Research Article

Research on Strategies of Promoting Mental Health of Higher Vocational College Students Based on Data Mining

Lingfei Yang

Department of the Student Affairs Office, Medical College of Jiaxing University, 314001, Jiaxing City, Zhejiang Province, China

Correspondence should be addressed to Lingfei Yang; lingfeiyang@stu.wzu.edu.cn

Received 30 March 2022; Revised 22 April 2022; Accepted 3 May 2022; Published 19 May 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Lingfei Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since entering the 21st century, my country’s education is facing the transition from traditional exam-oriented education to quality education and personality education, and the problem of mental health has gradually attracted the attention of all sectors of society. Suspension, drop-out, suicide, and crime occur from time to time, and the number of students with psychological disorders is also increasing year by year. Therefore, in-depth research on the mental health of vocational college students and exploration of students’ psychological intervention models have become the focus of colleges and other education departments. For the focus of research works, the application of DM technology in the field of teaching and management in an institution of higher learning has also achieved initial results. Through consulting a lot relevant literature, this paper deeply studies the DM technology and tries to analyze the application of DM technology in the data of psychological problems of higher vocational students. Through experimental comparison, a single algorithm model is used to classify and predict the test data set, and the accuracy rate is 78% after comparing the classification results with known categories. The accuracy of classification results obtained by the fusion algorithm in this paper is 85.1%. The results show that this technology provides help for students’ mental health counseling, mental health education and the prevention of students’ psychological problems in higher vocational colleges. It provides new ideas and methods to solve the mental health problems of higher vocational students and makes the school mental health education more purposeful, targeted, and effective.

1. Introduction

Healthy psychology has a very important impact on the development and growth of college students and indirectly affects the stability of campus, family harmony, and even social stability [1]. In this way, the research on the factors affecting college students’ mental health has aroused widespread concern in society. According to the development of national economic form, the society needs a lot of high-skilled talents [2]. Higher vocational education is an important part of China’s higher education, aiming at cultivating a lot of high skilled talents with reasonable structure and excellent quality [3]. The proportion of professional talents exported by higher vocational colleges will account for a very large proportion under the requirements of the economic structure [4]. At the same time, most vocational college students have poor academic performance and have not developed good study habits. In addition, vocational college students have a shorter academic system and lower academic qualifications than undergraduate students. Therefore, vocational college students will face greater psychological pressure in studying, interpersonal communication, and employment. College students usually shoulder the hope of their families and even the country. They have generally received a good education and should become the pillars of the country and contribute their talents to the national construction. However, the emergence of psychological problems has led to various tragic events. It is particularly important to find and prevent the mental health problems of college students in time.

Due to the emergence and wide popularization of database technology in recent ten years, it is possible to store, manage, and operate data. However, if we want to find useful knowledge hidden in massive data from database technology, find out potential rules, provide the basis for decision-making, use data more efficiently, and improve the utilization rate of information, we cannot start with database technology [5].
The technology of data mining (DM) and the technology of knowledge discovery have emerged from this, showing a strong role in all walks of life and a great function in education and scientific research [6]. Discover key relational information or regular connections in attribute categories in big DM, and mine potential information, through association rule analysis, to further find out the events that occurred and lead to the formation of other events in sequence or time rules [7]. In recent years, the application of DM in China has become more and more extensive, such as the application in urban traffic, security and environmental protection, crime analysis, and financial risk, and it has been gradually applied to all aspects of work and life [8]. At the same time, the application research of DM technology in college students’ mental health has also made some achievements.

Compared with traditional psychological counseling and psychotherapy, DM technology has more advantages in prevention and identification [9]. Based on the psychological DM system of higher vocational college students, this paper puts forward some innovations:

(1) Based on the fact that the algorithms of most higher vocational college students’ psychological DM systems are relatively single, this paper decides to adopt the innovation of combining multiple algorithms

(2) In addition, based on the combination of the above three methods, the GUI tool design using MATLAB is added to the psychological DM system of higher vocational college students to improve the effectiveness of these three algorithms

The chapters of this paper are arranged as follows. The first chapter is the introduction, which discusses the background and significance of the topic selection and expounds the innovation of the article.

The second chapter is the main body of this paper, which mainly puts forward the innovative achievements and research ideas of this paper combined with the research results of DM technology in the field of mental health of higher vocational college students at home and abroad.

The third chapter of this paper is the method part, which deeply discusses the application and principle of related algorithms and puts forward a new model of psychological DM for higher vocational college students based on the previous research results and the innovation of this paper.

The fourth chapter of this paper mainly discusses the experimental part of algorithm application. Through the results of the experiment and on the basis of sorting out the data, the psychological DM model of higher vocational college students is established.

The fifth chapter is the summary part, which summarizes the research results and shortcomings of this paper, as well as the prospect of follow-up research.

2. Related Work

As a subject with vitality and practical value, school mental health education has received extensive attention and developed rapidly in developed countries. Conley et al. proposed that the purpose of mental health education is to adapt to the diversity, difference, and comprehensive development of students through mental health education [10]. My country’s mental health education started in the 1980s, and it started relatively late compared with developed countries. With the development of society and the continuous progress of science and technology, more and more problems of mental illness appear, and mental health education has gradually attracted the attention of the state and society. Fu et al. proposed that most colleges and universities should conduct psychological surveys on students from the early stage of enrollment, and established mental health records for students [11]. The research of Woodford et al. shows that in order to make students develop in an all-round way, they must first have a good and healthy mentality in order to form a stable and healthy personality [12]. Loder et al. proposed the establishment of a mental health counseling center to provide psychological counseling services for school students [13]. The research by Shorey et al. shows that there is no large-scale self-funded way to carry out data mining technology research work in our country, and most of the research still stays at the government funding, in the form of national plans [14]. Mushonga and Henneberger take a higher vocational college as an example, on the basis of the school’s original psychological evaluation system, using data mining-related technologies to build a psychological data mining model, mining potential psychological information of students, and showing it to psychological counseling work in time personnel [15]. Yang et al. researches show that data mining technology is applied to the development of mental health management system, and the massive, fuzzy, and random data information in students’ mental health files is processed through relevant algorithms, and the mining is carried out. To a certain extent, it will help psychological counselors to make psychological judgments and prevention of students more scientifically and quickly [16]. Wei Hai’s research shows that in my country, data mining technology is still in the initial stage of application, but it also shows great commercial value [17]. Jennings et al. can see that with the development of computer technology in our country, people are more and more familiar with concepts such as data warehouse, business intelligence, and enterprise resource management. Mining and its predictive analysis have become a core technology for scientific management in all walks of life [18]. Wattick et al.’s preliminary research shows that the mental health status of vocational students has its group characteristics, and interpersonal problems are more obvious. The results show that there is a significant correlation between them, with different attribute characteristics [19]. The data used by Wang and Du came from the data research of the self-assessment scale SCL-90 and basic personal information in the mental health evaluation system of Hunan Vocational and Technical College of Industry, Hunan Vocational and Technical College of Science and Technology, and Hunan Engineering Vocational and Technical College. The psychological assessment database involving vocational colleges uses SQL Server 2008 to store and manage [20]. Eisenberg’s research shows that data mining
technology has its unique advantages over other technologies. Data mining can find the relationship between students’ psychological problems and their basic information and the main factors that lead to psychological problems from statistical data [21].

On the basis of the above-mentioned related research, this paper determines the positive role of DM technology in the field of mental health of higher vocational college students, constructs a DM system combining various algorithms, and makes deep analysis and research on the acquired and collected data by using DM technology, so as to make more effective use of the data, mine the valuable knowledge hidden behind the data, and discover and find the potential problems that affect the mental health of higher vocational college students.

3. Methodology

3.1. An Overview of DM Techniques. DM is a cutting-edge discipline in computer science, also known as data knowledge discovery, which refers to the process of mining useful information hidden in massive data or discovering potentially valuable relationships. It is an interdisciplinary subject that integrates multidomain knowledge such as database, machine learning, AI, and information retrieval [22]. With the development of computer network and storage technology, more and more data are accumulated every day. Data is specific, and knowledge is concepts, patterns, or laws, but data cannot help us directly; only data can help people make predictions and decisions by generating knowledge. DM is a data management, analysis, and processing technology that comes into being in response to this demand and is a key step in knowledge discovery [23]. Its main applications are correlation analysis, classification, clustering, forecasting time series model, deviation analysis, etc. Figure 1 shows a model of DM.

The above figure shows the basic process of DM: a lot of original data are extracted according to the data definition of the mining manager, and the data with correct and reliable format and easy to be processed by the Mining Wizard are obtained after data preprocessing and then enter the mining kernel to obtain the pattern set. After the pattern set is expressed and interpreted, it is obtained after the pattern screening of the mining manager, which is easy for users to understand knowledge base that can be generalized.

3.2. Basic Classification in DM. Classification is one of the important functions of DM and has become an important part of many DM systems. Classification is to construct a classification function based on given training data and use this function model to classify unknown data records to achieve the purpose of prediction. In general, there are many kinds of data, so it is very difficult to get the effective data we want quickly. Therefore, we can build a predefined model to describe the characteristics of different types of data sets or concept sets. Each record in the predefined data set has a corresponding category number; that is, it belongs to a known class. The model is built by analyzing the attribute characteristics of each record in the database.

3.3. Application of DM Technology in Mental Health of Higher Vocational Students

3.3.1. Demand Analysis. The competition in today’s society has evolved into a talent competition, which requires not only the competition of knowledge and skills but also the competition of psychological quality. Healthy psychology plays an important role in cultivating skilled and quality talents in higher vocational colleges, and it is also one of the basic qualities that higher vocational students should have to adapt to the society. Adopting reasonable psychological counseling methods and educational methods has far-reaching significance for mental health education in higher vocational colleges.

The school mental health management system has completed the collection, sorting, query, and simple result analysis of the mental health assessment of vocational college students, but only the data of the mental health surface is obtained. If we want to get more important knowledge and clarify the correlation between data, it is necessary to use DM technology to analyze and find out the relevant and valuable knowledge information among the attributes of students’ mental health data. At present, the popular symptom checklist 90 (SCL-90) is the symptom checklist 90 (SCL-90), which measures a person’s psychological symptom level in a certain period of time (usually one week).

Suppose that the total score measured by SCL-90 scale is \( T \), the total average score is \( T_a \), the item score is \( t_i \), an attribute factor is \( T_p \), the item score of the factor is \( T_{im} \), and the number of items constituting the factor is \( n \). \( T \) is the sum of the scores of 90 items. \( T_a \) is equal to the total score of each item constituting the factor divided by the number of items, that is,

\[
T = \sum_{i=1}^{90} t_i, \quad T_a = \frac{T}{90},
\]

\[
T_p = \frac{1}{n} \sum_{m=1}^{n} t_{im}, \quad \text{among } t_{im}, T_i \in [1, 5].
\]

3.3.2. Evaluation Process. Because of many iterations in DM, the mining tools used in various fields are different among widely used DM technologies. Choosing appropriate DM algorithms and DM tools to guide computer analysis and obtain valuable knowledge information is always an important part of DM. In this paper, the decision tree algorithm, BPNN, and Apriori algorithm are used to mine the mental health data of higher vocational students. The DM process of higher vocational students’ mental health evaluation is shown in Figure 2 below.

3.4. Mental Health Analysis Algorithm Based on DM Technology. Decision tree algorithm is a relatively classic classification DM algorithm. It generally uses a top-down recursive form to build generative decision trees from a lot of cases. The classification model of decision tree algorithm
is a directed acyclic tree composed of root node, parent node, child node, and leaf node. The decision tree algorithm is used for the prediction and analysis of mental health data of higher vocational students. The general idea is as follows: first, analyze and calculate which attribute is most relevant to psychological problems, as the root node of the decision tree, and then, use the iterative recursive method to analyze the rest in the same way. Attributes are classified, form decision trees, and build classification tree models for predictive analysis. The construction of decision tree is recursively realized by top-down greedy algorithm. At each internal node, select the test attribute with the best classification effect to classify the training sample set, and recursively call this process to construct the following subbranches until all attributes are used up or all training samples belong to the same category. If the data instance is of the same type as the node in the decision tree, it is classified into the same class; if the two are different, the instance is placed in the corresponding decision tree as a new node. Repeated operations can expand a decision tree containing only one root node into a complete decision tree, as shown in Figure 3.

Assuming that $S$ is a collection of $s$ data samples, then there are $m$ category attributes with different values: $C_i (i = 1, 2, \cdots, m)$. Let $s_i$ be the number of samples in $C_i$ class. For a given sample, the total information entropy is

$$I(s_1, s_2, \cdots, s_m) = - \sum_{i=1}^{m} P_i \log_2(P_i).$$

(2)
Among them, \( P_i \) is the probability that any sample belongs to \( C_i \), which can usually be expressed as \( s_j/s \). Let \( s_{ij} \) be the number of samples in the subset \( S_j \) with the category \( C_i \) and the information entropy value of the samples obtained by division is

\[
E(A) = \frac{1}{a} \sum_{j=1}^{a} \frac{s_{ij} + s_{2j} + \cdots + s_{mj}}{s} I(s_{1}, s_{2}, \ldots, s_{mj}). \quad (3)
\]

Finally, the information gain \( S \) of the sample set \( G \) obtained by division is

\[
G(A) = I(s_{1}, s_{2}, \ldots, s_{m}) - E(A). \quad (4)
\]

It shows that if the value of information entropy \( E(A) \) decreases, the value of information entropy \( G(A) \) will increase. Then, the uncertainty of test attribute \( A \) to classification will become very small.

A three-layer BP neural network is programmed with MATLAB. Since each student’s record includes 11 attributes, the input layer contains 11 neuron nodes. This problem is a binary classification problem in which students’ mental health status is “normal” or “abnormal,” so the output layer only contains one neuron. The number of neurons in the hidden layer is related to the network performance and needs to be determined experimentally. It has been found that in many applications where the cross-validation method is used to determine how many gradient descent iterations should be performed, a certain minimum number of hidden units is needed to learn the objective function accurately, and the redundant hidden units beyond this number will not obviously improve the generalization accuracy but will obviously increase the training time required by the network.

Assuming that there are \( C \) training sample sets with \( S \) results, the number of training samples is expressed by \( |S| \), and the training sample sets are divided into different classes \( C_1, C_2, \ldots, C_n \), and there are \( n \) different classes marked as \( |C_1|, |C_2|, \ldots, |C_n| \); then, the probability of class \( C_i \) for the training sample sets can be expressed as follows:

\[
P(S) = \frac{|C_i|}{|S|}. \quad (5)
\]

Then, the average information entropy can be expressed as

\[
E(S|C_1 C_2 \cdots C_n) = \sum_{i=1}^{n} \left( \frac{|C_i|}{|S|} \log_2 \frac{|S|}{|C_i|} \right). \quad (6)
\]

For two variables that can take a finite number of values, the conditional succession of \( X \) can be defined as

\[
E(X|Y=y_j) = \sum_{i} P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} . \quad (7)
\]

If \( Y \) is a known number, the average conditional entropy of \( X \) can be defined as

\[
E(S|A) = \sum_{j=1}^{m} P(S_j) \sum_{i=1}^{n} P(C_i|S_j) \log_2 \frac{1}{P(C_i|S_j)}. \quad (8)
\]

Subset \( S_j \) appears in training sample class \( C_i \) at a rate of \( P(S_j) = |C_i|/|S|, P(C_i|S_j) \).

To find the largest item set, you need to distribute the data set. The first step is to make simple statistics to obtain the frequent item set whose support is not lower than the minimum support set by the user. In the second step, it is necessary to construct association rules in the frequent item sets identified in the first step, with the requirement that the confidence level in the frequent item sets is not lower than the minimum confidence level set by the user. First, generate frequent item sets; then, expand the frequent item sets until they cannot be expanded. In the \( k \)th cycle, first, generate a candidate item set; then, scan the whole database; and find the \( K \)-dimensional frequent item sets by the generated support test.

4. Result Analysis and Discussion

The common method to deal with the inconsistency of BP neural network training results is to generate the network model twice with the highest accuracy and save it through many experiments. After repeated experiments, the accuracy of the network model is shown in Figure 4. The average value of accuracy is 0.9416, and the minimum value is 0.9295.

In this paper, the multialgorithm combination model is adopted. By adding social texts and observation texts published by vocational college students in a period of time collected in advance, the proportion of emotional vocabulary in the system is increased, which greatly reflects the control of emotional accuracy of the system. The experiments use 60%
of the data as training data and 40% of the data as test data. In addition to analyzing the overall accuracy of the model, the paper also simulates the effects of the three parameters that affect the psychological evaluation effect on the model performance. The results are shown in Figures 5–7 that affect the psychological evaluation effect on the model performance. The results are shown in Figures 5–7.

It can be seen from the figure that the accuracy of the multialgorithm combination algorithm on the data set is lower than that of the model, which is caused by the coincidence degree between the data and the emotional vocabulary system which is higher than that of the model. For \( \lambda_1 \) and \( \lambda_2 \), it is insensitive to the model accuracy within the value interval of [15, 25]; for \( \beta \), it is also insensitive to the model accuracy within the value interval of [10, 30]. Therefore, in general, the combination of multiple algorithms is insensitive to the changes of model migration parameters, which greatly reduces the training requirements of the model.
Figure 6: Relationship curve between $\lambda_2$ and classification accuracy.

Figure 7: The relationship between $\beta$ and classification accuracy.

Figure 8: The relationship between the classification accuracy of decision tree and the parameter splitmin.
5. Conclusions

This paper discusses the application of DM technology in the prevention of college students’ psychological crisis and completes the following research: (1) describe the research status of DM technology at home and abroad in recent years and its application in the prevention of college students’ psychological crisis and analyze the feasibility of applying DM technology in psychological crisis prevention and the commonly used model and system design methods. (2) Based on the previous research results and theories, the college students’ psychological DM models based on decision tree algorithm, BPNN algorithm, and Apriori algorithm are designed and implemented, respectively, and the feasibility is proved. (3) Based on the same training data and test data, the classification tests of the three classification mining models are carried out, respectively, and the classification accuracy of the models is compared. The posterior probabilities of the three classification models are compared by using formula theory, and the advantages and disadvantages of the three different models are compared. (4) On this basis, a DM model using three methods is proposed. (5) Use the GUI tool of MATLAB to verify the effectiveness of the system. Based on the structure, this paper designs the higher vocational students’ mental health evaluation system, applies the DM technology to the psychological system, realizes the automation of the collection of higher vocational students’ information and psychological evaluation information, increases the depth of psychological problem data analysis through the application of DM technology, and verifies the feasibility of using DM technology in the mental health evaluation system. Realize the automation of information collection and psychological evaluation of higher vocational students. The data analysis of mental health problems has been greatly improved through the application of DM technology, which verifies the reliability of DM technology applied to mental health evaluation system. Through this study, two-thirds of the data are selected as the training sample set to generate the algorithm model and algorithm rules. The remaining one-third of the data is used as the test data set, and the status of sensitive factors of interpersonal relationship among students in the test data set is known. A single algorithm model is used to classify and predict the test data set, and the accuracy rate is 78% after comparing the classification results with known categories. The accuracy of the classification results obtained by the fusion algorithm in this paper is 85.1%. It can be seen that the classification accuracy of the combined algorithm model is higher and the efficiency is obviously higher than that of the single-generation algorithm, which achieves the expected effect. Therefore, the results of classification DM using this algorithm are feasible for the analysis of psychological problems.

Data Availability

The data used to support the findings of this study are included within the article.
Conflicts of Interest
No competing interests exist concerning this study.

References


