ICIC Express Letters
Volume **10**, Number **5**, May **2016**

 ISSN 1881-803X
pp. 1181–1186

PARTIAL ORDERED STRUCTURE RADIAL TREE: A NEW METHOD FOR BIG DATA VISUALIZATION

Wenxue Hong[1,3,*], Fang Sun[1], Shaoxiong Li[1]
Jialin Song[1] and Cunfang Zheng[1,2]

[1]School of Electrical Engineering
[2]Liren College
Yanshan University
No. 438, West Hebei Ave., Qinhuangdao 066004, P. R. China
[*]Corresponding author: hongwx@ysu.edu.cn

[3]Big Data Visualization Technology Center
Northeastern University at Qinhuangdao
No. 143, Taishan Road, Qinhuangdao 066004, P. R. China

Received October 2015; accepted January 2016

Abstract. *Partial ordered structure diagram has been widely applied to knowledge discovery, data mining and many other domains due to its better visualization effect. However, when researching and analyzing a large-scale formal context, some shortcomings of early partial ordered structure diagram have led to indistinct visualization results. Through improving from the early diagram, a new visualization method called partial ordered structure radial tree is proposed. Applying this method, a partial ordered structure radial tree is mapped and compared with early diagram in this paper. The comparison results show that this new approach can surely solve some key problems when dealing with big data, it is clearer and more intuitive than early diagram, and visualization effect is better; thereby the new diagram can show important information of large-scale formal context more effectively. The raise of partial ordered structure radial tree provides a useful way for knowledge visualization, and it is of great significance to knowledge discovery and data mining of big data.*

Keywords: Visualization, Big data, Partial ordered structure, Knowledge discovery

1. **Introduction.** With development of Internet advance of computer and modern sensor technology, a data-driven era in which abundant data is continuously generated and acquired for varieties of purposes has come. Therefore, the ability to analyze data and make timely decisions based on available data is very crucial. How to use and present data effectively, finding useful information and knowledge from mass data, is one of the major problems that need to be solved in this big data era [1].

Human eye is a high-bandwidth parallel processor which can process a huge amount of visual signal, it has strong pattern recognition abilities, and the speed of perceiving visual symbols is much faster than perceiving numbers or texts. Therefore, the main way of presenting data is data visualization [2]. Using high flux characteristic of human visual perception, visualization presents internal laws and potential knowledge of information in a graphical view, and it offers people an essential way to understand complex phenomena and interpret complex data. At present it is necessary to do some research about visualization and develop visual analysis software autonomously in China.

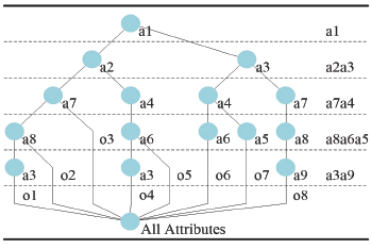
Partial Ordered Structure Diagram (POSD) is a knowledge visualization method proposed by Professor Hong and his research team in 2012, and it is one powerful visualization tool developed on the basis of the theory of partially order [3]. Partial Ordered Structure Diagram includes attribute partial ordered structure diagram and object partial ordered

structure diagram. Object partial ordered structure diagram is generated after transposing the formal context matrix, it applies the same algorithm as attribute partial ordered structure diagram, hence this article used attribute partial ordered structure diagram as an example for discussion, and the conclusion equally applies to object partial ordered structure diagram. Partial ordered structure diagram not only has a good visualization effect, but can achieve the cluster of hierarchical attributes. It is conducive to analyze concept constitution. Now partial ordered structure diagram has been widely used in many fields and all the researches have received gratifying results.

However, when facing formal contexts formed from big data, the readability of partial ordered structure diagram becomes lower and lower as the scale of formal context increases. In order to solve this problem, a method called partial ordered structure circular diagram [4] has been put forward by some researchers not long ago, but it also has its own limitations; therefore, more reasonable solutions need to be proposed. This article firstly introduces the theory of Partial Ordered Structure Diagram (POSD), then points out some shortcomings and challenges of this method when researching big data, and lastly puts forward a new visualization method – partial ordered structure radial tree. Through partial ordered structure radial tree, potential correlation of complex big data can be clearly presented, and valuable information can be extracted and used for further study. The raise of partial ordered structure radial tree provides a useful way for knowledge visualization, and it is of great significance to knowledge discovery and data mining of big data.

2. **Partial Ordered Structure Diagram.** Formal Concept Analysis (FCA) is a rigorous mathematical theory proposed by Wille in 1982 [5], it is a principled way of automatically deriving ontology from a collection of objects and their attributes [6]. Partial ordered structure diagram is developed on the basis of formal concept analysis theory. Formal concept analysis can effectively extract implied knowledge from data, especially in the process of handling practical science issues. Therefore, it is achievable to apply the theory of formal concept analysis to obtaining useful information from big data. The main advantage of this theory is that it can completely present the internal logic and organization structure of data (whether it is superficial or implied) and thus provide a systemic visualization tool to conceptual data analysis.

As one data mining method, partial ordered structure diagram determines the order of data characteristics according to the frequency of each characteristic, it builds partial ordered structure based on the correlation of data and achieves clustering according to partial order and aggregation formed from partial ordered structure, and then the universality and specificity of things can be found. Compared with traditional data mining methods, partial ordered structure diagram not only has a great visualization effect, but also is a good knowledge discovery method that integrates all kinds of statistical analysis approaches into a framework.



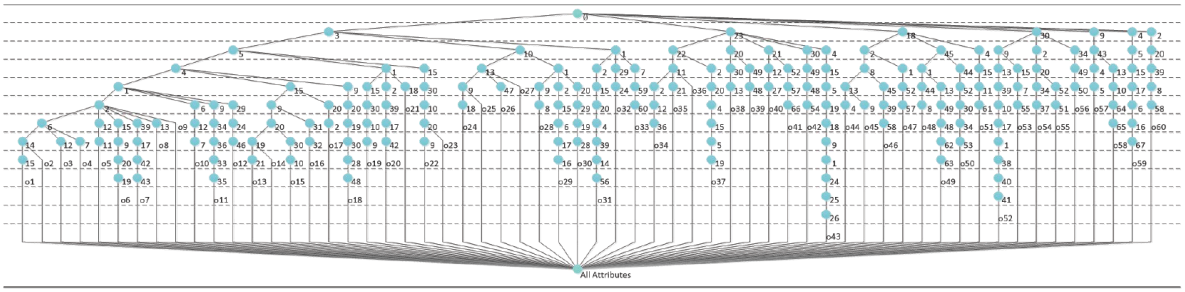
where: a1:need water, a2:aquatic, a3:terraneous, a4:chlorophyll, a5:dicotyledonous, a6: monocotyledonous, a7: mobile, a8: four limbs, a9: mammals; o1:frog, o2:giant salamander, o3:leech, o4:bulrush, o5: float grass, o6:corn, o7:bean, o8:dog.

FIGURE 1. Attribute partial ordered structure diagram of *Lives in water*

Applying the theory of partial ordered structure diagram, an attribute partial ordered structure diagram of *Lives in water* [7] is drawn and shown in Figure 1.

As Figure 1 shown, partial ordered structure diagram is a tree structure diagram which is closed and not circulatory. Each node corresponds to at least one branch and all the branches are unidirectional; it means that the path is irreversible from the initial node to the final node. Besides, partial ordered structure diagram has a clear flat structure, and there are not crossed lines. One node stands for an attribute and each branch stands for one object, it indicates that different kinds of creatures have different attributes and relations among different levels are obvious, and it also helps people discover more potential knowledge system.

3. **Limitations of Partial Ordered Structure Diagram in Large-Scale Formal Context.** At present, there is not an accurate definition for "big data"; it is generally believed that big data is a collection of data sets that are too large and complex to be managed using commonly used software tools. The features of big data can be summarized as the "4Vs": volume (amount of data), velocity (speed of data in and out), variety (range of data types and sources) and value (data information) [8]. Besides, big data itself is only a large amount of complex data, and the potential knowledge behind it is the huge



where: 1:Cassia twig, 2:Peony, 3:Licorice root, 4:Fresh ginger, 5:Jujube, 6:Chinese ephedra, 7:Gypsum, 8:Atractylodes rhizome, 9:Radix Ginseng, 10:Dried ginger, 11:Officinal magnolia, 12:Apricot kernel, 13:Aconite, 14:Gegen, 15:Pinellia ternate, 16:Schisandra fruit, 17:Asarum herb, 18:Tuckahoe, 19:Bupleurum, 20:Baikal Skullcap Root, 21:Mirabilite, 22:Citrus aurantium, 23:Rheum officinale, 24:Dragon Bone, 25:Minium, 26:Calcined Concha Osterae, 27:Pepperweed seed, 28:Trichosanthes Kirilowi Root, 29:Oyster shell, 30:Coptis root, 31:Xuanfuhua, 32:Haematitum, 33:Radix Rehmanniae, 34:DHG, 35:Maimendong, 36:Maziren, 37:Jizihuang, 38:Dark plum,39:Chinese angelica, 40:Shujiao, 41:Huangnie, 42:Tongcao, 43:Wuzhuyu, 44:Grifola, 45:Zexie, 46:Shuqi, 47:Pig bile, 48:Cortex Moutan, 49:Radix rehmanniae, 50:Common anemarrhena, 51:Cape jasmine, 52:Tree peony, 53:Talc, 54:Typha orientalis, 55:Bamboo leaf, 56:Xiongqiong(芎藭), 57:Peach seed, 58:Xiongqiong(芎劳), 59:Bamboo Shavings, 60:Cynanchum atratum, 61:Perilla leaf, 62:Common Yam, 63:Cornus officinalis, 64:Langdu Root, 65:Croton Seed, 66:Chinese Waxgourd Semen, 67:Blackberry lily; o1:Gegen and Pinellia ternate Decoction, o2:Gegen Decoction, o3:Cassia twig Chinese ephedra each half Decoction, o4:Guizhier Yuebiyi Decoction, o5:Cassia Twig Decoction plus Officinal magnolia and Xingzi, o6:Bupleurum and Cassia twig Decoction, o7:Sini Decoction plus Wuzhuyu and Fresh ginger, o8:Cassia Twig Decoction plus Aconite, o9:Cassia Twig Decoction, o10:Daqinglong Decoction, o11:Licorice root Decoction, o12:Cassia Twig Jiuni Decoction minus Peony & plus Shuqi Dragon Bone and Oyster shell, o13:Bupleurum Decoction plus Mirabilite, o14:Minor Bupleurum Decoction, o15:Fresh ginger Xiexin Decoction, o16: Xuanfu dai zheshi Decoction, o17:Baikal Skullcap Root Decoction plus Pinellia ternate and Fresh ginger , o18:Minor Bupleurum Decoction plus Coptis root and Tree peony, o19:Coptis root Decoction, o20:Chinese angelica Sini Decoction, o21:Tuckahoe Cassia twig Licorice root and Jujube Decoction, o22:Pinellia ternate Xiexin Decoction, o23:Licorice root Xiexin Decoction, o24:Tuckahoe Sini Decoction, o25:Sini Decoction plus Radix Ginseng, o26:Sini Decoction plus Pig bile, o27:Sini Decoction, o28:Cassia twig and Radix Ginseng Decoction, o29:Xiaoqinglong Decoction, o30:Bupleurum Cassia twig and Dried ginger Decoction, o31:Bentun Decoction, o32:Cassia twig Licorice root Dragon Bone and Oyster shell Decoction, o33:Zhupi Big Pill, o34:Maziren Pill, o35:Dachengqi Decoction, o36:Xiaochengqi Decoction, o37:Major Bupleurum Decoction, o38:Aconite Xiexin Decoction, o39:Rheum officinale Baikal Skullcap root Radix rehmanniae and Tree peony Decoction, o40:Daxianxiong Pill, o41:Rheum officinale and Tree peony Decoction, o42:Rheum officinale and Typha orientalis Decoction, o43:Bupleurum Decoction plus Dragon Bone and Oyster shell, o44:Zhenwu Decoction, o45:Aconite Decoction, o46:Chinese angelica and Peony Powder, o47:Cassia twig and Tuckahoe Pill, o48:Wuling Powder, o49:Xianqi Pill, o50:Grifola Decoction plus Coptis root and Tree peony, o51:Pinellia ternate Officinal magnolia Tuckahoe and Fresh ginger Decoction, o52:Dark plum Pill, o53:Radix Ginseng and Gypsum Decoction, o54:Coptis root and DHG Decoction, o55:Coptis root Baikal Skullcap Root Cape jasmine Tree peony and Peony Decoction, o56:Radix rehmanniae Common anemarrhena Coptis root and DHG Decoction, o57:Wuzhuyu Decoction, o58:Jiutong Pill, o59:Blackberry lily and Chinese ephedra Decoction, o60:Chinese angelica Powder.

FIGURE 2. Attribute partial ordered structure diagram of Chinese medicine rescriptions

value. Hence the key to researching big data is deriving relevant content from a mass of data and forming meaningful and valuable information [9].

Formal concept analysis takes formal context and concept generation as a core, the formal context must be very large when researching big data. The number of concept increases exponentially with the growth of context's size [10], so partial ordered structure tree diagram has certain limitation when processing big data. For instance, with increase of the number of objects and attributes, the readability and comprehensibility of structure diagram will become lower and lower. This shortcoming influences the effectiveness of extracting useful information and discovering potential knowledge from large number of complex data to some extent.

For example, *Shang Han Za Bing Lun* [11] written by Zhongjing Zhang in *Han* dynasty recorded many traditional Chinese medicine prescriptions which are used to treat typhoid disease. Now take 67 kinds of Chinese medicine as attributes and 60 kinds of prescriptions as objects from the book, and then build a formal context. The attribute partial ordered structure tree diagram formed from the formal context is shown in Figure 2, each node in this diagram represents one kind of medicine and each branch stands for one kind of prescription.

Figure 2 roughly reflects the whole structure and the relations between Chinese medicine and prescriptions, but cannot show the specific information clearly in certain area which is not big enough. Partial ordered structure diagram makes all nodes aligned both vertically and horizontally, this kind of orthogonal packing keeps consistent with coordinate axis and it coincides with people's visual habits, so it is very intuitive to general users. However, for the large hierarchical structure, especially for the structure whose breadth is very big, this kind of layout will lead to unreasonable aspect ratio. As is shown in Figure 2, though the diagram has already made full use of the transverse area, there is not enough space to show the information of each node clearly that causes the contradiction between the shortage of used space and the waste of available space. When the number of data becomes larger and data structure becomes more unbalanced, the contradiction will be more obvious.

4. **Partial Ordered Structure Radial Tree of Big Data.** In order to solve problems of unreasonable aspect ratio, underutilization of visualization space and bad visualization effect when using early partial ordered structure diagram to show large formal context, a new visualization method aiming at representing big data partial ordered structure radial tree is proposed. Figure 3 shows the attribute partial ordered structure radial tree which is improved from attribute partial ordered structure diagram of Chinese medicine prescriptions. Partial ordered structure radial tree uses concentric circles to represent attribute layers, the node in the center of the circle is root node (top-level node in early diagram), the outermost layer is object layer and all objects equally distribute on the circle. Then divide the central angle equally according to the number of objects, and put the attribute nodes of different levels on concentric circles whose radiuses are different. The specific location of each attribute node on the same level depends on the number of objects that have this attribute, the corresponding concentric circle is also divided into different areas and every area corresponds to one attribute node in this layer. Since the length of each orbit increases with the radius, there tends to be more room for the nodes.

The pseudo-code of big data partial ordered structure radial tree generation method is as follows.

Algorithm: *Generation of partial ordered structure radial tree*

Input: *The number of corresponding objects M, the number of attribute layer L, center coordinates O, x, y of radial tree, radius R of radial tree*

Output: *Partial ordered structure radial tree*

1 ***For*** $r = 1$ *to* $L + 1$;
2 *Draw concentric circles with center at O and radius equal to r;*
3 ***End For***
4 *Build polar coordinate system with center at O;*
5 *Define the outermost circle as object layer, let all objects equally distribute on the divided circle;*
6 *Put attribute nodes of different levels on concentric circles whose radiuses are different;*
7 *Ensure the specific location of each attribute node on the same level according to corresponding object and the number of objects that have this attribute;*
8 *Color the nodes according to different attributes;*
9 ***End***

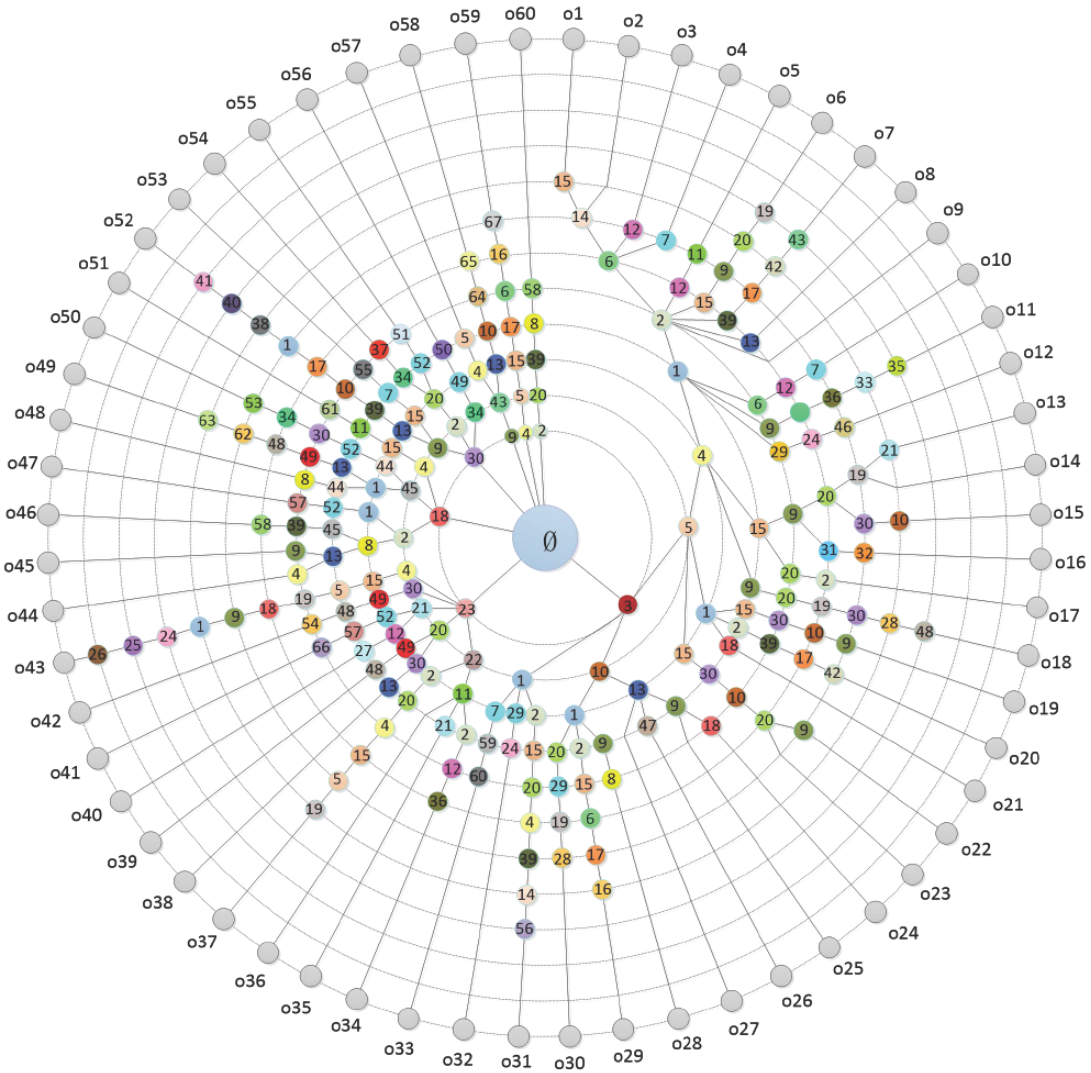


FIGURE 3. Attribute partial ordered structure radial tree

Since most large-scale formal context of big data is sparse, that is to say, the number of attribute layers is much less than the number of objects, now compare the capacity of objects in partial ordered structure diagram with that in radial tree diagram: take

for example the current mainstream screen, its resolution is 1440×900, suppose that the spacing of discernible object is d, then the maximum number of objects that partial ordered structure diagram contains is $1440/d$, while in the same space, the maximum number of objects that radial tree contains (the number of objects that the outermost object circle contains) is $900\pi/d$, that is to say, the number of objects in radial tree is almost two times of that in early diagram. For most paper publications, the width of available space is less than the length; under this situation, the number of objects that partial ordered structure diagram can contain will be less and the advantage of radial tree will be more obvious.

5. **Conclusion.** Partial ordered structure radial tree of big data improves readability and comprehensibility of knowledge discovery; it can be applied to analyzing large-scale formal context and has a higher efficiency. Applying this method to building partial ordered diagram avoids the emergence of crowded nodes, making hierarchical relations of attributes more obvious, and structural relation is easy to understand. It is important for development of theory of partial ordered structure and visualization analysis of big data.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grant 61273019, 61201111, 81273740 and 81373767. It is also supported by the Natural Science Foundation of Hebei Province under Grant F2013203368. The authors gratefully acknowledge the supports. The authors also gratefully acknowledge the reviewers for their valuable comments and suggestions to the paper.

REFERENCES

[1] X. Cheng, X. Jin, Y. Wang et al., Survey on big data system and analytic technology, *Journal of Software*, vol.25, no.9, pp.1889-1908, 2014.
[2] W. Chen, Z. Shen, Y. Tao et al., *Data Visualization*, Publishing House of Electronics Industry, Beijing, 2013.
[3] 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.
[4] C. Zheng, S. Li, J. Luan, J. Song and W. Hong, Partial ordered structure circular diagram: A new method for partial ordered structure visualization of big data, *Journal of Yanshan University*, vol.38, no.5, pp.409-415, 2014.
[5] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, in *Ordered Sets*, I. Rival (ed.), Dordrecht-Boston, Reidel, 1982.
[6] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundation*, Springer-Verlag, New York, 1999.
[7] *http://www.upriss.org.uk/fca/examples.html.*
[8] Big data, *Wikipedia*, http://en.wikipedia.org/wiki/Big_data.
[9] G. Li and X. Cheng, Research status and scientific thinking of big data, *Bulletin of Chinese Academy of Sciences*, vol.27, no.6, pp.647-657, 2012.
[10] Y. Ma, Z. Zeng, C. Chi et al., *Formal Concept and Its New Progress*, Science Press, Beijing, 2011.
[11] *http://www.zysj.com.cn/lilunshuji/shanghanzabinglun512/.*