A Study on the Application of the Decision Tree Algorithm in Psychological Information of Vocational College Students

Dongmei Cheng, Tan Li & Lihua Niu
Department of Public Basic, Langfang Health Vocational College, Langfang, Hebei, China

ABSTRACT: This paper discusses the basic operating principle and the development status of data mining technology, analyzes the insufficiency of the existing psychological management system, and proposes the development trend of psychological health education in colleges. According to an analysis on factors affecting college students' mental health and the deviation between the reality and the current number of students with psychological abnormality, this paper studies the application of data mining technology and puts forward a system based on data mining that combines the classified data mining technology with the existing psychological management system.

Keywords: decision tree algorithm; psychological information of college students; data mining

This analysis on college student's psychological resources with decision tree algorithm is to explore and analyze some problems in the psychological management system of students with the method of data mining. It is able to extract knowledge for the use of decisions from psychological measurement scales and basic information of students by embedding the data mining technology into the psychological management system. This not only has a theoretical value but also has practical significance in the digitalization construction of the mental health education. Practices indicate that the application of data mining in the psychological management system of students is the guidance for the mental health education of students. A correct technical route and a reasonable development mode can strengthen the functions of the psychological management system and improve the work efficiency of related staff workers, which helps to establish a stable, reliable, efficient and expandable psychological assistance system as well as a harmonious campus.

1 AN ANALYSIS ON THE APPLICATION OF DATA MINING IN THE PSYCHOLOGICAL MANAGEMENT SYSTEM OF STUDENTS

1.1 *A brief introduction and the research status of data mining technology*

The specific content of data mining is that those unknown and valuable rules and patterns are extracted from a relatively small, noisy and fuzzy database with a large capacity. Data mining is a complex process which is combined with database technology, statistics, artificial intelligence and other aspects. There are mainly two aspects in the realization of data mining. One is description and the other is prediction. The descriptive mining is able to describe general natures of mining data. The prediction methods should be selected if data needs to be inferred and predicted. In current studies, the data mining can be classified in line with categories of knowledge, such as association rules, classification and clustering rules, the neural network, generalization, prediction and so on.

1.2 *Application status and the key technology of the psychological management system*
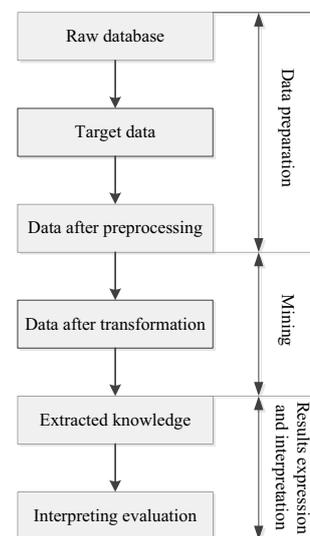

Figure 1. Implementation steps of the key technology

In the psychological management system of college students, factors related to college students' psychol-

ogy should be selected from a large amount of data so as to obtain more valuable data. The key technology is the data mining. It is able to master the whole psychological archives system of students and construct a scientific psychological archives service system for college students so that the psychological archives system can play a role with high efficiency and convenience in mental health education in colleges and the quality of mental health education for college students can be improved as well. Implementation steps are shown in Figure 1.

2 DATA MINING RULES OF STUDENTS' PSYCHOLOGICAL MANAGEMENT SYSTEM AND THE IMPLEMENTATION ALGORITHM

2.1 Features of the students' psychological management system database

The major purpose of the psychological archives system of students is to set up psychological archives for college students. Psychological archives are real records of such situations of college students as individual mental life experience, psychological characteristics and mental health consultation. Frequently used psychological measurement scales including the personality test scale UPI (University Personality Inventory) and the Symptom Checklist 90 SCL-90 (symptom check scale). But students' psychology changes with various aspects of the society. The mental health status of students can be practically reflected by scales with small contents. Thus, the mental health status of college students should be reflected by comprehensive analysis on other scales and basic information of students. It provides an important reference basis for the evaluation and prediction on college students' psychological behaviors and lays a foundation for psychological counseling and crisis intervention.

2.2 Application of decision tree in the student psychological management system

Decision tree learning is an inductive learning method. The most influential early-learning method is ID3 algorithm, which searches in all possible decision tree spaces through the top-down method. C4.5 algorithm is not only the improvement of ID3 algorithm but also the basis of other decision tree algorithms. The method of constructing a decision tree is the top-down recursive construction, which is shown in Figure 2.
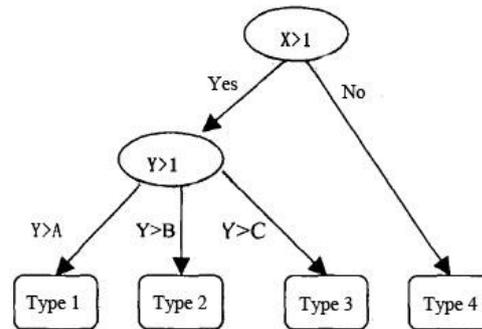

Figure 2. Decision tree

As for attributes of branches, C4.5 measures the division of branches with the information gain ratio. The valuable information included in branches accepts the impact of the information gain ratio. The larger the information gain ratio is, the more the useful data there is in branches.

In this paper, the data source is the mental health status of 20 students sampled from the 2010 class in a college. Table 1 is the training dataset.

The three attributes (harmonious relationship of parents, introverted, family poverty) in the training set

Table 1. Training dataset

No.	Introversion	Harmonious relationship of parents	Economic income	Psychological abnormality
1	No	Yes	Yes	No
2	No	No	No	No
3	Yes	Yes	No	No
4	No	Yes	No	Yes
5	No	No	No	No
6	No	No	No	Yes
7	Yes	No	No	Yes
8	Yes	No	Yes	Yes
9	No	Yes	No	No
10	No	Yes	Yes	No
11	No	Yes	Yes	No
12	No	Yes	No	No
13	Yes	Yes	No	No
14	Yes	No	No	Yes
15	No	Yes	Yes	No
16	Yes	Yes	No	No
17	No	Yes	No	No
18	Yes	No	No	Yes
19	No	Yes	No	No
20	No	Yes	No	No

of the 20 samples are calculated with the decision tree algorithm. The class attribute of 6 samples is "positive" and the class attribute of 14 samples is "negative". Results are provided in Table 2.

Table 2. Information amount of categorical attributes

Attributes / Information amount	Relationship of parents	Introversion	Family poverty
Expectation	0.5563	0.7471	0.8691
Gain	0.3250	0.3141	0.0121
Gain ratio	0.5842	0.1795	0.139

The attribute with the largest information gain ratio is selected as the root node through the comparison of the information gain ratio of each tested attribute. Apparently, the information gain ratio of "relationship of parents" is the largest. So, it is selected as the root of the decision tree. Divide the two values of the attribute "harmonious relationship of parents". Take "introversion" and "poverty" as root nodes of the two branches. The decision tree is shown in Figure 3.
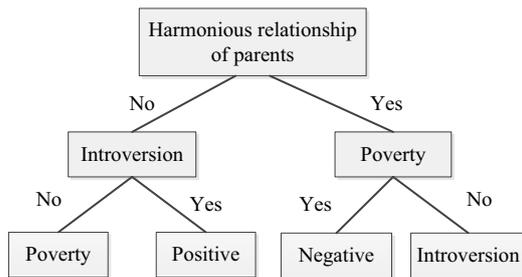

Figure 3. Generated decision tree.

2.3 Establishment of data mining rules and objectives of student psychological management system

The generation of a decision tree means that the knowledge represented by the decision tree can be extracted. The classification rule in the form of IF-THEN is expressed as follows:

If "harmonious relationship of parents"="no" And "introverted"="yes", then the proportion of students with psychological abnormality is large.

It can be seen from the above rules that the mental health of college students has a lot to do with the harmonious relationship of parents. It is unavoidable that conflicts between rural and urban economic concepts exist in college students' study, life and communications. Students from poor families are prone to the inferiority complex. In interpersonal communications, they have such behaviors affecting their mental health as taciturnity, being unsociable and eccentric, and so on. According to these situations, psychological counselor should carry out some educational work of

psychological health for introverted or impoverished students or students from single parent families and less harmonious families.

3 APPLICATION IMPLEMENTATION OF DATA MINING IN STUDENT PSYCHOLOGICAL MANAGEMENT SYSTEM

3.1 Architecture of data mining in the psychological management system

Decisions of high quality are dependent on data of high quality. Data preprocessing is especially needed in the mining of data with noises or incomplete and even inconsistent data. With the help of these data processing technologies before the data mining, the quality of the data mining model can be improved so as to reduce the time of actual mining. The structure is shown in Figure 4.
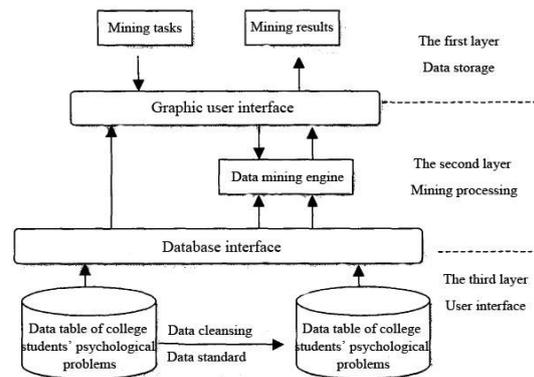

Figure 4. Structure of data mining for college students' psychological problems

In order to improve the efficiency of data mining, a great deal of research should be done before mining data. Data extraction is designed in line with the collected information, and required data sources which are determined according to the defined tasks so as to collect and extract data. In this paper, data attributes related to mining tasks are retrieved from the database according to features of college students' psychological data, such as results of SCL-90, results of UPI, family income and gender. These attributes provide a basis for the establishment of the categorical decision tree of reasons for students' psychological problems.

There are generally such problems as "missing values" and "multiple values" in selected data. The problem that interesting attributes lack attribute values should be dealt with by the rule of data cleansing. As for large-scale data sets, the process of data cleansing is automatic. First, define and determine the type of an error. Second, search and identify instances of the error. At last, correct the error.

(1) Data discretization: The continuous discretiza-

tion process of attributes is to divide the attribute space with a certain closed value. For example, values of factors such as anxiety and depression can be divided into five sections in line with the T of each dimensionality, namely no symptom, possible symptoms, weak symptoms, slightly severe symptoms and severe symptoms.

(2) Variables conversion: Convert the nationality of a student (the Han nationality or the national minority) into 11<10. This kind of conversion will bring a convenience for future data-mining and provide a guarantee for the rapid process of data.

3.2 *System implementation and application*

The main development tool used in the implementation of embedding data mining technology in the student psychological archives system is Macromedia Dreamwerver MX. Requirements on the operating system include: higher versions of Windows2003 Serve; support asp.net; SQL Server 2003 or higher versions for the backend database system. In the interface of Figure 3, click the "preprocessing" and then click the data preprocessing button after selecting the database required by mining, basic attributes and decision attributes. The preprocessing of data mining at the previous stage can be completed by data extraction, data cleansing and data standardization.

Construct the decision tree of the data after preprocessing with the C4.5 mining algorithm. The result is shown in Figure 3. Psychological counseling should be conducted in line with corresponding results after psychological counselors analyzing the rule obtained from data mining. The application of the C4.5 algorithm of data mining in the student psychological archives system is realized in accordance with the needs of college students' psychological health education. During the development stage of the embedding module, relevant verifications are accomplished with the combination of actual situations in colleges. SCL-90 psychological measurement scale results of the psychological investigation of new students, results of UPI and basic information of students are taken as the training set. Relevant information of 170 students is extracted as the training data to construct the decision tree of related attributes. The data mining results are consistent with the above experimental results.

According to practical verifications, reflections from psychological counselors and student management staff are good. It can be seen that studies on embedding data mining technology into the student psychological archives system not only are feasible but also make demands on psychological counselors and student management staff that relevant staff workers must cooperate with each other. Various methods are applied to the mental health education of college students.

4 CONCLUSIONS

Due to limited time and the insufficient experience of embedding a certain technology into the original system, the application of data mining in the students' psychological management system is still in its trial operation stage. Some functions are still imperfect and the data preprocessing can be only applied in the SCL-90 Symptom Checklist scale and the UPI scale. This topic requires further research and development in future work and studies. Follow-up work of mental health education should be also further improved, and problems of the following aspects will be emphatically considered in the future:

1. Add new functions on the basis of existing functions to further improve the system and provide more and better mental health services for psychological counselors, student management staff and students.

2. Carry out further studies on related algorithms of data mining technology and optimize part algorithms of data mining so that the data mining technology can play its role in more aspects. Facilitate all work in colleges with these advanced technologies.

3. Study the data mining technology based on the network environment, especially the data mining server established on the Internet, and cooperate with the database server so as to realize the perfect combination of date mining and mental health education.

REFERENCES

[1] Zhang, J.J. & Zhang, H. 2012. A study on the application of the decision tree algorithm in the student psychological archives system, *Value Engineering*, (17): 194-195.

[2] Wu, X.G., Zhou, P. & Peng, W.H. 2011. The application of the decision tree algorithm in college students' mental health evaluation, *Computer Applications and Software*, (10): 119-120.

[3] He, A.X. & Yuan, X.S. 2012. A study on the application of C4.5 decision tree algorithm in the employment management of application-oriented universities, *Journal of Chuzhou University*, (05): 113-114.

[4] Zhang, Q.C. 2014. A study on the application of the decision tree algorithm in students' performance analyses of secondary vocational schools, *Computer Programming Skills & Maintenance*, (10): 162-163.

[5] Hao, R.X. 2010. A study on the application of the decision tree algorithm in the score analysis system of computer basic courses, *China Educational Technique & Equipment*, (24): 103-104.

[6] He, G.D. 2013. The application of data mining in psychological problems of college students, *Wireless Internet Technology*, (02): 153-154.

[7] Dai, L.L. 2009. An exploration on the application of data mining in psychological problems of college students, *Reform and Opening Up*, (08): 192-193.

[8] Gao, Y.P. & Ding, Z. 2011. The application of C4.5 algorithm in the psychological education of college students, *Jiangxi Science*, (06): 205-206.