

## Data Mining and Management System Design and Application for College Student Mental Health

Jiang Qinghua

UniversityLongyan,Fujian, UniversityLongyan,Fujian364000; Institute of Psychology.CAS, Beijing 100101  
jqh00@163.com

**Abstract**—Developing college students' psychological data management system is one of the important means to prevent college students from psychological crisis. So, data mining technology based on BP neural network is used in the current students' psychological data management system to improve its operation effectiveness. Design method and the feasibility of applying data mining technology in psychological crisis prevention is investigated. MATLAB 2014a is used to realize the kernel of the psychological data mining and the kernel is embedded into student psychological management system to improve the effectiveness of data mining technology in psychological crisis prevention.

**Keywords**—data mining; BP neural network; mental health

### I. INTRODUCTION

With the development of China's higher education process shifting from elite education to mass education, serious incidents due to psychological problems happened frequently, which not only brought tremendous pressure and grief on the family, but also brought adverse impact on the normal teaching and management order, triggering concerns of the community[1,2]. It is an emphasis and difficult point to screen out the students with mental health problems correctly and implement different levels of attention and custody to different classes of students in college students' psychological crisis prevention. Developing college students' psychological data management system is one of the important means to prevent college students from psychological crisis. However, most mental health analysis system stay at the level of data operation and management, which is of low application efficiency, providing a broad space and resources for the development and application of data mining technology in this area[3].

Personality questionnaire survey or symptom self-assessment scale (SCL-90) will be done for most of the freshmen's entrance in a month, and then the students are divided into different categories. At present, the more popular is the self-assessment list of symptom, also known as symptom checklist 90(SCL -90), which contains 90 projects[4-6]. Each object has five or four level of score, which was proposed by L.R.Derogatis in 1975, and the measured results is psychological symptom level of a person in a certain period of time. SCL-90 scale is simple and easy to understand, the measure angle is more comprehensive than UPI scale[12]. It has very good ability to distinguish people in the psychological barriers, and is applied to survey the crowd who may have psychological

problems[13-16]. SCL-90 scale includes 10 subscales, a total of 90 questions, namely the somatization, interpersonal sensitivity, forced symptoms, depression, hostility, anxiety, paranoia, horror, psychoticism and other symptoms such as sleep assessment. Scoring high or low reflects the severity of the symptoms. In addition, it also considers the total average score in the analysis of SCL-90 measuring. Therefore the data used in this article is of 11 dimension.

In the next section, psychological crisis prevention model based on BP artificial neural network is put forward. In section 3, data mining of psychological crisis prevention system is realized. In the end, some remarks are given.

### II. PSYCHOLOGICAL CRISIS PREVENTION MODEL BASED ON BP ARTIFICIAL NEURAL NETWORK

So far, clustering analysis algorithm has been applied in many fields, and achieved high grades[7-11]. In business, the clustering analysis can be used for the enterprise or the mall to find different customer base, and depict the characteristics of the different customer base by buying patterns. And clustering analysis is an effective tool for niche market, also can be used to study consumer behavior, looking for new potential market to be as the pretreatment of the multivariate analysis. Geographically, the clustering analysis can identify area with similar land use. On the Internet application, cluster analysis can be used to classify different types of documents and information. On e-commerce, grouping can be used to calculate customers with similar browsing behavior and analyzes the common features of the customer, can better help e-commerce users understand their customers, so as to provide better service to the customer. This can help decision makers to understand the needs of the market, and take the appropriate promotional gimmick, and gain more benefits. On the biological, clustering analysis is used in classification of plants and animals, and gain understanding of the intrinsic structure of the population, and provide scientific analysis for biological research. In short, the application range of clustering analysis is very wide, which can serve as a separate data mining module for classification, and can also be used as preprocessing step of other analysis algorithm. College students' mental health analysis put forward more requirements.

Data comes from 941 records of students' psychological census database in some university. The measuring tool of students' psychological survey is SCL-90 scale. In the students' psychological survey database, it records the students' psychological health indicator, and the table structure is number, name, gender, age, major, grade, time consuming, somatization, interpersonal sensitivity,

forcing symptoms, anxiety, depression, hostility, paranoia, horror, psychoticism, positive average score, total score, and state. The psychological data has the following features.

It has strong scientific feature. SCL-90 psychological measurement scale is widely used in the field of psychology, and has high well-known degree. It also can quantize the student behavior and psychological symptom, and the data has strong professional and scientific feature. It has high reliability. This part of the measurement data is stored in the database of relevant part of the university. These data is obtained is under the unified leadership of student work department. The normal data has 865 piece and the abnormal data has 76 piece. It will affect the feasibility and the accuracy of the testing results of data mining. The data preprocessing includes data purification, data discretization and data balance. These three methods can ensure the validity of the data. After data preprocessing, we obtain a data collection containing 1360 piece of data, which is simple, and has high reliability.

$$T = \sum_{x=1}^{90} t_x = \sum_{i=1}^{10} \sum_{m=1}^n t_{im}.$$

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

$T$  represents total score,  $T_a$  represents total average score,  $t_x$  represents score of the item,  $T_i$  represents score of some attribute factor and item score corresponding to this attribute factor is  $T_{im}$ . The number of item that constitutes this factor is  $n$ .

We use MATLAB to realize a three layer BP neural network. Because each student's record includes 11 properties, the input layer contains 11 neuron node. The problem for students' mental health status is normal or abnormal binary classification problem, so the output layer consists of a single neuron. Hidden layer neuron number is related to network performance, which need to be determined by experiment. Student psychological test results based on BP is shown in figure 1 and convergence curve is shown in figure 2 showing that the algorithm has good convergence performance. The number of hidden layers should be determined. Normally, hidden layer number is 1 or 2. Use 1 or 2 hidden layer network is very common, and occasionally we use 3 layers. We don't usually use more hidden layer, because the training time will be long and the artificial neural network only containing a hidden layer can do arbitrary function approximation. The layer 3 network already can represent considerable objective function. We intend to use the BP neural network with one hidden layer. Then we determine the number of hidden layer neuron. BP neural network is feed-forward back propagation network. The determination of the number of hidden layer neuron has not yet formed a perfect theory instruction system, which is usually designed through a lot of experimental data. After testing, the network has good performance when the number of hidden

layer neuron is 20. The real value of sample 1 to 144 is 0 which is normal category. It can be seen from the predictive value, there are seven samples wrongly labeled to be 1(abnormal category). The real value of sample 145 to 156 is 1(abnormal sample), which are correctly labeled to be 1(normal category). The exception classification accuracy is 100%, and this is the advantage of the model. Psychological prediction model based on BP artificial neural network can successfully predict student psychological status. It proves the feasibility and effectiveness of its application in terms of psychological problem.

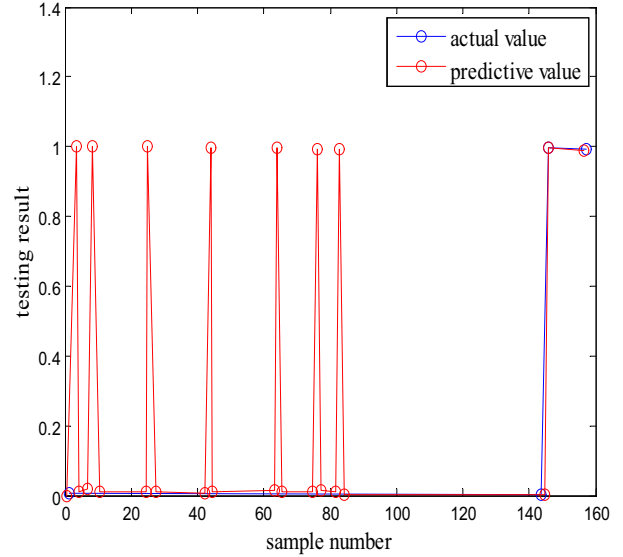


Figure 1. Student psychological test results based on BP

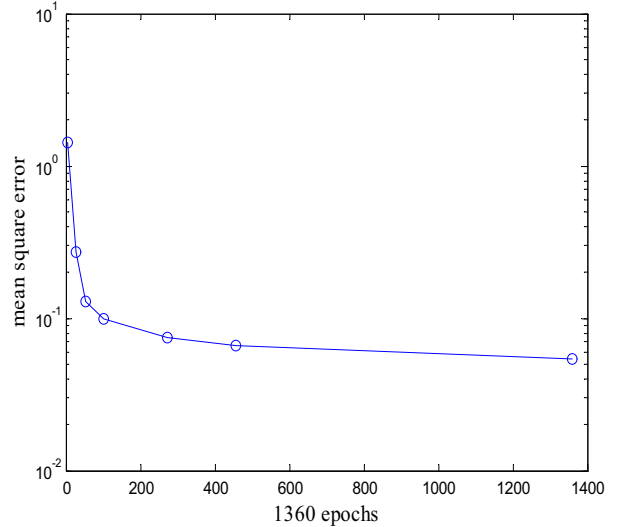


Figure 2. Convergence curve

### III. REALIZATION OF PSYCHOLOGICAL CRISIS PREVENTION SYSTEM BASED ON BP NEURAL NETWORK

Existing psychological management system provides the student/teacher service module, psychological consultant module, parent module and administrator module. The teacher service module provides students basic information data management, student measurement management, measurement data management and other management four modules. In the measurement results management module, the teacher can filter, query, sort, modify, and delete data. The system main function is to collect information about students, carry out psychological health level survey, set up student psychological archives, statistics and analysis of students' psychological services is also included. The system has a unique management ability of psychological measurement scale, powerful statistics, analysis, screening, convenient query and print function. Psychological crisis prevention system structure is shown in figure 3 and data preprocessing flow is shown in figure 4.

The system should have the following function according to the current system. The information transmission between each module is smooth, data interaction between each model is smooth, system response speed is fast, user interaction interface is friendly, and operation is easy. The embedded data mining module has good compatibility with the original system, data exchange is well, self protection of the embedded module and reliability requirements is high. It can implement a more complete data mining model, which can satisfy mining result comparison of different data mining model.

The system is divided into data layer, data mining layer and user interface layer. The data layer is mainly used to store students psychological raw data. The original data is dealt to get target mining data collection. Data mining layer carries out data mining for target data set through the data mining model. User interface layer can view the data mining classification results, evaluate and analyze the mining results and mine the knowledge rules. Students complete the test through psychological management system under the organization of university, and the data is stored in the psychological census database. Mental work evaluates psychological state of the students. After data preprocessing, we obtain mining data set. BP neural network is used for data mining. The final result is provided for mental workers.

With the continuous development of MATLAB and artificial neural network, MATLAB toolbox is integrated with neural network algorithm toolbox. Developers integrates correlation function implementation in MATLAB toolbox, and personnel can invoke the correlation function according to their own need, which makes personnel out from complex programming. We only need to use related toolbox function according to the guide of GUI. So we can concentrate and solve the related problem, which simplifies the development difficulty and improves work efficiency.

In the MATLAB R2014, we can use CBPDMtool command to open the design kernel. This module provides the trained network model, displays the network structure, and display results of classification image. This module can complete the establishment of BP network classification model, at the same time, it can classify students psychological data according to the existing network classification model.

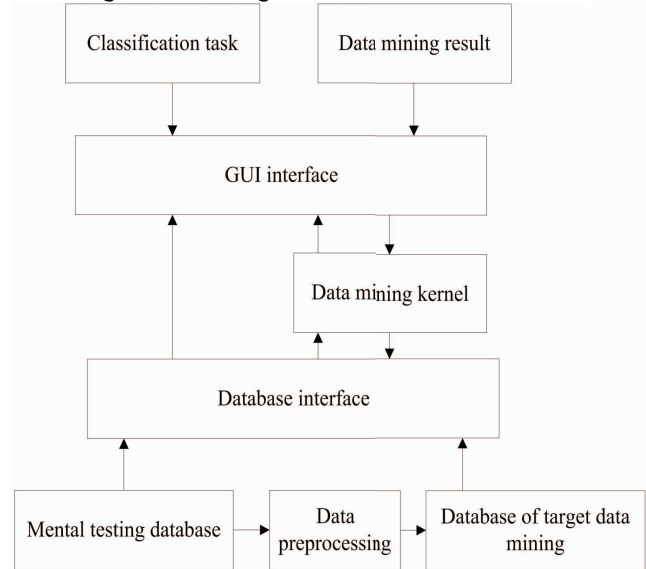


Figure 3. Psychological crisis prevention system structure

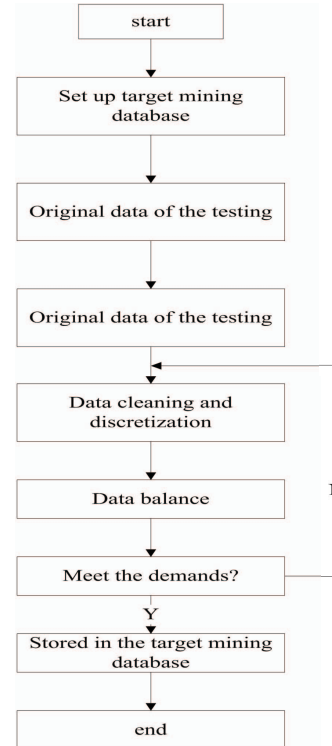


Figure 4. Data preprocessing

#### IV. CONCLUSIONS

We mainly discusses how to apply data mining technology to the current students' psychological data management system to improve its operation effectiveness. The research status of data mining technology at home and abroad in recent years and its application status in college students' psychological crisis prevention are expounded. Data mining technology is applied in the psychological crisis prevention and the feasibility of the model and the system design method are discussed. The abnormal category data is too little. In order to try to reduce the influence of unbalanced data on the result of the experiment, we repeatedly copy the only abnormal data set. In the classification test, the candidate abnormal data category is less, which has certain influence on the accuracy of experimental results. In order to make better application of data mining in this environment, we can promise not to reveal personal information, does not infringe the interests of others to establish standard psychological anomalies database within the limits of legal.

#### REFERENCES

- [1] Li-xin Huang, Hai-qiong Du, "A Web-based management system on undergraduates psychological archives establishment", *Journal of Baoding Teacher College*, vol. 19, no. 2, pp. 104-106, Aprl. 2006..
- [2] Wang Yu, Xiaofan Lv, "Application and research of data mining in cultivation of students' mental health", *Computer Programming Skills& Maintenance*, no. 4, pp. 89-90, 2013.
- [3] Yanping Gao, Ding Zhi, "the research of C4.5 algorithm on phychological health of college students", *Jiangxi Science*, no. 6, pp. 811-814, 2011.
- [4] Xin Li, Lishuo Zhang, "Research of data mining on psychological health of college students", *Computer Knowledge and Technology*, vol. 5, no. 21, pp. 5998-5999, 2009.
- [5] Linling Dai, "Preliminary exploration of data mining with its application to mental health of college students", *Reform & Openning*, vol. 16, 2009.
- [6] Xiaogang Wu, Ping Zhou, Wenhui Peng, "Application of decision tree algorithm in survey and evaluation of college students' mental health", *Computer Applications and Software*, no. 10, pp. 240-244, 2011.
- [7] Yuan Cui, Junpeng Zhang, Huanting Feng, "The application of data mining technology to life insurance dataset analysis", *Journal of Chengdu Medical College*, no. 2, pp. 137-140, 2006.
- [8] Dragan Gambergera, Nada Lavrac, Goran Krstacic, "Active subgroup mining: a case study in coronary heart disease risk group detection", *Artificial Intelligence in Medicine*, no. 28, pp. 27-57, 2003.
- [9] Rong Lei, Yuan Suo, Caina Li, "Relationship between mental health and family function teacher-student relations and five-factor personality of adolescents", *Chinese Journal of Clinical Psychology*, vol. 19, no. 5, pp. 687-689, 2011.
- [10] Xiujuan Li, Tian Xin, Feng Xin, "Research on Classification Technology in Data Mining", *Modern Electronics Technique*, vol. 20, pp. 86-88, 2010.
- [11] Laomo Zhang, Ying Ma, Guodong Wang, "Study of personalized and intelligent recommendation system based on data mining", *Modern Electronics Technique*, vol. 16, no. 31-34, 2011.
- [12] I. Yaroslavsky, J. Rottenberg, M. Kovacs, "The utility of combining rsa indices in depression prediction", *Journal of abnormal psychology*, vol. 122, no. 2, pp. 314, 2013.
- [13] S. W. Smith, R. Koppel, "Healthcare information technology's relativity problems: a typology of how patients' physical reality clinicians' mental models and healthcare information technology differ", *Journal of the American Medical Informatics Association*, vol. 21, no. 1, pp. 117-131, 2014.
- [14] D. M. Ndeti, R. Jenkins, "The implementation of mental health information systems in developing countries: Challenges and opportunities", *Epidemiologia e Psichiatria Sociale*, vol. 18, no. 01, pp. 12-16, 2009.
- [15] J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, A. d. Mendona, "Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy sensitivity and specificity of linear discriminant analysis logistic regression neural networks support vector machines classification trees and random forests", *BMC Research Notes*, vol. 4, no. 1, pp. 299, Aug. 2011.
- [16] P. I. Chow, B. W. Roberts, "Examining the relationship between changes in personality and changes in depression", *Journal of Research in Personality*, vol. 51, pp. 38-46, 2014.