Design and Implementation of Chinese Traditional Medicine Association System Based on Improved Association Rules Algorithm

Ruixiang Wang[a]

Department of Management, Shandong University of Traditional Chinese Medicine, jinan, 250355, China.

[a] Corresponding author: wrx731115@163.com

Abstract. As an important descriptive data mining method, association rules based on apriori algorithm have been widely used in the discovery of the core prescriptions of traditional Chinese medicine, and the correlations between symptoms and syndromes, syndromes and modern examination results, and symptoms and traditional Chinese medicine and other practical issues. However, we should see that the method of traditional association rules based on the frequent item set cannot fully reflect the complexity of traditional Chinese medicine. Chinese culture is an inclusive culture. Both effective and ineffective experience are written in the book. Many useful experiences are not frequently displayed. If the traditional association rules are followed, many important information will be abandoned. This study inherits the concept of support degree of traditional association rules and adds the parameters of asymmetric relative degree (JACCARD) as a supplement to the frequent item set. The new algorithm is supported by the Python platform and the traditional Chinese medicine association service system has been finished, well-performed, able to discovery secret, yet valuable associative knowledge. The system is expected to find effective experiences from the experiences of Chinese history and contribute to the inheritance and innovation of Chinese medicine.

Keywords: Data mining; Association rule; Algorithm; Service system; Traditional Chinese medicine.

INTRODUCTION

An authoritative view of Chinese medicine is that inheritance is the foundation of innovation [1]. This view holds that all discoveries based on clinical and modern science should be based on the premise and condition of the records in ancient books. This view was confirmed in two examples. One example is Mrs.Tu Yuyu found that the use of Artemisia annua to treat malaria was documented in ancient books. Moreover, in the books written by Ge Hong in the Tang Dynasty, she found out how to concoct Artemisia annua in order to play its role in the treatment of malaria. Thus, Tu Yuyu eventually broke through the technical bottleneck and extract anti-malarial drug Artemisinin from the juice of Artemisia annua using ether according to records of ancient Chinese medicine records. Mrs. Tu was awarded the 2015 Nobel Prize in Physiology and Medicine for her outstanding contribution to the discovery of artemisinin [2]. The second example is Zhang Tingdong et al.'s extraction of arsenic trioxide from arsenic, based on the ancient "fight poison with poison" principle and the association between symptom and Chinese medicine in numerous medical records. Arsenic trioxide was eventually used in the treatment of leukemia and achieved very good results. The above two findings are the most significant findings in the modernization of traditional Chinese medicine. The two major discoveries, without exception, were inspired and guided by the experiences recorded in ancient books [3].

Ancient Chinese medicine is a huge treasure house. However, we should also realize that the truths and lies, science, and superstition among the records recorded in ancient books do not allow the truth to be obtained through simple, frequent ensemble calculations. For example, the disease of malaria is associated with "cold and hot exchanges" in ancient books. In traditional Chinese medicine, the use of Artemisia annua to treat malaria is not a

mainstream treatment. After Emperor Kangxi of the Qing Dynasty had acquired malaria, a large number of doctors could not use Chinese medicine to relieve his symptoms, and finally he was cured by taking quinine brought by the missionaries. By the same token, Chinese medicine is often taken after boiled with water, and the method of soaking Artemisia annuus with cold water has only appeared once in massive ancient medical books. Therefore, it is impossible to obtain the truth of Chinese medicine for the treatment of malaria simply by analyzing the association between disease and Chinese medicine and between the Chinese medicine processing methods and Chinese medicine. The above facts show that the current method of association rules cannot obtain the truth we need in the experience of the ancients, and we need to establish a whole new method for the knowledge mining of traditional Chinese medicine.

INTRODUCTION OF ASSOCIATION ANALYSIS

In short, an association rule is an expression $X \Rightarrow Y$ where X and Y are items set. The purpose of association rules is to find possible connections between things through data.

If S is a set of items, $X \subseteq S$ is called an itemset, and X is called a k-itemset if the cardinality of itemset X is k. assuming that T is a set of transactions, SUPPORT(X) means the probability of X appearing in T. In other words, the support of an itemset measures how often X occurs in the transactions. In addition, the association rules also have two important parameters: confidence and lift. Confidence reflects the accuracy of the association rules, lift reflects whether the front and the back of the association rules are positively correlated, negatively correlated, or irrelevant. The calculation formulas for these three parameters are as follows [4]-[5]:

$$SUPPORT(X \Rightarrow Y) = \frac{num(X,Y)}{num(I)} \tag{1}$$

$$CONFIDENCE(X \Rightarrow Y) = P(Y \mid X) \tag{2}$$

$$LIFT(X \Rightarrow Y) = \frac{CONFIDENCE(X \Rightarrow Y)}{SUPPORT(Y)} \tag{3}$$

Num (X, Y) refers to the number of transactions that include both X and Y; num (I) refers to the total number of transactions in transactions.

If support(X) >= minsupport, we say that X is a frequent itemset. Finding frequent itemset is the premise of finding association rules. At present, the most commonly used algorithm for finding frequent itemset is the apriori algorithm, which continuously builds Cartesian products to find n-itemset that satisfy the conditions.

Though support, confidence, and lift are the three main parameters of the association rules, in many applications of TCM association rules, the purpose of the researcher is simply to find the probability that different items co-occur, so confidence and lift are meaningless. Therefore, it is sufficient to find frequent item sets based on the association rules algorithm without confidence and lift. At the same time, in order to enhance the discovery of Chinese medicine knowledge, we have introduced the concept of asymmetry similarity (JACCARD). JACCARD similarity coefficient as displayed below [6]:

$$JACCARD_{sim}(X,Y) = \frac{num(X,Y)}{num(X) + num(Y) - num(X,Y)} \tag{4}$$

Num (X, Y) is the number of transactions that contain both X and Y; num(X) is the number of transactions that contain X; num(Y) is the number of transactions that contain Y.

From the content of the JACCARD coefficient, it can be seen that if the two items are sufficiently close, even if their frequency of occurrence is not high, the JACCARD coefficient will be large enough. We provide that as long as the support is greater than the minimum support or the JACCARD coefficient is higher than the specified threshold, as long as any one of the conditions is satisfied, the set of items found is the set of items we need.

ALGORITHMIC DESCRIPTIONS

In the section, we describe the improved apriori algorithm based on traditional apriori algorithm [7] for TCM association system as follows:

L= {}; //L should contain all itemsets that meet the conditions

L1 = Finding frequent 1-item set of T;

L=L $\bigcup$ L1;

For (k = 2; Lk-1 is non-empty; k++) {

Lk= Cartesian Product (Li,Lj);//Li,Lj⊆Lk-1;

If (support (Lk)>=minsupport)

L=Lk $\bigcup$ L;}

t =total (item); //get the total of the no-repeat items;

While ((i! =j 0<=i,j<=t) and (JACCARDsim(A[i],A[J])>=minJaccard))//A[i],a[j]is no-repeat item

L=L $\bigcup$ (A[i], A[j]);

From the algorithm, it can be seen that the support-oriented apriori algorithm finds a frequent set that satisfies the condition and whose support is greater than the minimum support. On this basis, we calculate the JACCARD between any two non-repetitive single item and pick out binomial sets with all JACCARD values greater than the established threshold, and finally add to the frequent item set just calculated.

IMPLEMENTATION OF TCM ASSOCIATION SERVICE SYSTEM BASED ON IMPROVED APRIORI ALGORITHM

This research bases on the improved apriori algorithm, adopts python3.4 platform, adopts django's MVC architecture, and finally forms the platform of Chinese medicine associated service system.

System interface is shown in Fig.1.

We used more than 300 prescriptions in the Chinese classic book "SHANGHAN ZABING LUN" as the entry target. The format is shown in Fig. 2.

We set when the minimum support to 0.2 and the minimum Jaccard coefficient to 0.9, the part analysis results are as follows:

Licorice, Chinese date. Support: 0.21;

Guizhi, Licorice. Support: 0.22;

White Quartz, talactites. Support: 1.0;

bignoniad, moutan bark. Jaccrd coefficient: 1.0;

Coltsfoot, Aster. Jaccrd coefficient: 1.0;

Qinpi, pulsatilla. Jaccrd coefficient: 1.0;

Coltsfoot, blackberry lily. accrd coefficient: 1.0;

Antelope Horn, Cinnamon. Jaccrd coefficient: 1.0;

Blackberry lily, aster. Jaccrd coefficient: 1.0.

From the final analysis results, it can be seen that with the minimum support level of 0.2, the associated Chinese medicine pair that can be obtained using the support threshold is very limited, and because the licorice has the function of regulating herbal property in TCM theory. Therefore, most of the rules involving licorice are meaningless. So, valuable pairs of drugs can only be obtained through the JACCARD threshold in this example.

From the results, we can see that Coltsfoot and Aster are closely related Chinese medicines. At the same time, by collecting literature, we found that some people have done experimental research about Coltsfoot and Aste. The researchers found that after collocation of Coltsfoot flowers and Aster, inhibited LPS and stimulated NO production of macrophages, which significantly enhanced the anti-inflammatory effect [8]. The above research shows that the improved mining algorithm for TCM data is valuable, and Chinese medicine researchers can use the results as a clue to design experiments to discover the nature of prescription compatibility rules and develop new effective Chinese herbal compound.
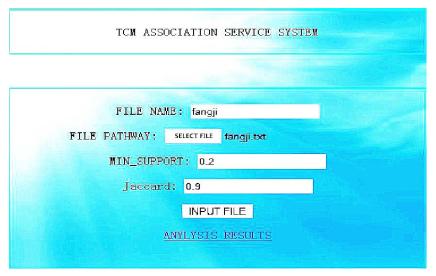
FIGURE 1. TCM association service system interface

```
rhizoma_anemarrhenae licorice plaster polisthed-round-grained-rice cassia
rhizoma_anemarrhenae licorice plaster polisthed-round-grained-rice ginseng
rhizoma_anemarrhenae plaster licorice polisthed-round-grained-rice
Lily Rehmannia
Lily Chicken_Yellow
Scanllion ginger monkshood
Scanllion dried_ginger monkshood pig_bile
Pulsatilla licorice Ejiao Qini Hunglian Phellodendron
Rhubard monkshood Asarum
```

FIGURE 2. Part TCM classic prescription in "SHANGHAN ZABING LUN"

REFERENCES

1. WU Jing-jing, ZHANG Honglei, GAO Shan. "Inheritance and Innocation: Exploration and Practice of Enhancing the Cultural Soft Power of Traditional Chinese Medicine in Higher Medical Colleges and Unviversities" in J.China Health Management, 2017, 12 (354):936-939,960.
2. LIU Jitang, LING Ziping." Historical Tracking on the Invention of Artemisinin and its Enlightenment" in. JOURNAL OF NEW CHINESE MEDICINE, 2018, 50 (3), pp.1-5.
3. RAO Yi,LI Runhong, ZHANG Daqing. " The scientific research monument of Chinese medicine" in Technology Review, 2015, 33 (20), pp.132-136.
4. R.Srikant,R.Agrawal."Mining Databases for Association Rules" in VLDB 1996:134-145.
5. C.C. Aggarwal, P. S. Yu: "Mining Large Itemsets for Association Rules" in Data Engineering Bulletion 1998,21(1):23-31.
6. Suphaki Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, etc. "Using of Jaccard Coefficient for Keywords Similarity" in Proceedings of the International MultiConference of Engineers and Computer Scientists, 2013, March 13-15, Hong Kong.
7. Mohammed Al-Maolegi, Bassam Arkok."AN IMPROVED APRIORI ALGORITHM FOR ASSOCIATION RULES" in International Journal on Natural Language Computing, 2014, 3 (1), pp.21-29.
8. LI Cong, HUANG Fang," DOU Chang-gui, etc. Anti-inflammatory effects of the compatibility of Radix Asteris and Flos Farfarae" in Chin J Clin Pharmacol Ther,2009,14(2), pp.155-159.