

Knowledge Discovery of Prescription for Spleen Deficiency Syndrome Base on Attribute Partial Ordered Structure Diagram

Hui Meng, Wenxue Hong, Cunfang Zheng, Fengjie Fan, Jialin Song

Institute of Electrical Engineering, Yanshan University, China

*menghui_ysu@163.com

Keywords: knowledge discovery; attribute partial ordered structure diagram (APOSD); spleen deficiency syndrome; prescription compatibility

Abstract. In recent years, the application of knowledge discovery method to the field of the traditional Chinese medicine (TCM) gains more attention to researchers. It is necessary to have a visual mining for hidden knowledge in the massive and complicated data of prescription, which embody important knowledge of treatment for different syndrome. The spleen deficiency syndrome is one of the common syndromes of TCM, but there have many problems needed to be discovered and perfected further. In this paper, we present a visual analysis method named attribute partial ordered structure diagram (APOSD) for mining knowledge of prescription compatibility law for spleen deficiency syndrome. A formal context is formed based on prescriptions and drugs recorded in literature. Then it is optimized and computed to construct APOSD for visualization. The useful knowledge can be mined from the APOSD. The result shows that APOSD can discover new knowledge of prescription compatibility.

1 Introduction

Knowledge discovery in databases (KDD) is a process that extracts novel, useful and understandable knowledge from large numbers of data. As a promising technique, KDD has been increasingly taken in traditional Chinese prescription (TCP) researchers [1, 2]. Formal concept analysis (FCA) is a powerful tool in data analysis and visualization, and has been applied to data mining, knowledge discovery and many other fields since proposed. Moriki and Yoshida proposed a document clustering algorithm based on FCA [3]. Yashinaga and Nobuhara used concept lattice to classify web pages [4]. The various applications have mainly focused on the use of concept lattice that can explain the relationship among the concepts in a formal context. However, in the concept lattice, the complex relations between concepts make the lines rather complicated and crossed. Aiming to delineate the relations among attributes and distinguish distinctive objects, the attribute partial ordered structure diagram (APOSD) is proposed, and its construction method is described [5]. APOSD is a closed acyclic tree-similar structured diagram. All the paths starting from the top node and ending at the bottom node are one-way ones.

The spleen deficiency syndrome is one of the common syndromes of TCM. The treatment of spleen deficiency syndrome is recorded as prescriptions. Prescription is the combination of various drugs. In this paper, we apply APOSD for knowledge discovery on prescription of spleen deficiency syndrome. Knowledge of the original data can be found from structural relation of data features in aspects of cluster sets, cluster subsets, edges and nodes in APOSD. According to the analysis of APOSD, the primary drugs, the medicine groups and their compatibility law can be mined.

2 Method

2.1 Formal Context. A formal context can be described as a set structure $K=(U, M, I)$, where $U=\{u_1, u_2, \dots, u_n\}$ is a set of n elements, called objects; $M=\{m_1, m_2, \dots, m_k\}$ is a set of k elements, called attributes; and I is a relation between U and M , $I \subseteq U \times M$. A formal context can be represented by a cross table, where the rows indicate the objects and the columns indicate the attributes. If an object has an attribute, the intersected cell can be labeled with a symbol "×". In order to satisfy with the computer

process, a formal context needs to be standardized in accordance with certain rules. All the objects and attributes are replaced by serial labels. All the symbol "×" are converted to 1 and others to 0 to form the binary relation.

2.2 Formal Context Optimization. If m is an attribute in a formal context K , then the objects set corresponding to m is $U=g(m)$.

Definition 2.1. The number of objects those corresponding to an attributes is defined as degree of the attribute, denoted by D_m , $D_m = |g(m)|$.

Definition 2.2. Formal context $K=(U, M, I)$ and $K_1=(U_1, M_1, I_1)$, if the two formal contexts meet: $U_1 \subseteq U, M_1 \subseteq M, I_1 \subseteq I$, then we claimed that K_1 is a sub-formal context of K , denoted by $K_1 \subseteq K$.

For an original formal context, the procedure of optimization is as follows:

Step 1. Sort attribute columns in descending order according to the degrees of attributes. The attribute column which covers the most objects is put to the forefront (the left most) of the formal context, and the attribute which covers least of objects is put to final end (the right most) of the formal context.

Step 2. Do all-dimensional sorting operation. This is to cluster objects in sub-formal context under some constraints. Then make all the 1 in every column distributed continuously under the conditions of the previous attribute.

2.3 Construction Algorithm of APOSD. APOSD is a visual representation of the optimized formal context. The definition of the maximum common attribute, common attribute and unique attribute refer to literature [6]. Then we can construct APOSD as follows:

Step 1. Compute the maximum common attribute M_0 , the corresponding node denoted as A_0 .

Step 2. Compute the set of common attributes $\{m_{11}, m_{12}, \dots, m_{1p}\}$ and the set of unique attributes $\{s_{11}, s_{12}, \dots, s_{1q}\}$ under the maximum common attribute, their corresponding nodes denoted as $\{A_{m1}, A_{m2}, \dots, A_{mp}\}$ and $\{A_{s1}, A_{s2}, \dots, A_{sq}\}$. Then draw a line connected A_0 with common attribute or unique attribute respectively.

If the nodes in this layer satisfied: $\bigcup_{i=1}^p g(m_{1i}) \cup \bigcup_{i=1}^q g(s_{1i}) = U$, then go to step 3. Otherwise, there are k edges connected A_0 with the bottom node.

Step 3. Compute the set of common attributes $\{m_{t+1,1}, m_{t+1,2}, \dots, m_{t+1,l}\}$ and the set of unique attributes $\{s_{t+1,1}, s_{t+1,2}, \dots, s_{t+1,k}\}$ under each common attribute m_{ti} , $t \in (1: p)$ in upper layer, their corresponding nodes denoted as $\{A_{mt+1,1}, A_{mt+1,2}, \dots, A_{mt+1,l}\}$ and $\{A_{st+1,1}, A_{st+1,2}, \dots, A_{st+1,k}\}$. Compute the set of unique attributes $\{sc_{t+1,1}, sc_{t+1,2}, \dots, sc_{t+1,q}\}$ under each unique attribute in upper layer, their corresponding nodes denoted as $\{As_{t+1,1}, As_{t+1,2}, \dots, As_{t+1,q}\}$.

Step 4. Stop when there are no common attributes and unique attributes under the node of upper layer. Otherwise, go to step 3.

2.4 Knowledge Discovery in APOSD. According to the relationship between data, APOSD present a hierarchy structure expressing the degree of attributes covered objects. Attributes cover most objects locate in the upper layer while objects cover most attributes locate in the lower layer. The higher the nodes locate, the more universal they are. On the contrary, the lower the nodes locate, the more individual they are. The most common analysis method is top-down model, which also named as concept driven model. The nodes in upper layer have several branches in lower layer. All the nodes are connected in accordance with the hierarchy and the correlation so as to form a partial order structure relationship.

A certain number of edges that express the common attributes of the objects are congregated into a cluster set, which can be divided into cluster subsets gradually. The cluster subsets can express local structure relationship of the characteristics of some objects. The similarity of objects can be found in a cluster subset and different subsets express different structure. Each edge is an expression of an object and is regarded as a pattern distinguishing different objects. Each node represents an attribute. By the relation between the nodes in the diagram, one can see the structural relationship among the attributes. A top-down edge expresses all attributes contained in an object.

3 Application

In order to carry out APOSD on compatibility of prescription, 99 prescriptions in General Records of Holy Universal Relief for spleen asthenia syndrome are selected as the research samples. According to Dictionary of Chinese Traditional Medicine, the name of drugs, the name of efficacy and dosage are standardized, quantified and conversed, becoming the scientific and analyzable data. Then an Excel datasheet is established to deposit data. The rows indicate prescription names and the columns are drugs contained in all prescriptions. If a drug is used in a prescription, the intersected cell is assigned to 1 and others to 0. By standardization, the formal context is formed. Based on the optimized formal context followed methods mentioned above, we can construct APOSD as shown in Figure 1.

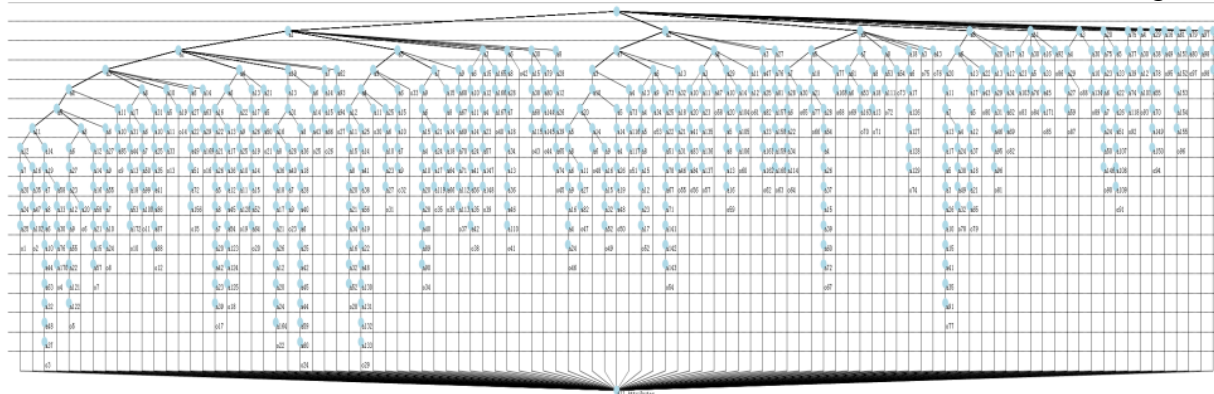


Figure 1. APOSD of 99 prescriptions for spleen asthenia syndrome

The diagram from top to bottom is divided into 19 layers. The first layer is $\{\phi\}$, which means that there is no drug contained in all prescriptions. There are 15 nodes in the second layer. The nodes located on the left are most commonly used drugs in the selected prescriptions and they have more sub-nodes. Several top nodes on the left most edge formulate primary drugs, that are chenpi, gancao, renshen, baishu and houpu. Chenpi can regulate qi, dispel dampness and resolve phlegm. Gancao can tonify spleen for nourishing qi and relieve spasm and pain. It can also mediate property of various drugs. Renshen has the effect of tonifying spleen for nourishing lung, promoting fluid production to quench thirst and calming the nerves for nootropic effect. Baishu is an important drug for reinforcing qi to invigorate the spleen and also can dispel dampness. Houpu can promote qi circulation to relieve flatulence, calm the adverse-rising energy and stop cough. These five kinds of drugs are compatible, which can tonify spleen, regulate qi and dispel stagnation.

From the perspective of cluster set, there are three main clusters in Figure 1. They are the cluster set followed $\{a1\}=\{\text{chenpi}\}$, the cluster set followed $\{a2\}=\{\text{gancao}\}$ and the cluster set followed $\{a9\}=\{\text{fuzi}\}$. The cluster set followed $\{a1\}=\{\text{chenpi}\}$ is shown in Figure 2, which can be subdivided into four subsets. The first subset followed $\{a1, a2, a3, a5, a4\}=\{\text{chenpi, gancao, renshen, baishu, houpu}\}$ comprise six prescriptions from $\{P1\}$ to $\{P6\}$ named as $\{\text{hewei pill}\}$, $\{\text{baishuhuoxiang decoction}\}$, $\{\text{baidoukou pill}\}$, $\{\text{huoxiang decoction}\}$, $\{\text{qingmuxiang pill}\}$ and $\{\text{shenpu decoction}\}$. The main therapy method in this cluster subset is regulating qi and also has the effect of reinforcing qi and dispelling dampness. The second subset followed $\{a1, a2, a3, a8\}=\{\text{chenpi, gancao, renshen, muxiang}\}$ comprise three prescriptions from $\{P8\}$ to $\{P10\}$ named as $\{\text{erju pulveres}\}$, $\{\text{baidoukou pulveres}\}$ and $\{\text{renshenfuling decoction}\}$. Muxiang can promote qi circulation and evacuate stagnancy. The main therapy method in this cluster subset is regulating qi and reinforcing qi. The third subset followed $\{a1, a2, a5\}=\{\text{chenpi, gancao, baishu}\}$ comprise five prescriptions from $\{P15\}$ to $\{P19\}$ named as $\{\text{chenjupi pill}\}$, $\{\text{jinye pill}\}$, $\{\text{fuzi decoction}\}$, $\{\text{yixiao pulveres}\}$ and $\{\text{baidoukou pulveres}\}$. The forth subset followed $\{a1, a2, a19\}=\{\text{chenpi, gancao, gaoliangjiang}\}$ comprise four prescriptions from $\{P20\}$ to $\{P23\}$ named as $\{\text{wenqi pulveres}\}$, $\{\text{dingxiang decoction}\}$, $\{\text{muxiang pill}\}$ and $\{\text{doukou decoction}\}$. Gaoliangjiang can relieve pain, which compatibility with chenpi and gancao has the effect of reinforcing qi and invigorating the spleen. The main therapy method in this cluster set is regulating qi.

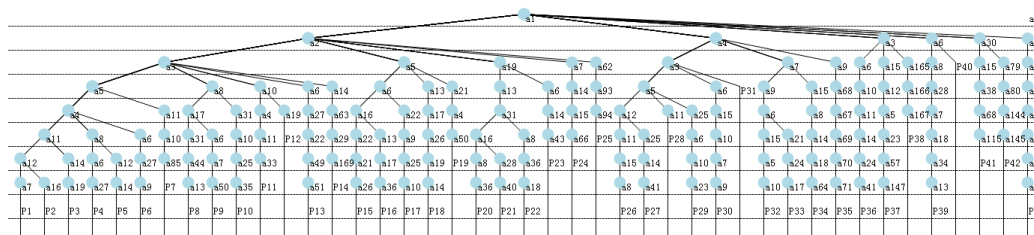


Figure 2. The cluster set followed $\{a1\}=\{\text{chenpi}\}$

The main therapy method in cluster set followed $\{a2\}=\{\text{gancao}\}$ is reinforcing qi. The main therapy method in cluster set followed $\{a9\}=\{\text{fuzi}\}$ is warming yang. So we can conclude that the formula of compatibility for spleen deficiency syndrome is regulating qi, reinforcing qi and warming yang. Based on the theory of TCM, the spleen is the source of qi and blood. The compatibility of qi-regulating medicine and qi-reinforcing medicine can tonify spleen as meanwhile dispel stagnation. Qi and yang are interdependent in pathology. Assisting drugs for warming yang have better curative effect.

4 Conclusions

In this paper, we presented an efficient method for the visualization and analysis of prescription compatibility of spleen deficiency syndrome. The proposed method of APOSD based on optimization of formal context can reflect the characteristics of the drugs, and show the relationships between the prescriptions and the drugs, the relationships among the drugs clearly. We have described the steps of formal context optimization and the construction algorithm of APOSD. With a strict structure and hierarchical division, the APOSD is well suited for knowledge discovery of prescription compatibility. Its distinct hierarchy, clear structure, uncrossed edges provide a better visualization. The generation of the APOSD requires simple computation that makes a large potential in allusion to big data.

Acknowledgment

This work is supported by National Science Foundation of China (NSFC) under Grant No. 61501397 and 61273019.

References

- [1] She K.K., Hu K.F., Wang Z.. Research on Chinese prescription compatibility based on variable precision tolerance model and attribute sensitivity reduction. World Science and Technology-Modernization of Traditional Chinese Medicine, 2014, 16(6):1222-1228.
- [2] Ma J.G., Hu Z.S., Cao H., et al. Preliminary study of prescription compatibility law based on association rule mining. Chinese Journal of Experimental Traditional Medical Formulae, 2013, 19(7):351-353.
- [3] Moriki A., Yoshida S.. Order-based clustering using formal concept analysis. Proc. of World Automation Congress, Kobe, Japan, 2010: 376-388.
- [4] Yashinaga N., Nobuhara H.. Formal concept analysis based web pages classification/visualization and their application to information retrieval. Proc. of the 10th International Symposium on Communications and Information Technologies, Tokyo, Japan, 2010: 153-157.
- [5] Hong W.X., Li S.X., Yu J.P., et al. A new approach of generation of structural partial-ordered attribute diagram. ICIC Express Letters, Part B: Applications ICIC International, 2012, 3(4): 823-830.
- [6] Hong W.X., Li S.X., Zhang T., et al. Generation principle of partial ordered structure towards big data. Journal of Yanshan University, 2014, 38(5):388-393.