree, T_t represents the subtree with node t as the root, $r(t)$ represents the number of instances misclassified at node t, and $n(t)$ represents the number of all instances covered at node t. The classification error rate at node t is:

$$Y(t) = \frac{r(t)}{n(t)_{\mathrm{v}}} \tag{7}$$

The PEP algorithm corrects it to:

$$Y'(t) = [r(t) + 1/2]n(t) \tag{8}$$

The PEP algorithm is faster and more effective than other algorithms because each sub-tree can be accessed at most once during pruning. After constructing the decision tree, we use the iForest algorithm to recognize the abnormal psychology of students.

The iForest forest is composed of decision tree units, which are constructed as follows:

(1) Selecting abnormal characteristics of students' psychological state from the data set;
(2) Randomly selecting a value of the feature;
(3) Classify the data according to the characteristics, put the data less than this random characteristic value on the left, and put the data greater than or equal to the right;
(4) Then the left and right branches are constructed recursively until they are satisfied. The height of only one data or tree in the input dataset has reached the limit.

The structure of iForest is similar to that of a random forest, all of which are randomly selected features that break each tree apart. There are differences between the trees, which are built by combining n iTree trees into iForest.

In this paper, due to the imbalance of student data, the use of undersampling will eliminate too much sample data, which may lead to the loss of key sample information

and the lack of sample size, which will lead to the low accuracy and robustness of the training classification model. But only using over-sampling technology will expand a few samples and result in the over-fitting of classification. In order to solve the above problems, this section proposes a mixed sampling method for extremely unbalanced data. Through the mixed sampling of sample data, the problem of information loss caused by under-sampling and model over-fitting caused by over-sampling are avoided. There are two parts in the process of psychological state data collection. Firstly, the SMOTE algorithm is used to oversample the minority abnormal samples, that is, to interpolate each sample with its K samples to form a new minority sample. Secondly, the K-Means method is used to undersample most normal samples, that is, the data sets are clustered by K-Means method to eliminate redundant points in each class by calculating the cluster center.

Given a 2D array dataset, put the array into iForest for outlier identification. Build a binary tree with iTree n and iTree's sample count. The initial height of the tree defaults to 0. Through the mental view of students and the extraction of psychological characteristics of students, recursive, the distribution of psychological state of students under the abnormal point detection and recognition. Build n trees to form a forest. Because there are far fewer psychological anomalies than normal. Based on the principle of iForest priority, the algorithmwill run quickly and the average depth of iTree will be shallow. Setting the number of samples properly can effectively control the maximum depth and reduce the running time under the condition of ensuring the recognition effect. After several iterations of classification output, we can get the recognition result of abnormal psychology of students in the process of mobile teaching of physical education courses.

3 Method Testing

Middle school students are in an important stage of physical and mental health development. While teaching students the theoretical knowledge and skills, they also need to pay attention to their mental health. In this section, we will design the test scheme of the recognition method, and through the analysis of the test data, we can verify whether the recognition method can be used in the management of students' mental health.

3.1 Test Readiness

Because of the large number of middle school students in China, their abnormal psychological tendency may develop into serious psychological problems. School education should guide students to master mental health knowledge and adjust psychological methods, and give understanding and guidance when necessary, so as to promote the healthy growth of middle school students. According to the professional psychological evaluation form, we choose to manually organize the data of students with abnormal psychology to avoid errors in data entry due to different data sources. Specific dataset parameters are shown in Table 1 below.

Table 1. Data Sets of Students' Learning Behavior in Physical Education

Field name	Field settings	Field type
User_ id	bigint	Student ID
Age	intefer	Age of student
class	varchar(10)	A student class
S_act	varchar(32)	Student behavior
S_t	varchar(32)	Physiological parameters of students
PE_S	char(8)	Achievement of students in physical education
PE_SP	varchar(32)	PE activity performance rating

Firstly, the original data is preprocessed by removing the heavy and empty data. Then, the abnormal psychological recognition methods based on data mining, SVM and feature extraction are used to analyze the artificial data and identify the students who have abnormal psychology. By comparing the recognition result of the method with the known result, the recognition accuracy of the method is obtained. By comparing the recognition accuracy of this method, we can verify whether this method is helpful to the construction of students' mental health in teaching activities.

3.2 Test Results

The accuracy of the two abnormal mental recognition methods for different grades of students is shown in Tables 2, 3 and 4.

Table 2. Comparison of Abnormal Psychological Recognition of Grade One Middle School Students

Serial number	Method in this paper	Recognition method based on SVM	Recognition method based on feature extraction
1	86.94	73.08	66.55
2	86.56	72.95	69.08
3	87.21	74.07	66.26
4	86.12	74.26	66.93
5	87.63	74.61	69.37
6	88.09	74.26	66.13
7	87.14	74.14	68.57
8	88.81	74.63	68.71
9	87.98	72.82	66.42
10	86.34	74.45	66.84

Table 3. Comparison of Abnormal Psychological Recognition of Second Grade Middle School Students

Serial number	Method in this paper	Recognition method based on SVM	Recognition method based on feature extraction
1	88.76	76.15	66.46
2	87.61	75.44	67.14
3	89.14	76.58	66.65
4	89.25	74.51	67.29
5	87.13	75.62	65.94
6	85.65	74.65	69.42
7	85.72	76.35	69.26
8	88.48	74.76	65.83
9	87.57	76.20	66.11
10	85.74	74.92	67.13

Table 4. Comparison of Abnormal Psychological Recognition among Junior Middle School Students

Serial number	Method in this paper	Recognition method based on SVM	Recognition method based on feature extraction
1	89.26	75.90	69.87
2	89.19	76.68	72.30
3	88.93	75.12	69.92
4	86.86	75.74	70.73
5	87.38	75.56	70.58
6	87.67	74.01	70.12
7	88.84	75.15	72.34
8	86.93	75.93	72.75
9	86.95	75.47	71.76
10	87.52	77.12	72.23

Analyzing the data in Tables 2, 3 and 4, we can see that the accuracy rate of abnormal mental recognition is higher than that of the other two methods in different grades. Among them, the recognition rate of the method based on feature extraction is obviously improved because of the psychological abnormality caused by the pressure of entering a higher school. The average recognition accuracy of the method is 87.28%, the average recognition accuracy of the method based on SVM is 73.93%, and the average recognition accuracy of the method based on feature extraction is 67.49%. The average recognition

accuracy of the method is 87.51%, the average recognition accuracy of the method based on SVM is 75.52%, and the average recognition accuracy of the method based on feature extraction is 67.12%. The average recognition accuracy of the method is 87.95%, the average recognition accuracy of the method based on SVM is 75.67%, and the average recognition accuracy of the method based on feature extraction is 71.26%.

To sum up, in the process of PE mobile teaching, the method based on data mining proposed in this paper is more accurate and the evaluation result is more stable.

4 Conclusions

Mental health education in primary and middle schools is to meet the needs of the times and the healthy growth of students. It is a process in which the school exerts direct or indirect influence on students in a purposeful and organized way to improve their psychological quality and promote their all-round development. In this paper, we use data mining technology to study the method of abnormal mental recognition of students in mobile teaching of physical education. The test results show that the method has good recognition precision, and improves the reliability of the results of abnormal mental state recognition of students to a certain extent. And the recognition method proposed in this paper has strong reusability and extensibility.

Fund Project. Project of Shaanxi Provincial Department of Education: Research on the Intrinsic Mechanism of the Impact of College Students' Sports on Mental Health (Project No.: 20JK0288).

References

1. Xiang, S., Lin, W., Liu, J.: A randomized controlled study of effect of group counseling on facial expression recognition in college students with social delay [J]. Chin. Ment. Health J. **35**(03), 177–181 (2021)
2. Liu, D., Bao, L., Wan, C., et al.: Multi-layer partial information fusion model for psychological crisis identification of online forum users. J. Chin. Comput. Syst. **42**(04), 690–699 (2021)
3. Chakladar, D.D., Dey, S., Roy, P.P., Dogra, D.P.: EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm. Biomed. Signal Process. Control **60**, 101989 (2020)
4. Fei, Z., Yang, E., Li, D., et al.: Deep convolution network based emotion analysis towards mental health care. Neurocomputing **388**, 212–227 (2020)
5. Su, Y., Liu, M., Zhao, N., et al.: Identifying psychological indexes based on social media data: a machine learning method. Adv. Psychol. Sci. **29**(04), 571–585 (2021)
6. Shang, Y., Yang, N.: An improved particle swarm optimization back propogation neural network algorithm for psychological stress identification. Sci. Technol. Eng. **20**(04), 1467–1472 (2020)
7. Zhao, T., Huang, Z., Wang, X., et al.: Recognition research on concealing behavior in psychological test. Comput. Eng. Appl. **56**(20), 158–164 (2020)