

A METHOD OF ATTRIBUTE PARTIAL ORDERED STRUCTURE DIAGRAM FOR THE COMPOSITION STRUCTURES OF PRESCRIPTION AND KNOWLEDGE DISCOVERY

FENGJIE FAN^{1,*}, WENXUE HONG^{1,2}, JIALIN SONG¹
SHAOXIONG LI¹ AND CUNFANG ZHENG¹

¹School of Electrical Engineering
Yanshan University

No. 438, Hebei Ave., Qinhuangdao 066004, P. R. China

*Corresponding author: ffjmz@126.com

²Big Data Visualization Technology Center
Northeastern University at Qinhuangdao

No. 143, Taishan Road, Qinhuangdao 066004, P. R. China

Received August 2015; accepted October 2015

ABSTRACT. *Data mining can find out the potential valuable information from massive data. It is one of the most active research fields. With the development of data mining technology and traditional Chinese medicine information system, many data mining methods have been applied to traditional Chinese medicine research. Because of the huge-ness and complexity of data, it is difficult to study the compatibility of Chinese medicine formulae. In this paper, a visualization method of attribute partial ordered structure diagram for the prescription compatibility is proposed based on the formal concept analysis theory. Formula data was chosen from the classic medical books of Zhang zhongjing pre-scription, and composition structures of prescription-drug, syndrome-drug were analyzed by attribute partial ordered structure diagram. The result shows that the attribute partial ordered structure diagram provides a new method to discover the new knowledge of pre-scription compatibility.*

Keywords: Data mining, Formal concept analysis, Attribute partial ordered structure diagram, Compatibility of prescription, Knowledge discovery

1. **Introduction.** Data mining, which is also called knowledge discovery in database, is a process that extracts connotative, unknown, potentially valuable information and knowledge [1]. The extracted knowledge can be expressed as the concepts, rules, regularities, patterns and so on. Now statistical methods, machine learning, neural computation and visualization are commonly used methods of data mining. In recent years, data mining is widely applied in the field of traditional Chinese medicine researchers. Shang et al. analyzed the incompatibility characters and methods of Chinese herbal medicine by data mining method. The results show that the specific property combinations of nature, flavor and channel tropism are verified to exist in incompatible drugs pairs, which is evidently different from those of common drug pairs [2]. [3] presented a method of knowledge discovery on compatibility laws of traditional Chinese medicine prescription for stroke disease based on biclustering method. Results prove that this method can be well applied for prescription-herb knowledge discovery in traditional Chinese medicine prescription databases. The prescription-herb information can be effectively extracted, and laws of prescription-herb and syndrome-treatment can be mined. According to the characteristics of Chinese medicine, the demand for drugs group and Frequent Pattern-growth algorithms, Han et al. proposed the traditional Chinese medicine-Frequent Pattern (FP) tree [4]. Wang et al. evaluated the efficacy of three categorized formulas for tonifying Shen yang, i.e., Shenqi Pill, Yougui Pill, and Yougui Drink based on rough set, and explored

the law of compatibility between the core herbs and the edge herbs as well as the law of compatibility between the yang-tonifying herbs and the yin-tonifying herbs in the core herbs [5]. [6] presented the Joint Primary Medicine Extraction Algorithm which combines the syndrome factors to analysis and the similarity of “prescription name-medicine name”.

Formal concept analysis is a branch of a mathematical concept and concept hierarchy based on the application of lattice theory, which was put forward in 1982 by Professor Ganter and Wille [7]. It is a powerful tool for data mining and visualization. Formal concept analysis has its roots in the field of mathematics; however, it has changed its orientation towards the computer science area [8]. Formal concept analysis has become very attractive as a knowledge representation method for computer science, providing a support for the field of Conceptual Knowledge Processing [9]. The basic idea is to establish a conceptual hierarchy based on the relationship between objects and attributes, and generate the concept lattice from the data set. Concept lattice is the core data structure of formal concept analysis. It describes the relationship between objects and features. According to this principle, Hong et al. put forward the method of generating multi-level complex concept network based on attribute partial order structure [10]. It is attribute partial ordered structure diagram. Song et al. mined knowledge of traditional Chinese medicine by this method [11-13]. In this paper, the knowledge of prescription compatibility is mined by the visualization method based on attribute partial ordered structure diagram. This method provides a new idea for the prescription compatibility research.

The paper is organized as follows. Section 2 introduces the theory of formal concept analysis and related concepts. Section 3 generates the attribute partial ordered structure diagram according to the formal context based on the data of *Xiaoqinglong* decoction, and discovers prescription compatibility knowledge from the attribute partial ordered structure diagram. Section 4 presents the conclusions and directions for future research.

2. Theory of Attribute Partial Ordered Structure Diagram. Formal concept analysis is described as a set structure $K = (U, M, I)$ where U is a set of objects, M is a set of attributes and I is a relation between U and M . In order to express the relationship I between object g and attribute m , we write gIm or $(g, m) \in I$. A formal concept of the formal context K is a pair (A, B) with $A \subseteq U$, $B \subseteq M$, $A = B'$, and $B = A'$. A' is the set of attributes common to the objects in A ; and B' is the set of objects which have all attributes in B . The sets A and B are called the extent and the intent of the formal concept (A, B) . A formal context can be represented by a cross table, where the rows represent the object names and columns represent the attribute names. Table 1 shows an example of the formal context.

A formal context consists of two sets $U = \{u_1, u_2, \dots, u_n\}$ and $M = \{m_1, m_2, \dots, m_k\}$ and a relation I between U and M . The elements of U are called the objects and the elements of M are called the attributes of the context. Suppose $m \in M$, and we define $f(m) = \{g \in U | (g, m) \in I\}$ which denotes the object set of attribute m ; meanwhile, suppose $g \in U$, and we define $g(u) = \{m \in M | (g, m) \in I\}$ which denotes the attributes set of object u . The concept lattice according to the formal context in Table 1 is shown in Figure 1. Figure 2 shows the corresponding attribute partial ordered structure diagram by the method of constructing in literature [10].

According to Figure 2, attribute partial ordered structure diagram is a closed tree structure. There is only one node at the top and bottom in the graph, and the line direction is one-way, which cannot return. The path is called from one node to another node. Any node must be able to find a path, which is from the top node to the bottom node. The top node a is called the maximum common attribute of the all attributes. The

TABLE 1. An example of the formal context

objects/attributes	a	b	c	d	e	f	g
1	x	x	x	x			x
2	x	x	x	x	x		x
3	x	x	x				x
4	x	x	x	x			x
5	x				x	x	
6	x	x		x	x	x	
7	x	x	x	x	x	x	

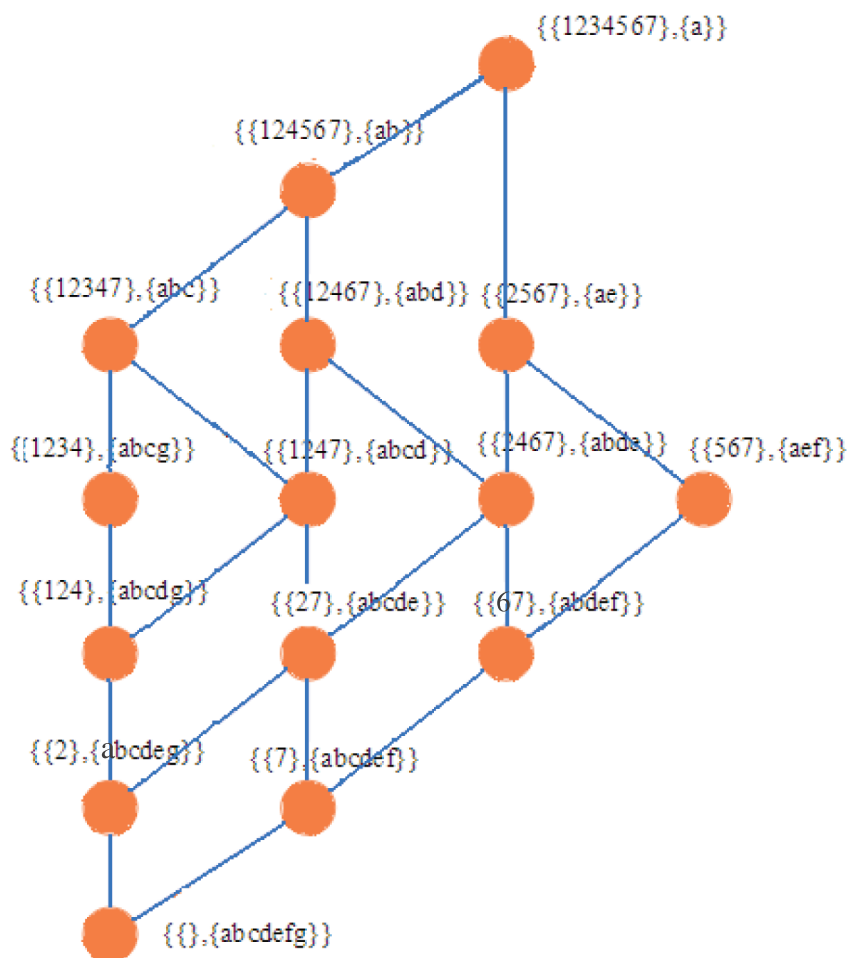


FIGURE 1. The concept lattice of Table 1

node *b* of the second layer is called the common attribute, and the node *e* is called the unique attribute.

3. Case Analysis. We establish the formal context based on the data of *Xiaoqinglong* decoction, which was chosen from the classic medical books of Zhang zhongjing prescription. In the formal context, the objects are names of all the prescription or syndrome of this book, and drugs are regarded as the attributes. The tool for data storage is Excel 2007. Then we can generate the attribute partial ordered structure diagram according to the formal context. Figure 3 shows the attribute partial ordered structure diagram between prescription and drug. The attribute partial ordered structure diagram between syndrome and drug is shown in Figure 4. The attribute partial ordered structure diagram generated from optimized context can be used to discover knowledge.

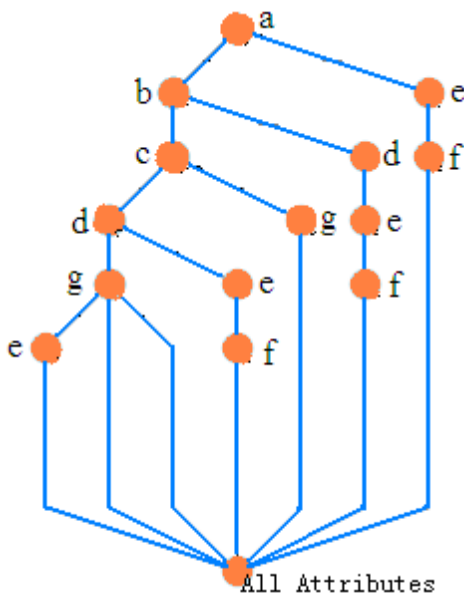


FIGURE 2. The attribute partial ordered structure diagram of Table 1

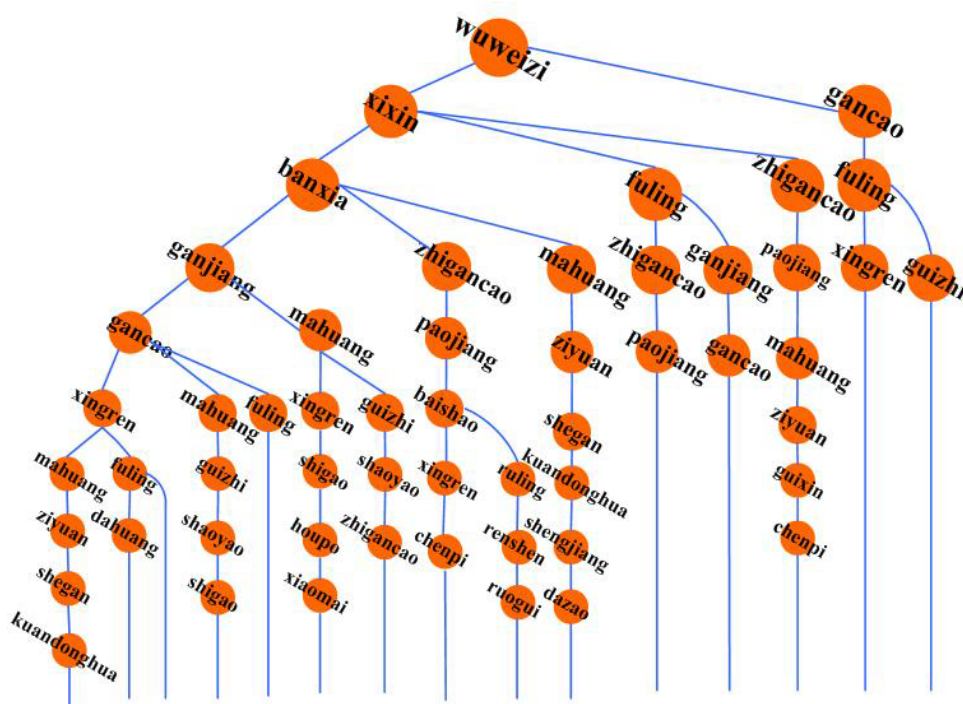


FIGURE 3. The attribute partial ordered structure diagram between prescription and drug

3.1. Analysis of attribute partial ordered structure diagram between prescription and drug.

3.1.1. *Analysis from the hierarchy.* According to different hierarchies and covered branches, we can find the core drug, basic drug and drug group in the attribute partial ordered structure diagram.

(1) The maximum common attribute is the core drug: in Figure 3, $\{Wuweizi\}$ is the first layer of 14 layers. According to the maximum common attribute definition, each of the prescription contains *wuweizi*. That is to say, *wuweizi* is the core drug of *Xiaoqinglong* decoction prescription.

(2) The common attribute is the basic drug: in the attribute partial ordered structure diagram, high frequency properties appear in the high level, so the drug at the higher level is the basic drug. $\{Xixin\}$, which is the second layer of 14 layers in Figure 3, appeared in the 01-013 prescription. There are many times in the prescription of *xixin*, so *xixin* is the basic drug. Similarly, $\{Banxia\}$, which is less than *xixin*, appears in the third layer, and 01-010 prescription contains *banxia*. So *banxia* is also the basic drug.

(3) The associated attribute is the drug group: in the attribute partial ordered structure diagram, attribute which is associated relationship must be in the same branch. In Figure 3, $\{Wuweizi\}$ - $\{Xixin\}$, $\{Wuweizi\}$ - $\{Banxia\}$, $\{Wuweizi\}$ - $\{Ganjiang\}$ are common drug groups of *Xiaoqinglong* decoction prescription.

3.1.2. *Analysis from the different cluster.* The similar objects appear in the same cluster in the attribute partial ordered structure diagram, and the smaller clusters are, the less covered branch is.

In Figure 3, there are five clusters: vertices of the first clusters are $\{Wuweizi, Xixin, Banxia, Ganjiang, Gancao\}$. This cluster contains 01-05 prescriptions. *Fuling gancao wuweizi* adding *shengjiang xixin banxia xingren dahuang* decoction, *bufei* decoction and *guling wuwei gancao* decoction subtracting *guizhi* adding *ganjiang xixin banxia* decoction can dispel cold, drinking, phlegm and swelling effects. *Xiaoqinglong* adding *shigao* decoction and *guling wuwei gancao* decoction subtracting *guizhi* adding *ganjiang xixin banxia* decoction can dispel drinking and remove heat effect.

Vertices of the second clusters are $\{wuweizi, Xixin, Banxia, Ganjiang, Mahuang\}$. In this cluster, there are 2 prescriptions, and they are all used when cold pathogenic and drinking pathogenic hurt the lung.

Vertices of the third clusters are $\{wuweizi, Xixin, Banxia, Zhigancao, Paojiang, Baishao\}$. In this cluster, there are 2 prescriptions. They are used when drink stopped because of Qi weak. Warm lung decoction can treat cough and asthma, and *xingzi* decoction can treat hemoptysis.

Vertices of the fourth clusters are $\{wuweizi, Xixin, Fuling\}$. In this cluster, there are 2 prescriptions. There are the common drugs of *wuweizi, Xixin, Fuling* in 2 prescriptions. *Wenfei* power also includes *zhigancao* and *paojiang*, and *fuling gancao wuweizi ganjiang xixin* decoction also includes *gancao* and *ganjiang*. *Ganjiang* and *paojiang* can warm lung and reduce watery phlegm. However, *ganjiang* has more heat than *paojiang*. *Gancao* has antitussive and expectorant effect, and *gancao* can *buzhongyiqi*. That is to say, *fuling gancao wuweizi ganjiang xixin* decoction has stronger effect.

Vertices of the last clusters are $\{wuweizi, gancao, fuling\}$. There are 2 prescriptions: *xingren wuweizi* decoction and *guizhi fuling wuwei gancao* decoction. *Wuweizi, gancao, fuling* are the common drugs of 2 prescriptions. *Xingren wuweizi* decoction also includes *xingren*, and *guizhi fuling wuwei gancao* decoction includes *guizhi*. The former has *jiangqi* effect, and the latter can dispel cough.

According to the above analysis, there is the similar treatment efficacy in the same cluster. However, prescription which has different effects appears in the same cluster. For example, 2 prescriptions are in the second cluster.

3.2. Analysis of attribute partial ordered structure diagram (between syndrome and drug). Figure 4 is the attribute partial ordered structure diagram between syndrome and drug. In the formal context of Figure 4, the syndromes are regarded as objects and the drugs contained in the prescriptions are attributes. Based on this formal context and the former theory [10], Figure 4 is generated.

From the top layer of the diagram, we can see that $\{wuweizi\}$, $\{xixin\}$ and $\{ganjiang\}$ are the basic drugs of *Xiaoqinglong* decoction prescription.

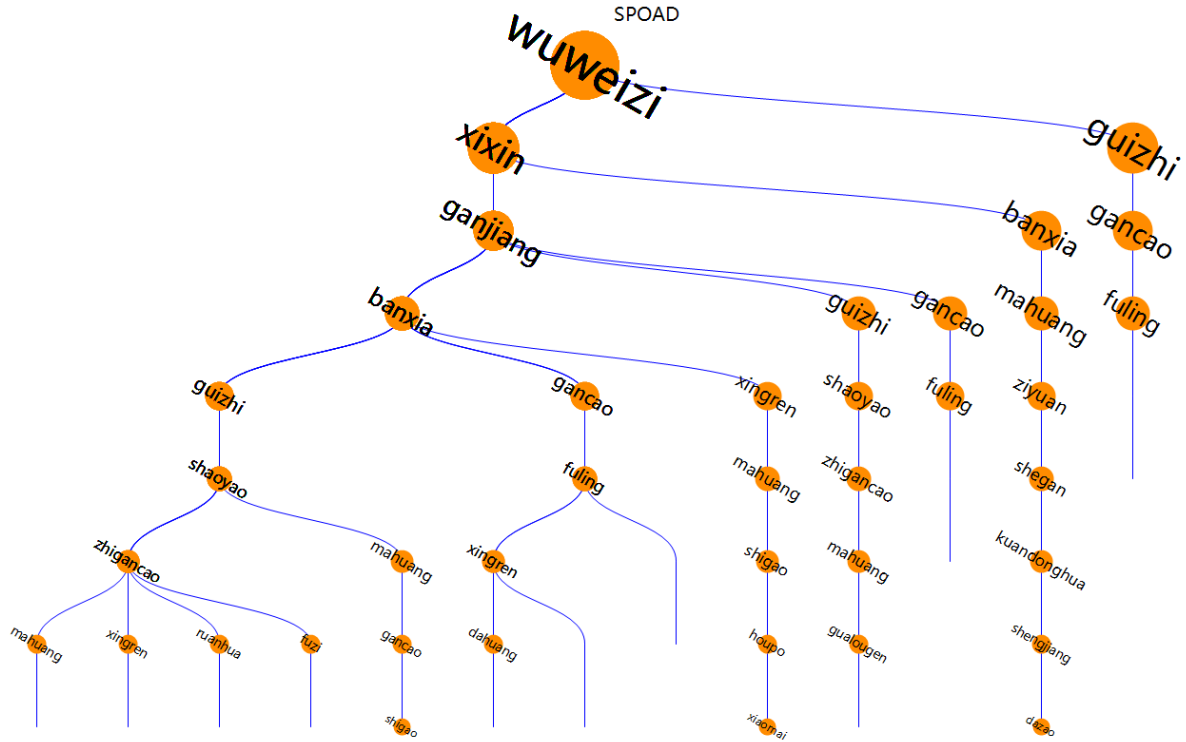


FIGURE 4. The attribute partial ordered structure diagram between syndrome and drug

There are two clusters in Figure 4: $\{wuweizi, xixin, ganjiang, zhigancao, banxia, guizhi, shaoyao\}$ and $\{wuweizi, xixin, ganjiang, zhigancao, banxia, fuling\}$. $\{wuweizi, xixin, ganjiang, zhigancao, banxia, guizhi, shaoyao, mahuang\}$ is the sub cluster of the first cluster. All the drugs of this sub cluster compose *Xiaoqinglong* decoction. It is the second branch. The pathogenesis for use of *Xiaoqinglong* decoction pattern is “exterior cold attack, cold rheum collecting in the lung, or exterior cold with interior fluid retention”. The characteristic symptoms treated by *Xiaoqinglong* decoction should be cough, cough productive of white thin sputum, shortness of breath, chest tightness, runny nose, and the secondary symptoms include diarrhea, sneezing, chills, fever, nasal congestion, tiredness, weakness, etc. If pathologic drinking may be transformed to pyretic toxicity, then *Xiaoqinglong* decoction adds *shigao*, which can clear away heat. This is the *Xiaoqinglong* adding *shigao* decoction. The book of Synopsis of Golden Chamber records that *Xiaoqinglong* adding *shigao* decoction was for the treatment of cold fluid and gloomy fever syndrome. If there is the symptom with difficulty in micturition, then *Xiaoqinglong* decoction that subtracts *mahuang* and adds *xingren* can be used. If there is the symptom with diarrhea, then *Xiaoqinglong* decoction subtracting *mahuang* and adding *ruanhua* is used. If there is the symptom with yege, then *Xiaoqinglong* decoction subtracting *mahuang* and adding *fuzi* can be used.

Vertices of the second clusters are $\{wuweizi, xixin, ganjiang, zhigancao, banxia, fuling\}$. It includes 06-08 braches. *Xiaoqinglong* decoction that subtracts *mahuang*, *shaoyao* and *guizhi*, then adds *fuling* is the composition of prescription in 08 brach. This prescription can treat giddy and vomiting. If there are the symptoms with edema and the face of fever, then adding *xingren* and *dahuang* can be used.

In addition, prescription for treating other branches syndromes is composed of the drugs which add or subtract drug based on the *Xiaoqinglong* decoction. The prescription, which subtracts *mahuang*, *shaoyao*, *guizhi* and *banxia*, then adds *fuling*, can be used to

treat this symptom in case of the symptom of diaphragm retention of fluid relapse. If there is the symptom with thirsty, the prescription which subtracts *banxia*, then adds *gualougen* can be used. The prescription, which subtracts *shaoyao*, *guizhi* and *zhigancao*, then adds *xingren*, *shigao*, *houpo*, *xiaomai*, can be used if there is the symptom with chest congestion. If the symptom of wheezing sound appears, the prescription, which subtracts *shaoyao*, *ganjiang*, *guizhi* and *zhigancao*, then adds *shegan*, *shengjiang*, *ziyuan*, *kuandonghua*, *dazao*, can be used. In order to express this ideas as clearly as possible, the diagram of drug addition and subtraction is shown in Figure 5.

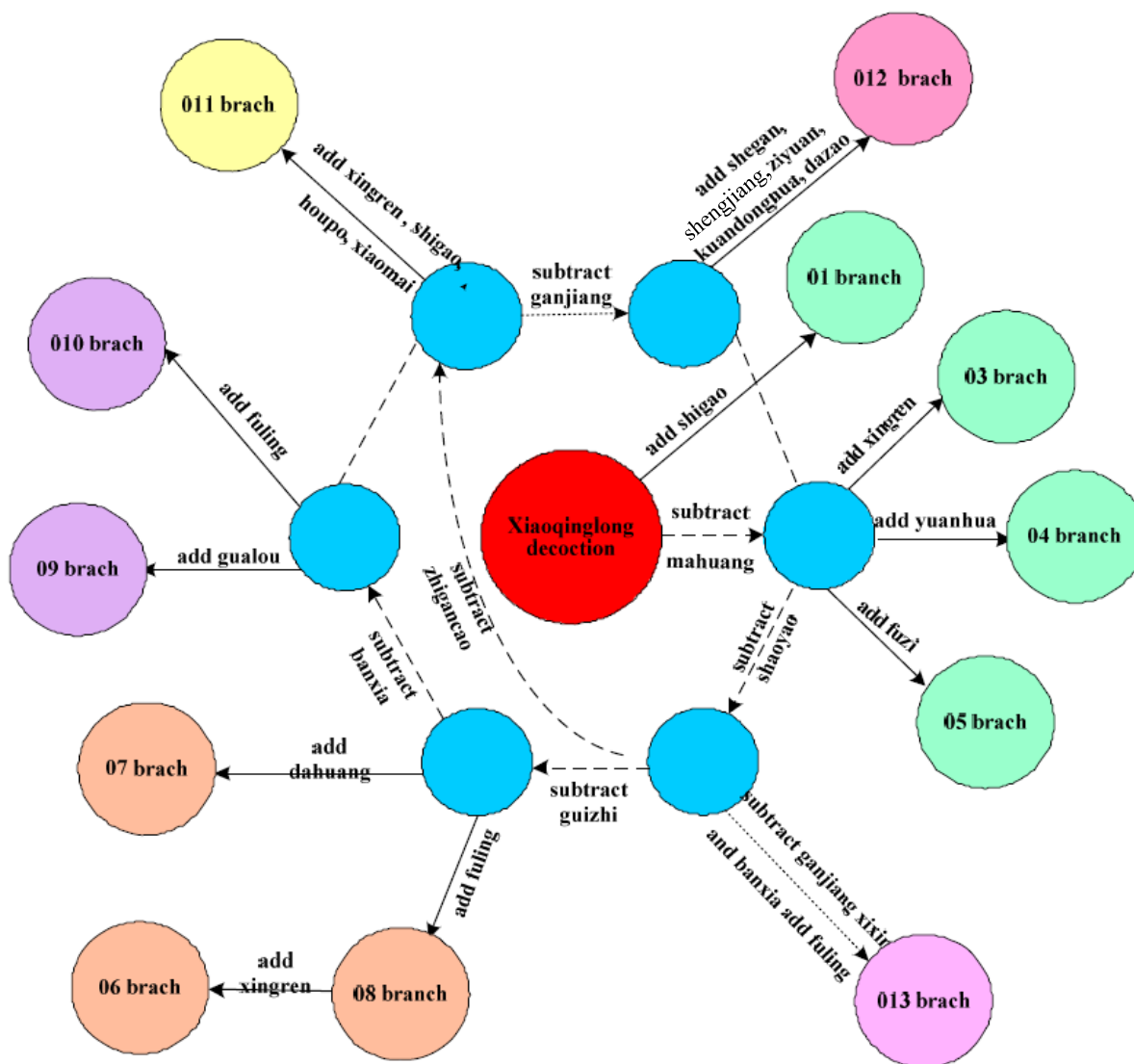


FIGURE 5. The diagram of drug addition and subtraction

4. **Conclusions.** In this study, we can find that attribute partial ordered structure diagram is a clear tree structure, and there is no cross lines in the diagram. It can express the relationship between objects and properties, category meaning is clear and hierarchy relationship is distinct. So it works well in extracting rules for the prescription. With the help of the attribute partial ordered structure diagram, we can easily find the group structure of compatibility between prescription and drug, syndrome and drug. So attribute partial ordered structure diagram should be a good tool for data mining in traditional Chinese medicine. We can make further research in the fields of traditional Chinese medicine, such as the relationship between the syndrome and disease, drug and drug properties, drug and channel tropism.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grant No. 61201111, 61273019, 81373767 and 81273740. It is also supported by the Natural Science Foundation of Hebei Province under Grant F2013203368. The authors gratefully acknowledge the supports.

REFERENCES

- [1] M. Fan and X. Meng, *Data Mining: Concepts and Techniques*, Machinery Industry Publishing, Beijing, 2001.
- [2] E. Shang, X. Fan, J. Duan and L. Ye, Data mining study on incompatibility characters of Chinese herbal medicine in accordance with association rules, *Journal of Nanjing TCM University*, vol.26, no.6, pp.421-424, 2010.
- [3] L. Wu and S. Li, Knowledge discovery on compatibility laws of TCM prescription for stroke disease based on biclustering method, *Chinese Journal of Information on TCM*, vol.20, no.11, pp.16-19, 2013.
- [4] L. Han, S. Ju and J. Zhu, Research on the mining of inter-dimensional maximal frequent itemsets algorithm of TCM-FP tree, *Journal of Jiangnan University (Natural Science Edition)*, vol.9, no.2, pp.185-190, 2010.
- [5] X. Wang, L. Yi, L. Zhang, L. Jin and L. Luo, Study on the law of compatibility in three categorized formulas for Tonifying Shen Yang: Shenqi Pill, Yougui Pill and Yougui Drink based on rough set, *Chinese Journal of Integrated Traditional and Western Medicine*, vol.33, no.1, pp.114-118, 2013.
- [6] X. Qin, J. Xiong, Y. Wang, Z. Yu, X. Zhang, S. Zou and Y. Jiang, Integrating syndrome factor analysis and "Prescription Name-Medicine Name" similarity to mine primary medicines in traditional Chinese medicine, *Journal of Sichuan University (Natural Science Edition)*, vol.48, no.1, pp.67-72, 2011.
- [7] B. Ganter and R. Wille, *Formal Concept Analysis, Mathematical Foundations*, Springer Publishers, New York, 1999.
- [8] G. Stumme, Formal concept analysis on its way from mathematics to computer science, *Proc. of the 10th International Conference on Conceptual Structures: Integration and Interfaces*, London, UK, pp.2-19, 2002.
- [9] J. Hereth, G. Stumme, R. Wille and U. Wille, Conceptual knowledge discovery and data analysis, *Proc. of the 8th International Conference on Conceptual Structures*, Darmstadt, Germany, pp.421-437, 2000.
- [10] W. Hong, S. Li, J. Yu and J. Song, A new approach of generation of structural partial-ordered attribute diagram, *ICIC Express Letters, Part B: Applications*, vol.3, no.4, pp.823-830, 2012.
- [11] W. Hong, J. Yu, F. Cai and J. Song, A new method of attribute reduction for decision formal context, *ICIC Express Letters, Part B: Applications*, vol.3, no.5, pp.1061-1068, 2012.
- [12] J. Song, H. Zhao, J. Yu, W. Hong and S. Li, Discovery of selection rules for acupuncture points of digestive system by structural partial-ordered attribute diagram, *ICIC Express Letters, Part B: Applications*, vol.3, no.5, pp.1165-1170, 2012.
- [13] J. Song, J. Yu, E. Yan, Z. Zhang and W. Hong, Syndrome differentiation of six meridians for warm disease based on structural partial-ordered attribute diagram, *ICIC Express Letters*, vol.7, no.3(B), pp.947-952, 2013.