A Distribution Model with Pattern Structure in Formal Concept Analysis for Meteorological Data Minging

Xiajiong Shen, Lei Zhang, Daojun Han and Peiyan Jia

Institute of Data and Knowledge Engineering, Henan University, Kaifeng, Henan 475004, China
E-mail: {shenxj, zhanglei, hdj, jpy}@henu.edu.cn

Abstract

Using data mining technology to analyze the huge amounts of meteorological data plays an important role in improving the accuracy of weather forecasts. After analyzed the features of meteorological data, a distributed meteorological data mining models using the pattern structure in formal concept analysis is proposed in this paper. Since there exists large numerical, boolean, and geographic concepts in meteorological data, using classic methods of formal concept analysis needs to build single-valued formal context. This paper adopts concept lattice pattern structure to avoid such conversions and the results of rules mining have higher readability and efficiency. This pattern structure of concept lattice is extended to the distributed model to improve data processing capability.

Keywords: formal concept analysis (FCA); concept lattice; Pattern Structure; Meteorological Data Minging

1. Introduction

The weather forecast work is running on the process of meteorological data analysis, and it is divided into short-time, short-term, and medium-term weather forecast[1]. At the same time, the climate prediction is also divided into short-time, short-term, and medium-term based on meteorological data analysis. Method based on statistics in the medium-term weather and climate prediction has a larger advantage. With the development of data mining technology, the research which utilize meteorological data mining using the computer and looking for the regularity of climate to give weather forecast and disaster early warning cause the attention of many researchers[1].

Garcia use visual data mining techniques to assess the quality of weather [2] data collected by weather stations. Kusiak proposed a data mining method to improve the wind farm power prediction accuracy, and set up time series model of four different forecast cycles for data mining [3]. Cortez proposed a meteorological data mining method to predict the burned area of forest fires [4]. Cheng Tao furtherly study the forest fire forecast mining method of time and space data [5]. Yang put forward a spatial data mining method based on satellite image sequence analysis to predict the storm [6]. Bartok introduced the framework of meteorological data mining and the method of predicting haze and low clouds in data mining meteor project [7]. Ganguly A R analysis challenge and opportunities of the data mining in the studies of the effects of global climate change [8].They put forward the mining algorithms to solve these problems, and give an illustrative case study.

Modern meteorological observation method makes the meteorological department has accumulated a lot of forms of meteorological data[9], which is a typical big data[10]: 1) the various types of data, including the stations data focus on observation data of the ground, high altitude, solar radiation, agricultural

meteorological stations and stations data of statistical processing products; the lattice data focus on assimilation analysis data of numerical model and numerical inversion products of remote sensing detection; image data focus on satellite cloud pictures and radar image; comprehensive meteorological data composed of comprehensive information of a certain area or range which is subject oriented and made up with a variety of data. 2) The great amount of data. The station received as high as 100 GB weather radar data every day; The central station received data is up to TB orders of magnitude every day, business application data up to PB orders of magnitude. 3) Timeliness requirements. The weather forecast and the geological disaster warning has strong real-time requirements for the analysis of meteorological data processing. 4) Huge value but low density. There are many factors influencing the weather, and the relationship between the factors is complex. The mining value of the local meteorological elements and part of the time meteorological data for meteorological association rule is not high.

The existing study can't meet the requirements of meteorological data mining in big data era. It is because that existing meteorological mining usually specific to a particular application and only analysis certain data sets and meteorological elements, but not all the meteorological data. In this paper, we established a meteorological data mining model integrated various meteorological and geographical data on the basis of pattern concept lattice, and use the distributed method to improve the time performance of the algorithm.

Concept lattice is the mainly data structure of formal concept analysis [9], it has been widely used in the field of data mining [11]. In concept lattice, we use the relationship between connotation sets to describe the association rule. It can be calculated by the inclusion (or approximate contains) relation of extension set. The unity of connotation and denotation on the concept node in concept lattice is the basic of extracting association rules. Because of relationship between concept node embodies the generalization and instantiated relation between concept, the concept lattice is also used as visual mining tool and classification navigation [12-13]. Classic concept lattice mining method need to process data as the single value form background, after that, construct the concept lattice for association rules mining[14]. There are many types of data in meteorological, it is hard to use concept lattice extension model of specific data types to manage these data unified.

Kuznetsov put forward a pattern structure of formal concept analysis which can solve the diversification of data type of knowledge discovery in large data, such as string, numerical interval, logic formula, figure and so on[15]. This paper focused on applying the method to the meteorological data mining. For two special kinds of meteorological data: numerical data and geographical information data, we define the similarity operation of structure model respectively. Finally, in order to take advantage of distributed hardware's processing capacity, we will further promote to the distributed model to improve data processing ability.

2. Basic Definition of Pattern Structure

Pattern structure is the extension of formal context and used for processing complex data. In the pattern structure, one object can be depicted by another, defining the similarity operation of two descriptions and constituting a semi-lattice with sets of the descriptions and similarity operation.

More generally, a pattern structure is a triple(G, D, δ) where G is a set of objects, D is a set of patterns, $D =(D, \cap)$ is a meet-semi-lattice of object descriptions or patterns, and $\delta : G \rightarrow D$ is a mapping providing any object $g \in G$ with a description $d \in D$.

The operation $\cap$ is a set of patterns, being used to compute the greatest lower bound of two patterns. Pattern structure was used for the studying of graphs at the earliest, the operation $\cap$ of two graphs is also the maximal attribute sharing of two patterns, therefore $\cap$ is also called similarity operation. The operation $\cap$ is idempotent, associative and commutative. It could define the order of D basing on operation $\cap$—containing relation $\subseteq$: given $c, d \in D$, it meets $c \sqsubseteq d \Leftrightarrow c \sqcap d = c$. The sets of patterns D and its operation $\cap$ construct a semi-lattice, i.e., every subset X of D has an infimum $\cap X$ in $(D, \cap)$. It's obviously that $(D, \cap)$ has the equivalence with $(D, \subseteq)$.

A pattern structure (G, D, δ) gives rise to the following derivation operators $(.)^{\diamond}$
$$A^{\diamond} = \prod_{g \in A} \delta(g)$$
$$d^{\diamond} = \{ g \in G \mid d \sqsubseteq \delta(G) \}$$

These operators form a Galois connection between the powerset of G and $(D, \subseteq)$. The former returns the maximum sharing descriptions (patterns) of sets of objects A, the latter returns all the objects whose description is d.

The pairs (A, d) satisfying $A^{\diamond} = d$, and $A^{\diamond} = d^{\diamond}$ are called the pattern concepts of (G, D, δ), with extent A and pattern intent d. In the pattern structure (G, D, δ), if $d^{\diamond\diamond} = d$, then pattern $d \in (D, \cap)$ is closed; if $A^{\diamond\diamond} = A$, sets of objects $A \subseteq G$ is closed. It's obviously that the extension and connotation of pattern concept is closed.

The definition of partial order in two pattern concepts is $(A_1, d_1) \leq (A_2, d_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow d_2 \sqsubseteq d_1$. Complete lattice being induced by partial order is called pattern concept lattice.

To the formal context (G, M, I) and pattern structure (G, D, δ), when I satisfying $xIm \Leftrightarrow m \sqsubseteq \delta(G)$, the two of them have congruent relation as follows: concepts on (G, M, I) and (G, D, δ) have one-to-one correspondence and the corresponding concepts have the same extension. Concept lattice induced by formal context (G, M, I) have isomorphism with the pattern concept lattice induced by pattern structure (G, D, δ).

There is a definition for applying pattern structure and pattern concept lattice to data mining as follows: $a, b \in D$, if $a^{\diamond} \subseteq b^{\diamond}$, then pattern contains $a \rightarrow b$ is established. It supposes that confidence coefficient is c, degree of support is s, if $s \geq | a^{\diamond} \cap b^{\diamond} | / |G|$ and $c \geq | a^{\diamond} \cap b^{\diamond} | / a^{\diamond}$, then the pattern association rule $a \rightarrow_{c,sb}$ is established.

3. Similar Computing in Meteorological Data

To structure the pattern concept lattice, we need to define the operation of the pattern according to the specific application environment. In this paper, we define $\sqcap$ operation of three kinds of data in meteorological data.

3.1. Numerical Interval

Numerical data is the most common type of data in meteorology. In station data of high attitude, ground, and radiation, the temperature, the humidity, the wind direction, and the rainfall is presented in numerical form. We need to convert it into numerical interval during data mining.

For numerical interval, we define the operator based on this understanding: two numerical interval are similar because they are included in a larger numerical interval, and the larger one should contain the two small numerical interval.

For two interval $[a_1, b_1]$ and $[a_2, b_2]$, and $a_1, b_1, a_2, b_2 \in \mathbb{R}$, so
$$[a_1, b_1] \sqcap [a_2, b_2] = [\min(a_1, a_2), \max(b_1, b_2)]$$

According to the definition of pattern in section 2, there is a relationship $\sqsubseteq$ between patterns. $c \sqsubseteq d \Leftrightarrow c \sqcap d = c$. So we can get the definition of two intervals' relationship from the definition of the $\sqcap$ operation.

$[a_1, b_1] \sqsubseteq [a_2, b_2] \Leftrightarrow [a_1, b_1] \sqcap [a_2, b_2] = [a_1, b_1] \Leftrightarrow [\min(a_1, a_2), \max(b_1, b_2)] = [a_1, b_1] \Leftrightarrow a_1 \leq a_2 \text{ 且 } b_1 \leq b_2 \Leftrightarrow [a_1, b_1] \sqsupseteq [a_2, b_2].$

There are usually several ranges of number of one record in meteorology data. So we must make definitions of $\sqcap$ operation in several ranges of number. We assume that data interval with same index in two same dimensions have the same meaning. In meteorology database, every column of data has same meaning, such as wind speed, air temperature and so on. In concept configuration, intervals with same index come from same column. So this assumption is reasonable. The $\sqcap$ operation in two numerical interval vectors is equal to the $\sqcap$ operation in two numerical intervals with same index.

We set $e=<[a_i, b_i]>_{i\in[1, p]}$, $f=<[c_i, d_i]>_{i\in[1, p]}$, as numerical interval vector of same level p, so the $\sqcap$ operation in same level numerical interval vector can be defined as follow:

$$e \sqcap f = <[a_i, b_i]>_{i\in[1, p]} \sqcap <[c_i, d_i]>_{i\in[1, p]} \Leftrightarrow e \sqcap f = <[a_i, b_i] \sqcap [c_i, d_i]>_{i\in[1, p]}$$

Similarly, two interval vectors satisfy the relationship $\sqsubseteq$ means that every numerical interval in vectors satisfies the relationship $\sqsubseteq$, the definition is as follow:

$$e \sqsubseteq f = <[a_i, b_i]>_{i\in[1, p]} \sqsubseteq <[c_i, d_i]>_{i\in[1, p]} \Leftrightarrow [a_i, b_i] \sqsubseteq [c_i, d_i], \ \forall i\in[1, p].$$

3.2. Geographical Regional Information

There is large amount of geographical regional information in weather forecast, such as the geographical regional of meteorological influence scope, the geographical regional of geological disaster, and the airline and flight information area in aeronautical meteorology etc. In order to make the geographical regional information to have comparable partial order relation, this article will present these geographical regionals with unified ontology. Then, we use ontology to realize the pattern similar computation of this kind of information.

Meteorological data of geography regional information can be divided into two kinds, one kind is common geographic name regional, which is named after administration or professional, they tend to have fixed meaning and location information. Another kind is meteorological phenomena such as precipitation, cloud area, which usually do not have a fixed name and can be described with the geographical coordinates.

Geography region ontology is established with location information and the containment relationship. Geography region ontology contains two attributes: name and location information. We only consider containment relationship of geographical location in semantic relationships between ontology. Containment relationship established by three ways: 1) We use its original containment relationship if the determine named regions do have the containment relationship, such as two administrative areas like Henan province and Zhengzhou; 2)we use the minimal outsourcing rectangular to describe the containment relationship between unnamed regions; 3)we also use the minimal outsourcing rectangular to describe the containment relationship between unnamed regions and named regions.

When the hierarchical relationship of ontology in database is already established according to the above principles, the partial order relation $\leq$ between ontology is also established according to the containment relationship. Partial order relation formal description of geography ontology is given below:

For geography region ontology o_1 and o_2, $o_1 \leq o_2$ is exists when o_1 is in the geography region of o_2.

$o_1 \vee o_2$ is the minimum upper bound, it means the minimum region ontology contains o_1 and o_2.

Function $S(z)$ represent geography region ontology corresponding geography region z. $T(o)$ represent geography region name corresponding geography region ontology o. E can be a geography region presented by place name, also can be the geography region presented by minimum outsourcing rectangular vector.

Here is definition of similarity computing in geography region information in weather database:

There are two geography region $z1$ and $z2$, so

$z_1 \sqcap z_2 = T(S(z_1) \vee S(z_2))$

In fact, similar computing of $\sqcap$ two geography region use the smallest geography region of the two geography region as similar computing result.

We can get the inclusion relation of two geography region according to $\sqcap$

$z_1 \sqsubseteq z_2 \Leftrightarrow z_1 \sqcap z_2 = z_1 \Leftrightarrow T(S(z_1) \vee S(z_2)) = z_1 \Leftrightarrow S(T(S(z) \vee S(z_2))) = S(z_1) \Leftrightarrow S(z_1) \vee S(z_2) = S(z_1) \Leftrightarrow S(z_2) \leq S(z_1)$

As for the similar computing $\sqcap$ of geography region vector, it can be defined as below:

We set E and Z as geography regions with same level. Their components are represented by e_i and z_i, and $i \in [1, p]$, so the $\sqcap$ operation of same level geography region vectors can be defined as below:

$E \sqcap Z = \{e_1,.., e_i, …, e_p\} \sqcap \{z_1,.., z_i, …, z_p\} \Leftrightarrow E \sqcap Z = \{e_1 \sqcap z_1,.., e_i \sqcap z_i,, …, e_p \sqcap z_p\} \Leftrightarrow \{ T(S(e_1) \vee S(z_1)),… T(S(e_i) \vee S(z_i)),…T(S(e_p) \vee S(z_p))\}$

It is same that two geography region vectors satisfy inclusion relation $\sqsubseteq$ means that every geography region in vectors satisfy inclusion relation $\sqsubseteq$, the definition is as below:

$E \sqsubseteq Z = \{e_1,.., e_i, …, e_p\} \sqsubseteq \{z_1,.., z_i, …, z_p\} \Leftrightarrow e_i \sqsubseteq z_2, \ \forall i \in [1, p]$.

3.3. Boolean Value

There is Boolean type of meteorological data in some weather report of automation acquisition device. Such as, whether there is the sky clouds, whether rainfall, and so on.

Boolean type of data is the basic form of formal context, which generally do not need special treatment. But in this paper, in order to be unified by the processing model of concept lattice model with the complex type data, we also need the definition of similarity computation.

Boolean value satisfies partial order relation itself, namely portrayed. So it is easy to be unified into the model concept lattice.

Here are two Boolean variables a, $b \in \{0,1\}$. The similar computing can be defined as below:

$a \sqcap b = a \wedge b$

Inclusion relation $\sqsubseteq$ can be defined like this:

$a \sqsubseteq b \Leftrightarrow a \sqcap b = a \Leftrightarrow a \wedge b = a \Leftrightarrow a \leq b$

It means that the operation $\sqcap$ and inclusion relation of Boolean variable reach agreement with operation $\wedge$ and relation $\leq$ of Boolean algebra. $0 \wedge 0 = 0$, $0 \wedge 1 = 0, 1 \wedge 1 = 1$, $0 \leq 0$, $0 \leq 1$, $1 \leq 1$.

3.4 Mixed Type Vector

One meteorological database data are often mixed with a variety of data types such as numerical interval, geographical information, and Boolean. We need to make the three types unified into one vector operation and inclusion relation. Operation $\sqcap$ of mixed type vector with same dimension means operation $\sqcap$ on every variable with corresponding index. Two mixed type vector with same dimension

satisfy containment relationship means of vector of each corresponding index variables meet the containment relationship.

4. The Construction of the Pattern of Concept Lattice

The construction of the pattern of concept lattice should carry on the pretreatment of data firstly, which will be converted to the pattern structure in the form of database. Then it constructs the pattern of concept lattice base on relative algorithms, and produces association rules according to the generated model of concept lattice.

4.1 The Conversions of Pattern Structure

It always needs a discretization of continuous data and converts continuous data into Boolean data putting numerical intervals as attributes. For example, adopting CP clustering algorithm to make discretization of continuous data, or being divided into several classes according to the data's aggregation degree, or generating multi-way trees.

In this paper, using the method of model structure does not need a discretization of clustering. 1) For numeric data, it can be directly represented in interval number of $[a, a]$. For example, some temperature data as 23.3, it can be directly represented in interval number of $[23.3, 23.3]$. 2) For geographic area data, it can be represented by adopting the name of the ontology or using the minimum outsourcing rectangular latitude and longitude coordinate vector. We can use the cities' name such as Zhengzhou or Kaifeng etc. Also we can use the coordinate vectors that describe the latitude and longitude rectangular range such as [112.42E, 114.14E, 34.16 N, 34.58N]. 3) For Boolean data, it can use 0 or 1 in Boolean algebra to represent 'yes' and 'no' or 'have' and 'no'.

It will generate pattern structure quickly by scanning and converting meteorological data that need to be performed according to the above three principles.

4.2 The Construction of the Pattern of Concept Lattice

The majority of concept lattice construction algorithm can be converted into construction algorithm of the pattern of concept lattice. In this paper, the famous Godin algorithm has been converted into construction algorithm of the pattern of concept lattice.

In the algorithm of Godin, it needs to put concepts which have the same amounts in connotation in an array. In the pattern structures, pattern connotation is not the traditional collection form, it can't be ordered by using the amounts of the elements in connotation collection. Thereby it needs to remold the Godin algorithm.

It defines the degree of $[a, b]$ is $b\text{-}a$, the limit of z in geography region is the depth of geography region noumenon being apart from its basic noumenon, the limit of Boolean value a is a. The limit of mixed type vector is the sum of all the elements.

Theorem 1 If there are two pattern concepts satisfying the formula $(A_1, d_1) \leq (A_2, d_2)$, then it's necessary to meet that the limit of d_2 is less than d_1.

Proof: According to the definition of limit, any two attributes like a and b, if $a \sqsubseteq b$, it's necessary to meet the limit of b is greater than a. Therefore, to the mixed type vectors x, y, if $x \sqsubseteq y$, it's necessary to meet the limit of y is greater than x. And because $(A_1, d_1) \leq (A_2, d_2) \Leftrightarrow d_2 \sqsubseteq d_1$, so it's necessary to meet that the limit of d_2 is less than d_1.

According to theorem 1, the pattern concepts of less limit is the sub-concepts of the greater limit.

Therefore the attributes amounts in connotation can be replaced by the pattern distance in the sorting in Godin algorithm. Then the following is the pattern concept lattice's Godin construction algorithm:

Algorithm 1: PatternGodin(L, x^*,$\delta(x^*)$)

```
BEGIN
  B[i]:={nodes' limit of pattern connotation is i};
  B'[i]=Φ;
  FOR i:= 0 to max(all the limits of pattern connotation) DO
      FOR C ∈ B[i] DO
        IF Intent(C) \subs⊆ δ(x*) THEN
          Extent(C):= Extent(C) ∪ { x*};
          Put C into B'[the limit of connotation of C];
          IF Intent(C) = δ(x*) THEN exit;
        ELSE
          Int:=Intent(C) ∩δ(x*);
         IF no exist C' ∈ B'[Int limit] so that Intent(C') =Int THEN
           Add new Cn:=(Extent(C) ∪ { x*}, Int);
           Then put Cn into B'[Int limit];
           Update the edges;
         ENDIF
        ENDIF
      ENDFOR
    ENDFOR
END
```

Classical Godin algorithm judges whether producing new concepts based on assembly intersection of connotation, PatternGodin judges whether producing new concepts based on intersection of pattern connotation. PatternGodin algorithm constructing concept lattice needs to produce a initial concept node, then add the objects in multiple-valued context to the constructed concept lattice one by one. Initial concept nodes are extending to empty sets, connotation is the maximal pattern among all the sets of patterns.

5. Experimental Results and Analysis

In order to illustrate the validity of this paper's method more intuitively, we choose an experiment in small data sets to verify the efficiency of pattern concept lattice. In order to reflect the universality of data sets, we produced multiple-valued formal context data sets. The objects' numbers start from 10 and increase 2 lattices every time until reach to 30. The attributes' number is 5(be ||5||), every attribute has 10 possible values (be ||S||=10) at most. We made the experiment with a common PC whose CPU is 2.3G, internal storage is 4G and operating system is Windows 7.
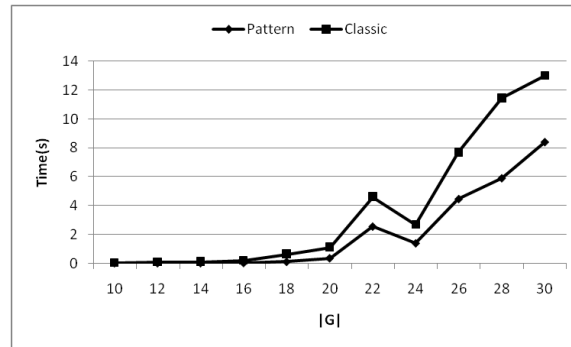
Figure 1. Performances of Pattern and Classic Concept Lattice Algorithms where ||M||=5, ||S||=10

The process of the experiment is as follows: Firstly, to construct concept lattice using pattern concept lattice construction algorithm in this paper mentioned and record time consume when constructing in every time. Then to convert multiple-valued formal context sets into monodrama formal context making use of tensor technology of formal context. After that to construct by adopting classical concept lattice construction algorithm AddIntent and record the time consume. The experiment result is as shown in Figure 1. We can find that the pattern concept lattice adopted in this paper has more quickly time performance than classical concept lattice and the performance characteristic is more obvious especially when the number of objects increases.

6. Conclusion

This paper has proposed a meteorological data mining method based on pattern structure and pattern concept lattice. This paper has also defined the similarity operation and partial ordering relation of three data types as numerical interval and geographic regional information and Boolean value contraposing the characterizes of meteorological data types, built the model structure of meteorological data unifying all kinds of data types and converted traditional concept lattice construction algorithm into pattern concept lattice construction algorithm basing on the characterizes of partial ordering relation. It differs from the traditional method that using tensor technology to convert multiple-valued formal context sets into monodrama formal context and then reconstructing concept lattice, the method in this paper can construct concept lattice on multiple-valued formal context directly. The experiment shows that the method in this papa has higher time performance.

In this paper, the further work is to apply the pattern structure and pattern concept lattice to picture data such as satellite cloud picture or radar map and data of meteorological data coding format proposed by the WMO, thus to full support the mining of all kinds of meteorological data.

Acknowledgements

References

[1] M. Mu, B. Y. Chen, F. F. Zhou and Y. S. Yu, "Methods and Uncertainties of Meterorolgical Forecas", Meteorological Monthly, vol.1, no.37, **(2011)**.

[2] J. R. M. Garcia, A. M. V. Monteiro and R. D. C. Santos, Visual Data Mining for Identification of Patterns and Outliers in Weather Stations' Data. Proceeding of the 13th Intelligent Data Engineering and Automated Learning-IDEAL, **(2012)**; Natal, Brazil.

A. Kusiak, H. Zheng and Z.Song, Short-term prediction of wind farm power: a data mining approach. IEEE Transactions on Energy Conversion, vol.1, no.24, **(2009)**.

[3] P. Cortez and A. J. R. Morais, "A data mining approach to predict forest fires using meteorological data", Proceedings of the 13th Portuguese Conference on Artificial Intelligence , **(2007)**; Guimarães, Portugal.

[4] T. Cheng and J. Wang, Integrated Spatio- temporal Data Mining for Forest Fire Prediction. Transactions in GIS, vol.5, no.12, **(2008)**.

[5] Y .Yang, H. Lin and Z. Guo, "A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis", Computers & geosciences, vol.1, no,33, **(2007)**.

[6] J. Bartok, O. Habala and P. Bednar, "Data mining and integration for predicting significant meteorological phenomena", Procedia Computer Science, vol.1, no.1, **(2010)**.

A. R. Ganguly and K. Steinhaeuser, "Data mining for climate change and impacts", Proceeding of the 8th IEEE International Conference on Data Mining Workshop, **(2008)**; Pisa, Italy.

B. Ganter and R. Wille, "Formal Concept Analysis: Mathematical Foundations", Springer, Berlin, **(1999)**.

[7] X. F. Meng and X. Ci, "Big Data Management: Concepts, Techniques and Challenges", Journal of Computer Research and Development, vol.1, no.50, **(2013)**.

[8] Z. P. Xie and Z. T. Liu, "Concept Lattice and Association Rule Discovery", Journal of Computer Research and Development, vol.12, no.37, **(2000)**.

[9] C. He, X. Q. Cheng and J. F. Guo, "Mining Hierarchical Concept Lattice for Faceted Navigation", Chinese Journal of Computers, vol.9, no.34, **(2011)**.

[10] C. Melo, M. A. Aufaure and C. Orphanides, "A conceptual approach to gene expression analysis enhanced by visual analytics", Proceedings of the 28th Annual ACM Symposium on Applied Computing, **(2013)**; Coimbra, Portugal.

[11] Y. H. Liu, W. Zhou and Z. T. Liu, "Interval Scaling Algorithm and its Concept Lattice Construction from Excented Formal Context", Computer Science, vol.10, no.36, (**2009)**.

[12] S. O. Kuznetsov, "Fitting Pattern Structures to Knowledge Discovery in Big Data", Proceeding of the 11th International Conference on Formal Concept Analysis, **(2013)**; Dresden, Germany.

Authors

Xiajiong Shen, he received Ph.D. degree in 2006. He is a professor at the School of Computer and Information Engineering, Henan University, Kaifeng, China. His major study fields include software engineering, knowledge discovery and spatial data process.
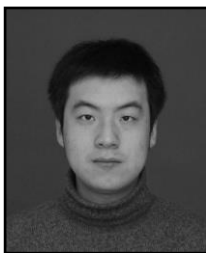
Lei Zhang, he received the M. Sc. degree from Henan University, Kaifeng, China, in 2006. He is a Ph.D. candidate of Harbin Institute of Technology, Harbin, China. His research interests include formal concept analysis, data mining, and information security.

Daojun Han, he received Ph.D. degree in 2011. He is an associate professor at the School of Computer and Information Engineering, Henan University, Kaifeng, China. His major study fields include information security, knowledge discovery and spatial data process.

Peiyan Jia, she received the M. Sc. degree from Henan University, Kaifeng, China, in 2006. Her research interests include formal concept analysis, data mining, and information security.