ISSN: 2664-9640

Summary of Research on Application of Decision Tree Algorithm and Association Analysis Algorithm in Warning of College Students' Psychological Crisis

Li Su and Yu Sun*

School of Information, Yunnan Normal University, Kunming, China;

*Corresponding Author: sunyu_km@hotmail.com

Abstract

The application of data mining technology to the identification of psychological crisis of college students is a relatively new warning method of psychological crisis. Different from traditional psychological crisis identification methods, data mining warning methods based on psychological evaluation data can effectively combine psychology and computer technology to conduct data mining and analysis on psychological evaluation data, thereby assessing the psychological status of college students from another aspect. Achieve warning of psychological crisis. After a brief introduction to traditional psychological crisis warning methods, this article focuses on the research of data mining in the warning of college students' psychological crisis, and introduces the commonly used features of decision tree algorithms and association analysis methods as well as the characteristics of these algorithms. The advantages and disadvantages of these methods are introduced.

Keywords

Data mining ,Psychological crisis warning, Decision tree, Association analysis.

1. Introduction

At present, with the rapid economic development, education, medical care, and culture have also made qualitative leaps. People enjoy unprecedented material prosperity, while mental health problems have gradually become a problem that people cannot ignore. As a group with a higher level of education, college students have more knowledge reserves, but countless examples prove that the psychological crisis of college students cannot be ignored. Compared with the elementary and middle school stage, the college stage is the stage of high incidence of psychological crises. On the one hand, due to the imbalance of educational resources, universities are mainly concentrated in big cities and provincial capitals. This indirectly leads to the fact that most college students need to stay away from their hometowns, relatives and former friends when they are going to school to face and deal with the difficulties encountered in the university alone. On the other hand, with the popularization of mobile phones and the Internet, information has been spread faster and more widely. Compared with the primary and secondary school students who are mainly facing academic pressure, college students are not only able to understand society more, but also face unprecedented anxiety.

The lack of social support has caused many college students' negative emotions to find no way out, which can easily lead to somatization symptoms such as anxiety, insomnia, headache, and fatigue. Even some college students have been suffering from psychological distress for a long time, and they have gone on a path of no return to destroying themselves. The fall of these young lives brought tremendous suffering to their families, but also caused society to suffer unmeasured losses. Therefore, universities, scholars in various fields and the whole society have gradually paid attention to and valued the psychological problems of college students, and

ISSN: 2664-9640

therefore invested huge manpower and material resources to deal with this problem. In fact, most scholars believe that the psychological crisis of most college students can be prevented. At present, colleges and universities deal with the psychological crisis of college students mainly including: first, every year when freshmen enter the school, use a psychological scale to conduct a psychological survey of students, screen positive individuals and make appropriate interventions; second, carry out regular psychological theme activities every year to popularize Psychological common sense; third, offer mental health courses to teach knowledge about mental health. Among them, the establishment of psychological files and the use of psychological scales to conduct psychological tests on students are the most important ways for colleges and universities to give warning to college students. This psychological warning method generally uses the College Student Personality Test Scale (UPI) or the Symptom Self-Rating Scale SCL-90 to test students and screen out positive individuals. It has the characteristics of fast speed, less workload, and higher accuracy.

2. Psychological crisis warning method

2.1. Traditional Warning Methods of Psychological Crisis

Traditional psychological crisis warning mainly relies on psychological evaluation methods, including investigation method, observation method, interview method, work analysis method, and psychological test method, as shown in the table below.

Table 1 Psychological assessment methods

method	introduction	advantage	shortcoming
Survey	Investigate the assessee's files, literature, experience, etc., and ask relevant personnel	Comprehensive content	The authenticity of the material is affected by the subjective factors of the investigator
Observation	Observe individual behavior	The observed behavior is true and reliable	Not easy to repeat, and the effect depends on the observer's own observation and analysis ability
Conference method	Face-to-face communication	Effectively obtain the true thoughts of the assessee	The effect depends on the evaluator's interview skills
Work analysis	Analyze the work of the assessee, including drawings, crafts, diaries, etc.	Beyond the limits of time and space	Easy to misunderstand
Psychological testing	Use standardized and quantitative questionnaires for testing, and the results can refer to the norm	High efficiency, repeatable, and effectively avoid the influence of subjective factors	Lack of flexibility

ISSN: 2664-9640

College psychologists use psychological evaluation methods to predict individuals with abnormal psychology among college students, and make appropriate interventions. It is still the most common psychological warning method used by colleges and universities. This method mostly uses the University Personality Inventory (UPI) or the SymptomChecklist90(SCL-90)to test students and give feedback to students who have a positive test result. the University Personality Inventory was compiled by the Japan University Health Care Management Association in 1966. It was translated and revised in 1993 by Fan Fuman and Wang Jianzhong of Tsinghua University. It is targeted at college freshmen to understand college students' psychosomatic disorders, neurosis, and schizophrenia. And other simple questionnaires for various troubles, confusion, dissatisfaction, conflicts, etc.0The SymptomChecklist90 is the most widely used mental disorder and mental illness outpatient checklist. There are 90 items, including 10 factors, which are anxiety, terror, compulsion, depression, paranoia, hostility, somatization, psychosis, interpersonal Relationship sensitivity and so on.

2.2. Data mining warning method based on psychological evaluation data

Freshmen in colleges and universities do psychological assessments every year, which produces a large amount of psychological data. Some scholars have begun to consider whether they can use data mining methods to mine and analyze these evaluation data and discover potential important relationships in order to help psychologists better understand and identify individuals with abnormal psychological conditions and realize psychological crisis warning.

Data mining methods mainly include classification, association, clustering, etc. Many domestic scholars use classification and association data mining methods to mine and analyze the data generated by psychological assessment, and achieve certain results.

2.2.1. Use decision tree model to analyze psychological assessment data

Decision tree is a basic classification and regression method. Decision tree learning usually includes three steps: feature selection, decision tree generation and decision tree pruning.

When using the decision tree method to analyze the psychological evaluation data, most of the relevant researchers use the decision tree ID3 algorithm or the C4.5 algorithm, taking basic information such as the student's gender, only child, major, and family location as the basic attributes. After preprocessing, the corresponding information gain or information gain rate is calculated respectively. Then build a decision tree with the psychological factors (such as somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, horror, paranoia, and psychosis in SCL-90) obtained from the student's psychological evaluation, and make a decision tree Prune, and finally form a rule.

Hou Zhen used the C4.5 algorithm to build a decision tree on the psychological evaluation data, and found that the main attributes related to the student's interpersonal sensitivity are whether it is only child, gender, single parent, and family economic level. After performing regression analysis on the relevant attributes of students, Zhao Dan proposed that the attributes that are most closely related to the psychological status of students are personality characteristics, family economic level, family relations, family composition, class status, and type of leave. Sun Weiping proposed that the main attribute of obsessive-compulsive disorder with students is gender, whether it is a left-behind child, family income, and the place of birth is rural or urban. Lu Wei proposed that the main attributes related to student depression are gender, whether it is an only child, family situation, and the place of birth is rural or urban.

2.2.2. Use association analysis model to analyze psychometric data

Association analysis methods are used to discover hidden relationships hidden in large data sets. The method is generally divided into two steps: using support to find frequent itemsets in the data set and using confidence to extract strong association rules from frequent itemsets.

ISSN: 2664-9640

Relevant researchers generally use the Apriori algorithm or FP-Growth algorithm in the association analysis method to conduct correlation analysis on students' psychological factors, and conduct correlation analysis on students' basic attributes and various psychological factors, in order to discover the correlation between psychological factors. The relationship between attributes and psychological factors.

After Yang Juan uses the Apriori algorithm to calculate the frequent itemsets of the psychological factors, he derives the association rules between the psychological factors: students with obvious somatization symptoms are more likely to be depressed, interpersonal sensitivity, etc.; at the same time, they are paranoid and anxious. Higher students are more likely to have somatization and depressive symptoms. After Chen Wan analyzed the data, it was found that individuals with obsessive-compulsive symptoms in the sample were more likely to be accompanied by symptoms such as depression, anxiety, sensitivity, phobia, and psychosis. These research results all show that there is indeed an association between psychological factors. Xie Qing proposed the relationship between individual basic attributes and psychological factors: girls are more likely to suffer from obsessive-compulsive disorder than boys]. Zhao Hongli uses the improved Apriori algorithm to propose that interpersonal sensitivity is related to major, gender, and age; students majoring in literature and history are more sensitive to interpersonal relationships than students majoring in science and engineering; compared with boys, girls are more sensitive to interpersonal relationships ; The older you are, the lower the sensitivity of interpersonal relationships.

3. Discussions

3.1. The shortcomings of traditional psychological warning methods

For traditional psychological warning methods, scholars Li Tongtong, Li Tan, Guo Xuning, etc. believe that the currently commonly used psychological crisis screening methods based on clinical diagnostic scales have problems such as large errors, low efficiency, and poor timeliness. The reason for the large error is generally due to the deviation of understanding and conception of some college students on psychological problems. Because most students are worried that if their psychological tests show abnormalities, they will be labeled as "depressed", "neuropathy", "mental disorder", and "mentally problematic" by the people around them. The consequence of this is that students are likely to face group isolation and even the consequences of school dismissal. In this way, it is not difficult to explain that many students will deliberately choose options with high social approval, and even provide false information. Of course, there are also some students who do not pay attention to psychological testing at all, and even have a psychological resistance, fill out questionnaires casually, or find classmates to complete the test instead. The low efficiency is because the previous psychological warning was mainly conducted through paper questionnaires for investigation and testing, which led to a lot of manpower, material resources and timeSince most colleges and universities have abandoned the traditional paper questionnaire survey method and switched to a psychological evaluation system, students can also log in to the system with a mobile phone or computer to conduct a questionnaire test. Now the efficiency of psychological warning has been greatly improved. Poor timeliness is caused by a combination of genetic, physical, psychological, family environment, education, interpersonal relationships, etc., and is always dynamic. Therefore, the psychological questionnaire can only preliminarily judge the psychological state of the individual in a period of time.

ISSN: 2664-9640

3.2. Insufficiency of data mining warning methods based on psychological evaluation data

From the research published so far, data mining warning methods based on psychological evaluation data mainly focus on using data mining related algorithms to find relevant characteristics of samples, establish psychological warning models through relevant characteristics, and conduct psychological warning analysis of samples. The psychological warning mechanism established by this method can be used as a supplement to the traditional psychological warning methods to help college workers better identify psychological problems. The shortcomings of related research are:

First, the relevant characteristic attributes of the constructed model are mainly derived from the most common and common student data information in student work, but in fact, the types of attributes that can give warning are far more than this information. If you can get more personal information, such as physical health indicators, diet, sleep and other action trajectories, it will help build a better model.

Second, when using the decision tree model for analysis, the decision tree model is prone to produce an overly complex model. Such a model will have poor generalization performance on the data, that is, it is prone to overfitting.

Third, when using data mining algorithms for model construction, computational overhead is high. For example, when using the association analysis algorithm to build a model, the algorithm itself scans the database too many times, which is computationally expensive; in addition, the rules found by the association analysis method may be false. Future research needs to further improve the efficiency of the algorithm and choose an objective and effective evaluation method.

Fourth, data mining warning methods based on psychological evaluation data are mainly based on supervised learning methods. The goal of prediction is to approximate the value measured by the user meter. Therefore, the best reliability and validity that the theory can achieve is the training sample used The reliability and validity of the scale.

4. Conclusions

The following table briefly summarizes the data mining warning methods based on psychological evaluation data and the data mining warning methods based on social media.

Table 2 Data mining warning methods based on psychological evaluation data

methods	Method principle	Feature	Algorithm
Based on psychological assessment data Data mining warning method	Starting from the root node, calculate the information gain for all features, select the feature with the largest information gain as the split feature, and build the node from the	Gender, whether he grew up with his parents, whether he was single parent, whether he grew up in a rural or urban area, the family's economic situation, family relations, personality characteristics,	ID3 Information entropy Entropy(t)=-$\sum_{i=0}^{c-1} p(i\|t)$log2p(i|t) 2) Information gainΔ=I(parent)-$\sum_{j=1}^{k} \frac{N(Vj)}{N}$ I(Vj) C4.5 Information entropy Entropy(t)=-$\sum_{i=0}^{c-1} p(i\|t)$log2p(i|t) 2) Information gainΔ=I(parent)-$\sum_{j=1}^{k} \frac{N(Vj)}{N}$ I(Vj) 3) Information gain rate =$\frac{\Delta info}{Split\ Info}$,

ISSN: 2664-9640

	value of this feature until the feature selection is completed or the information gain (C4.5 algorithm uses the information gain rate) Are very small to end the split.	whether to join a department, etc.	Partition information SplitInfo=- $\sum_{i=1}^{k}$ P(vi)log2 P(vi)
	Use association analysis algorithm: Through multiple scans of the database to calculate the support degree of the itemsets, all frequent itemsets are found, and then the association rules are generated.	Psychological factors: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, horror, paranoia, psychosis	Apriori Frequent itemsets generation 1: k=1 2: Fk={i|i∈I∧σ（{i}）≥N*minsup} { Find all frequent 1-itemsets } 3: repeat 4: k=k+1 5: Ck=apriori-gen(Fk-1) { Candidate set } 6: for Every transactiont∈T do 7: Ct=subset(Ck,t) { Identify the candidate set belonging to t } 8: for Each candidate setc∈Ct do 9: σ(c)= σ(c)+1 { Support count increase } 10: end for 11: end for 12: Fk={c|c∈Ck∧σ(c) ≥N*minsup} { Extract frequent k-itemsets } 13: until Fk ≠∅ 14: Result= ∪ Fk

As a supplement to the warning methods of psychological crisis for college students, data mining warning methods based on psychological evaluation data can make up for the shortcomings of traditional methods to a certain extent and identify some individuals in psychological crisis that cannot be identified by traditional psychological crisis warning methods. However, because these models based on data mining are modeled with traditional self-evaluation scales as calibration standards, the accuracy cannot surpass that of traditional psychological evaluation. How to choose new objective and effective evaluation indicators combined with data mining technology to establish a more accurate psychological crisis warning model will be a very important research direction.

Finally, researchers need to pay attention to the fact that psychological assessment data involves a large number of personal privacy and complex ethical and legal issues. In addition to the need for anonymization in data processing, relevant research also urgently needs relevant

ISSN: 2664-9640

workers to specify clear guidelines on authorization issues to protect personal privacy from infringement.

References

[1] Tian Wenhai,Liu Chunli.Problems and Countermeasures of Freshmen's Mental Health in Colleges and Universities[J].Journal of College Counselors,2020,12(05):75-80.

[2] Zou Zhiwen, Zhu Jinwei. Research and review of data mining algorithms [J] . Computer Engineering and Design, 2005, 26(9): 2304-2307.

[3] Hou Zhen. Design and implementation of college students' psychological evaluation and guidance system based on data mining[D]. Xidian University, 2020.

[4] Zhao Dan. Research and application of college students' psychological crisis early warning model based on decision tree[D]. Beijing Forestry University, 2020.

[5] Sun Weiping. Research on the Application of Decision Tree in Mental Health Evaluation of College Students[D]. Zhengzhou University, 2019.

[6] Lu Wei. Research on the Application of Data Mining Technology in the Mental Health Data of Vocational Students[D]. Anhui University, 2019.

[7] Yang Juan. Research on the Application of Data Mining Technology in Predicting Mental Health Problems of Vocational College Students[J].Science Consultation (Technology. Management), 2021, (03): 161-162.

[8] Chen Wan. Research on the Application of Data Mining in the Mental Health Management System of Vocational College Students[D]. Zhengzhou University, 2017.

[9] Xie Qing. Mechanism construction of psychological crisis prevention system based on association rule mining algorithm[J]. China New Communications, 2021,23(01):137-138.

[10] Zhao Hongli. Research on Improved Apriori Algorithm in Psychological Analysis of College Students[D]. Central China Normal University, 2015.

[11] Li Tongtong,Li Tan,Guo Xuning. Early warning of psychological crisis of college students based on social media big data[J/OL].Modern Distance Education Research: 1-12[2021-10-02]

[12] Zhang Lei,Chen Zhenxiang,Yang Bo.Analysis and Prediction of Personality of Social Network Users[J].Chinese Journal of Computers,2014,37(08):1877-1894.