2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT)

A new association rule mining method and its application in mental health data analysis

Xitang Zhang
Shandong Sports Lottery Management Center
Jinan, China
54413222@qq.com

Lei Zhang
Shandong Sports Lottery Management Center
Jinan, China
270500341@qq.com

Quanqing Yu
Shandong Sports Lottery Management Center
Jinan, China
184196444@qq.com

Zhe Zhang
Shandong Sports Lottery Management Center
Jinan, China
1433045832@qq.com

2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT) | 978-1-6654-8192-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/GCRAIT55928.2022.00117

***Abstract*-A lot of employees are plagued by all kinds of mental health problems. Mental health problems brought a lot of negative effects to them. Therefore, it is particularly important to find key laws by analyzing mental health data. Association rule algorithm can mine the potential relationship between the elements hidden in massive data, which is often used in data analysis in various fields. In this paper, we apply it to the analysis of mental health data in order to mine the correlation between various symptom factors and provide data-based scientific support for psychologists or relevant practitioners. However, traditional association rule mining algorithms will still produce a large number of redundant results even when the support or confidence is high enough, which undoubtedly adds obstacles to decision-makers in discovering key information. Therefore, in this paper, we propose a freely filterable association rule mining method, which can better locate the target results. Experimental results show that our method can mine association rules accurately among the various mental factors, and these rules can play a guiding role in the psychological construction of various organizations for their employees.**

Keywords-Association rule, Mental health, Data analysis, Association rule mining algorithms

I. INTRODUCTION

According to the relevant statistics, mental health problems of enterprise employees or students are becoming increasingly prominent. In view of this situation, many organizations regularly carry out mental health evaluation, and even provide mental health counseling free of charge. The research on mental health has been one of the most important issues that many researchers have focused on. For example, many enterprises have established their own employee mental health detection system, and employees can judge whether they have related symptoms by doing some simple question and answer tests[1]. But the hidden relationship between these symptoms cannot be found.

Association rule mining algorithms can mine the potential relationship between the elements hidden in massive data, which is often used in data analysis in various fields, such as financial[2], transportation[3], network log analysis[4], educational system[5], health care[6]. There are few studies on the application of association rules in mental health data [7,8,9]. Those studies, however, applied the traditional association rule mining algorithms in mental health data analysis, which produce a large number of redundant results even when the support is high enough, which undoubtedly adds obstacles to decision-makers in discovering key information.

Thus, in order to solve the problem of generating a large number of redundant rules in traditional algorithms, in this paper, we propose an association rule mining model that can freely filter the results, and apply it to the analysis of mental health data. The real data is collected from 500 employees' mental health test named SCL_90 from a company. Through our proposed method, we can simply and efficiently extract the correlation between the factors of mental illness. These rules may play a guiding role in how to regulate employees' mental health.

The remaining parts of this paper are organized as follows. Section II introduces our new model. Section III shows preprocessing of mental health data. Section IV presents the experimental results and analysis, followed by conclusions in section V.

II. OUR MODEL

In this section, we first introduce the related concepts of association rules in part *A*. Then we present our model in part *B*.

DOI 10.1109/GCRAIT55928.2022.00117

A. The concepts of association rules

Let $I = \{i_1, i_2,..., i_n\}$ be a set of n different items, where i_k (k = 1,2,..., n) is called item. The existing transaction database d is composed of transaction t (set of items, T_i), and its total number of transactions is represented by $| D |$. A is a set of items in I (itemset). If there are K ($1 \leq K \leq n$) items in itemset a, a is called k-itemset.

The support of itemset A is defined as the ratio of the number of transactions containing itemset a in database d to the total number of transactions in database d, i.e. $| D |$. Let $| a |$ represent the number of transactions containing a in database d, then the support degree $s(a)$ of itemset $a = | a | / | D |$.

Assuming that the minimum threshold given by experts to express their interest in itemsets, that is, the minimum support, is ms, the itemset satisfying $s(a) >= ms$ is called frequent itemset (FS).

When the implication $X \Rightarrow Y$ satisfies the conditions $X \subset I$, $Y \subset I$ and $X \cap Y = \Phi$, we call it association rule. Where X is called the antecedent of the rule and Y is called the consequent of the rule. In the same way as the calculation method of item set support, the support of rule $X \Rightarrow Y$ is the number of transactions containing X and Y in database D.

$$s(XY)=s(XY)=|XY|/|D| \quad (2.1)$$

The confidence of association rule $X \Rightarrow Y$ (abbreviated as c) refers to the ratio of the number of transactions containing X and Y ($| XY |$) to the number of transactions containing X ($| X |$), which is an extremely important measure of rule effectiveness.

$$c(XY)=s(XY)/s(X)=|XY|/|X| \quad (2.2)$$

B. The proposed model

Generally speaking, association rule mining can be divided into the following two steps: (1) generate frequent itemsets that meet the minimum support threshold ($s(XY) \geq ms$); (2) For each frequent itemset, an association rule satisfying the minimum confidence threshold ($c(XY) >= mc$) is generated. However, it will still produce a large number of redundant results even when the support or confidence is high enough. For example, there are one rule containing type A and 1000 rules containing type B in the mining results. But decision makers only want to focus on type A results. This makes it difficult for decision makers to find out the rules they focus on in many results. The model we proposed shown below can solve this problem well. The pseudo code of the model is as shown in Table 1.

Table 1. The pseudo code of the our model

Input：	*ms:* minimum support; *mc*：minimum confidence; *D*：Dataset; *I*: Items to be displayed
Output：	*Association rules(AR)*；
（1）	$AR = \Phi$
（2）	Generate *FIS* with the same steps as Apriori
（3）	**for** any *FIS X* **do begin**
（4）	**for** any itemset $A \cup B = X$ and $A \cap B = \Phi$ **do begin**
（5）	if $c(AB) >= mc$ then；
（6）	if I in $A \cup B$
（7）	return Set_AR ；
（8）	*End;*

III. EXPERIMENT DATA PROCESSING

The experimental data comes from the SCL-90 mental health evaluation results of 500 employees in a company. This is one of the most effective mental health evaluation forms in the world, and is widely used in outpatient examination of mental disorders and psychosis. It has the characteristics of rich reflection of symptoms and more accurate description of subjects' symptoms and other characteristics. It contains depression, anxiety and other mental symptoms.

Table 2 is part of the original data. We can see that all factors except gender factors are continuous values. However, the characteristics of association rule algorithm determine that its input must be discrete data, so we need discretize the continuous values.

Table 2. Part of the original mental health data

Gender	somatization	Obsessive	Interpersonal Sensitivity	Depression	Anxiety	Hostility	Fear	Paranoid	Psychoticism
Male	1.69	1.52	1.56	1.65	2.11	1.89	1.59	1.43	1.19
Male	1.36	1.71	2.29	1.21	2.05	2.33	1.45	1.76	2.05
Male	1.33	1.20	1.16	1.45	1.27	1.66	1.56	1.45	1.67
Male	1.67	2.23	1.78	1.89	1.89	2.33	1.47	2.09	1.25
Female	1.09	1.33	1.20	1.01	1.35	2.00	1.28	1.09	1.00
Female	1.82	1.78	2.00	1.56	1.31	1.56	1.83	1.17	1.35
Female	1.31	1.63	2.09	2.01	1.20	1.27	1.50	1.42	1.36

Table 3. The normal value

Factors	Normal value
Somatization	1.37
Obsessive	1.62
Interpersonal Sensitivity	1.65
Depression	1.5
Anxiety	1.39
Hostility	1.46
Fear	1.46
Paranoid	1.43
Psychoticism	1.29

Table 4. Data after discretization

Gender	somatization	Obsessive	Interpersonal Sensitivity	Depression	Anxiety	Hostility	Fear	Paranoid	Psychoticism
Male	A1	B0	C0	D1	E1	F1	G1	H1	I0
Male	A0	B1	C1	D0	E1	F1	G0	H1	I1
Male	A0	B0	C0	D0	E0	F1	G1	H1	I1
Male	A1	B1	C1	D1	E1	F1	G1	H1	I0
Female	A0	B0	C0	D0	E0	F1	G0	H0	I0
Female	A1	B1	C1	D1	E0	F1	G1	H0	I1
Female	A0	B1	C1	D1	E0	F0	G1	H0	I1

Table 3 shows the standard values of each factor. If any factor scores exceed the standard score, then we think that the student has a mental illness. The converted result is showed in Table 4.

IV. EXPERIMENT RESULTS AND ANALYSIS

In this section, we first compare the our model with traditional association rule algorithm [10] in number of excavations in part *A*. Then, the obtained results based on our model are given in part *B*.

A. Algorithm performance comparison

Our programs are performed in Java on a 64-bit Windows 10 Professional PC with 8 Intel Core i5 CPUs of 1.60 GHz, 8GB memory.

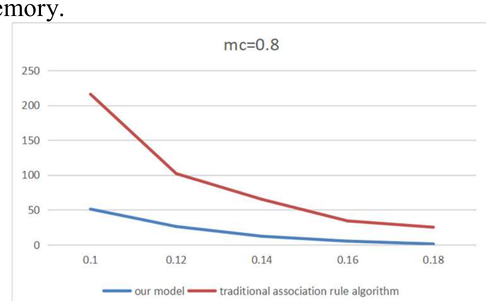

Figure 1. The number of rules changes with *ms*.

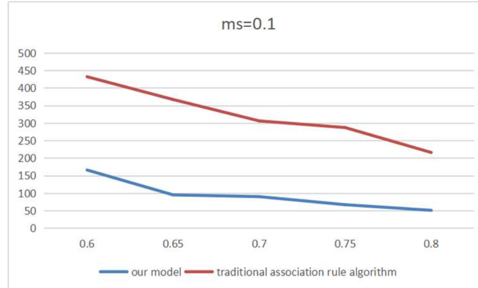

Figure 2. The number of rules changes with *mc*.

Figure 1 shows the comparison of the number of our model and the traditional association rule algorithm as the *ms* changes. And Figure 2 shows the comparison of the number of our model and the traditional association rule algorithm as the *mc* changes. As we can see that, under the same *ms* or *mc*, the number of our model is more than that of the traditional association rule algorithm. This is because our model can filter the results according to the input parameters .

B. The obtained results

According to the principle of the model, the higher the value of confidence, the stronger the correlation between the elements. Table 5 shows the higher confidence (>80%) level of the rule.

Table 5. Mental health data after discretization

Number	Association rules	Confidence
1	male⇒somatization, obsession	95.15%
2	female⇒anxiety	92.56%
3	obsession⇒paranoid	90.35%
4	obsession , anxiety⇒depression	86.37%
5	anxiety, interpersonal relationship sensitiveness ⇒depression, obsession	86.29%
6	depression, obsession ⇒anxiety, interpersonal relationship sensitiveness	82.78%
7	interpersonal relationship sensitiveness, fear⇒paranoid	81.50%

As shown in Table 5, there are different correlations among different factors. And there are obvious differences between men and women. Rule "male⇒somatization, obsession(conf=95.15%)" means through the analysis of our data samples, male have a 95.15% chance of suffering from somatization and obsession at the same time. Similarly, we can see female have a 92.56 percent chance of suffering from anxiety disorder. If a person is suffering from obsessive-compulsive disorder, the probability of having paranoid is as high as 90.35%. People suffering from both obsession and anxiety have a probability of 86.37% suffering from depression. From rules 5 and 6, we can see that there is a strong correlation between the four symptoms of anxiety, obsession, interpersonal relationship sensitiveness, and depression. The last rule in Table 5 tells us people suffering from both interpersonal relationship sensitiveness and fear have a probability of 81.50% suffering from depression.

V. CONCLUSIONS

In this paper, we propose an association rule mining algorithm that can freely filter the results, and apply it to the analysis of mental health data. The real data is collected from 500 employees' mental health test named SCL_90 from one company. Experimental results show that our method can mine association rules accurately among the various mental factors, and these rules can play a guiding role in the psychological construction of various organizations for their employees.

REFERENCES

[1] Chen H. Research on Psychological Health Education of College Students Based on Positive Psychology Theory[J], Management Science and Research. 2019, (1): 43-45.

[2] Sawangarreerak S, Thanathamathee P. Detecting and Analyzing Fraudulent Patterns of Financial Statement for Open Innovation Using Discretization and Association Rule Mining[J], JOItmC. 2021, 7(2):128.

[3] Hong J, Tamakloe R, Park D. Application of association rules mining algorithm for hazardous materials transportation crashes on expressway[J], Accident Analysis & Prevention. 2020, 142(5):105497.

[4] Li C, Li W. Automatic Classification Algorithm for Multisearch Data Association Rules in Wireless Networks[J], Wireless Communications and Mobile Computing. 2021, (1):1-9.

[5] Kittur J, Bekki J, Brunhaver S. Development of a student engagement score for online undergraduate engineering courses using learning management system interaction data[J], Computer Applications in Engineering Education. 2021.

[6] Zhou S, He J, Yang H, Chen D, Zhang R. Big Data-Driven Abnormal Behavior Detection in Healthcare Based on Association Rules[J], IEEE Access. 2020, 8:129002-129011.

[7] Qi W, Jie Y, Huang S, Lei G, Lu R. The Application of Association Rule Mining in College Students'Mental Health Assessment System[J], Journal of Hunan University of Technology. 2013.

[8] Wen-Juan Q I, Huang S C. Statistical Analysis and Association Rule Mining of Application in College Students' Mental Health[J], Computer Systems & Applications. 2014, 23(10):228-232.

[9] Meng Q , Sha J . Tree-based frequent itemsets mining for analysis of life-satisfaction and loneliness of retired athletes[J], Cluster Computing. 2017, 20:3327-3335.

[10] Agrawal R. Mining association rules between sets of items in large databases[C]. Acm Sigmod International Conference on Management of Data, 1993, 207-216.