A Rough Concept Lattice Model of Variable Precision

SHENG Yan LI Yun TIAN Su-fang LUAN Luan

(Institute of Information Engineering, Yangzhou University, Jiansu Yangzhou 225009)

E-mail: shengyan1985@gmail.com

Abstract: Classical formal concept analysis can't deal with rough and uncertain information in practice, so the research on rough concept lattice and information express is an important task. There is an interesting relationship between approximate space and concept lattice in the rough set theory. The upper-lower approximation and variable precision in rough set theory are introduced into concept lattice, this paper proposes a rough concept lattice model of variable precision. Two parameters β_1 and β_2 which represent the roughness of the objects and attributes respectively in a rough concept in order to change the traditional Galois connection into the approximate mapping between objected set and attribute set. According to the different parameter β_1 and β_2 given by user, the rough concept lattice with different roughness can be obtained.

Key word: Formal concept analysis, Rough set, Rough concept lattice, Variable precision

1 Introduction

Formal concept analysis (FCA) is the research on the knowledge representation. Concept lattice is the core structure of FCA, which describes the concepts and their relationship. In a certain extent, it is a kind of highly simplified describing form of the real world. At present a considerable amount of research about the mathematical concept lattice nature and the use of concept lattices as a mathematical tool are applied in many fields, such as knowledge management[1], machine learning[2], information retrieval[3], software engineering[4] and so on. Rough set theory[5] raised by Pawlak Z in 1982 is a mathematical tools handling imprecision, vagueness and uncertainty in data analysis and it is a definable subset created by the equivalence classes based on the database and other subsets called upper and lower approximations. It provides a mathematical method of knowledge discovery and has already been widely used in knowledge acquisition, machine learning and other fields.

Both formal concept analysis and rough set theory is based on data tables, so they have close contact. Y.YYao discussed the corresponding relationship between the concept lattice and rough set theory based on the object oriented concepts lattice[6]. Wolff put forward the relationship between the concept lattice and rough set theory in the multivalued formal context. J.Y Liang[7] uses the concept lattice to express some attributes of rough set theory including the equivalence class, upper and lower approximations.

The paper is based on the similarity of concept lattice and rough set and combine the two and put forward a rough concept lattice model of variable precision to handle uncertainty and vague information and defined the approximate mapping of the object set and the attribute set in rough concepts. Two parameter β_1 and β_2 showing the roughness between the extent and intent of concepts is introduced. According to the different value of two parameter β_1 and β_2 input by user, we can get different fuzzy concept lattice with different roughness. When two parameter β_1 and β_2 is both 0, the variable precision rough concept lattice degenerated to the standard concept lattice, thus, the variable precision rough concept lattice can handle some rough and imprecise information as a rough model of expansion which is more versatile and significantly enrich the expression of concept lattice.

2 Rough Set & Formal Concept Analysis

2.1 Rough Set

Rough set theory is characterized by that we don't need to give out the quantity of certain characteristics or attributes and from a given set of the problem description, confirm the approximate domain by the indiscernibility classes directly and then find out the inherent laws of problems.

Def 1: Assume that U denotes a non-empty finite universe, R denotes a set of equivalence relations on U, U/R is all of the equivalence classes on R, for any $X \subseteq U$, the lower approximated set and upper approximate set of X on R can be defined respectively as below:

$$\underline{R}(X) = \{Y \in U/R \mid Y \subseteq [x]_R\} \quad (1) \qquad \bar{R}(X) = \{Y \in U/R \mid Y \cap [x]_R != \Phi\} \quad (2)$$

$\underline{R}(X)$ is the maximum set of objects which belong to X on the basis of known knowledge; $\bar{R}(X)$ is the union set of the equivalence classes $[X]_R$ whose intersection with X is not empty and also is the minimum set of object which maybe belong to X. The boundary regions of set X is defined as $BN(X) = \bar{R}(X) - \underline{R}(X)$. Sets have uncertainty owing to the existence of the boundary regions. The bigger the boundary regions are, the lower their accuracy is, and the greater their roughness is. When condition $\bar{R}(X) = \underline{R}(X)$ is satisfied, set X is called the accuracy set of R; Otherwise when $\bar{R}(X) \neq \underline{R}(X)$, set X is called the rough set of R. The rough set can be depicted by the upper and lower approximations of precise set.

2.2 Variable Precision Rough Sets

Ziarko proposed the variable precision rough sets model[8] ,in order to improve upon the anti-jamming capability of rough set model. The variable precision rough sets model which is the extension of Pawlak's rough set introduced the definition of precision based on the traditional rough set and so has a certain tolerance.

Def 2: Assume that both X and Y are non-empty subsets on finite universe, $C(X,Y) = 1 - Card(X \cap Y)/Card(X)$, which $Card$ denotes the cardinality and $C(X,Y)$ denotes the relative classification error ratio of set X and set Y. On the basis of $C(X,Y)$, the β-lower approximate of X and the β-upper approximate of X can be defined respectively as below:

$$\underline{B}_\beta X = U\{E \in U/R \mid C(E,X) \le \beta \quad (3) \qquad \bar{B}_\beta X = U\{E \in U/R\} \mid C(E,X) < 1-\beta \quad (4)$$

β denotes a certain level of error in classification, commonly set $0 \le \beta \le 0.5$. When $\beta = 0$, the variable precision rough set model degenerated to the standard rough set model. So we say that the variable precision rough set model is the extension of the standard rough set model.

2.3 Concept Lattice

Def 3: A formal context is defined as a triple (G,M,I), where G is a set of objects, M is a set of attributes and I is a binary relation between G and M, for $\forall g \in G, m \in M$, if $(g,m) \in I$, object g has attribute m, namely $g\,\mathrm{Im}$.

Def 4: In the formal context, between the object set $A \in P(G)$ and the attribute set $B \in P(M)$, there are the connections like as follow:

$A^* = \{m \in M \mid \forall g \in A, g\,\mathrm{Im}\}$, $B^* = \{g \in G \mid \forall m \in M, g\,\mathrm{Im}\}$, then every couple (A,B) satisfied $A = B^*, B = A^*$ from the formal context is called a formal concept. A which is the element of the power set $P(G)$ is called the extent of concept (A,B); B which is the element of the power set $P(M)$ is called the intent of concept (A,B).

Def 5: Assume that the concept $C_1 = (A_1, B_1)$ and the concept $C_2 = (A_2, B_2)$ satisfy $A_1 \le A_2$, then (A_1, B_1) is called sub-concept, and (A_2, B_2) is called super-concept, namely, $(A_1, B_1) \le (A_2, B_2)$. In the formal context, the partial ordering relation (also called generalization and specialization relationship) between sub-concept and super-concept formed a lattice structure called concept lattice.

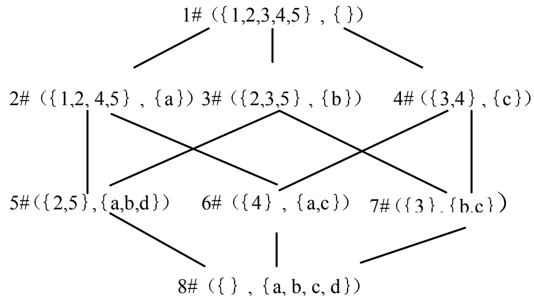
Here, the objects and attributes are based on the accurate sets, so the formal concepts are accurate.

The following table 1 shows a formal context and the corresponding concept lattice is shown in figure 1.

Tab.1 Sample of formal context

M \ G	a	b	c	d
1	1	0	0	0
2	1	1	0	1
3	0	1	1	0
4	1	0	1	0
5	1	1	0	1

Fig.1 the concept lattice of the sample in tab.1



3 A Variable Precision Rough Concept Lattice Model

Def 6: (the objects approximate, the attributes approximate) The formal context $k=(G,M,I)$, $g\in G, m\in M$, for $A\subseteq G, B\subseteq M$, $A^*=\{m\in M \mid g\,\mathrm{Im}\}$ and $B^*=\{g\in G \mid g\,\mathrm{Im}\}$, then $Co(A,B)$ denotes the approximate measure of object sets, for short, the objects approximate and $Ca(A,B)$ denotes the approximate measure of attribute sets, for short, the attributes approximate. Hereinto:

$$Co(A,B)=1-\frac{|A^*\cap B|}{Max(|A^*|,|B|)} \quad (5) \qquad Ca(A,B)=1-\frac{|A\cap B^*|}{Max(|A|,|B^*|)} \quad (6)$$

Def 7: (the objects roughness, the attributes roughness) In order to measure the similarity of sets, we choose two threshold β_1 and β_2 in the person of the roughness, which satisfy $0\le\beta_1\le 1, 0\le\beta_2\le 1$. β_1 and β_2 reflect respectively the roughness of the object sets and the attribute sets, which are correspondingly called the objects roughness and the attributes roughness.

Here, the objects roughness and the attributes roughness are commonly specified by user in different applications.

Def 8: The formal context $k=(G,M,I)$, $g\in G, m\in M$, for $A\subseteq G, B\subseteq M$, there is a connection called Galois connection $A^*=\{m\in M \mid g\,\mathrm{Im}\}$ and $B^*=\{g\in G \mid g\,\mathrm{Im}\}$. So two mapping relations can be defined, $\alpha: 2^U\to 2^M$ and $\beta: 2^M\to 2^U$. If it satisfied the two expressions below:

$$f(A)=\{m\in M \mid Co(A,B)\le\beta_1\} \quad (7) \qquad g(B)=\{g\in G \mid Ca(A,B)\le\beta_2\} \quad (8),$$

we defined $f(A)$ and $g(B)$ respectively as the approximate mapping of object sets and the approximate mapping of attribute sets.

β_1,β_2 are two adjustable roughness threshold, the parameter β_1 indicates the roughness of the intent of concepts and the parameter β_2 indicates the roughness of the extent of concepts, commonly, $0\le\beta_1\le 1$, $0\le\beta_2\le 1$. If two parameter $\beta_1=0, \beta_2=0$, the variable precision rough concept lattice degenerated to the standard concept lattice. The bigger the parameters β_1,β_2 are, the rougher the concept lattice is.

Def 9: (the rough formal concept) In the formal context $k=(G,M,I)$, if it satisfies $A=g(B), B=f(A)$, then we define the ordered pair $(A,B)=(g(B),f(A))$ as the rough formal concept, for short, the rough concept RFC and A is the extent of the rough concept, B is the intent of the rough concept.

Def 10: In the formal context $k=(G,M,I)$, two rough concept $C_1=(A_1,B_1)$ and $C_2=(A_2,B_2)$, if $A_1\subseteq A_2$ (or $B_2\subseteq B_1$), the rough concept C_1 is the sub-concept of C_2 and C_2 is the sup-concept of C_1, denoting as $C_1\le C_2$, $\le$ is called the partial ordering relation between the rough concepts. All rough concepts sets $RC(k)$ and partial ordering sets which are constituted of the partial ordering relations $\leq$ in k make up of the rough concept lattice(the rough lattice), denoting as $RL(G,M,I)=\{RC(k),\leq\}$.

From the formal context in Tab1, we set the value: $\beta_1=0.5,\beta_2=0.4$ and obtained the rough concepts as follows:

Tab.2 the rough formal concepts of the sample in Tab.1 ($\beta_1=0.5,\beta_2=0.4$)

Con	Extent	Intent	Con	Extent	Intent	Con	Extent	Intent
C1	{1,2,3,4,5}	{}	C7	{2,4,5}	{a,d}	C13	{2,3,5}	{b}
C2	{1,2,4,5}	{a}	C8	{1,2,5}	{a,d}	C14	{3,5}	{b}
C3	{2,3,5}	{a,b}	C9	{2,4,5}	{a,b}	C15	{}	{a,b,c,d}
C4	{2,3,5}	{b,d}	C10	{4}	{a,c}			
C5	{3,4}	{c}	C11	{3}	{b,c}			
C6	{1,2,5}	{a,b}	C12	{2,5}	{a,b,d}			

According to the following definitions, we can get rid of the redundant concepts from the rough concept set above.

Def 11: (the redundant object concept) For rough concept $C_1=(A_1,B_1),C_2=(A_2,B_2)$, if $B_1=B_2,A_1\subset A_2$, C_1 is defined as the redundant object concept of the concept C_2.

Def 12: (the redundant attribute concept) For rough concept $C_1=(A_1,B_1),C_2=(A_2,B_2)$, if $A_1=A_2,B_1\subset B_2$, C_1 is defined as the redundant attribute concept of the concept C_2.

For example, two concepts C3({2, 3, 5}, {a, b}) and C13({2, 3, 5}, {b}), $Extent(C3)=Extent(C13),Intent(C13)\subset Intent(C3)$, according to the Def 12, know that C13 is the redundant attribute concept of C3 , so can reduce C13 and only keep C3. The redundant object concept is the same.

Then we put forward the constructing algorithm of the variable precision rough concept lattice as follows:

```
Algorithm 1: RoughLattice(FormalContext, β1, β2)
Input: Formal Context, Two Adjustable Roughness Threshold
Output: The Rough Concept Lattice
Initial the Lattice and other preparation
For obj in Object Set:
    For H in All concepts which need to be regenerated in the old lattice:
        if H.intent <= intent(obj):    # If the intent of current concept contains the intent of the object, change the current concept
            H.addExtent(extent(obj))
        if H.intent == intent(obj):    # If the intent of two is equal, break the inner for and deal with the next obj
            break
        else:
            for intSet in Power set of intent(obj):
                Get the object set fxintSet corresponding to intSet
                for Hn in get_Hn_Concept(H.extent, intSet, fxintSet, β1, β2): # Obtain all candidate rough concepts satisfied the roughness threshold β1 and β2.
                    if Hn is the redundant object concept or the redundant attribute concept: continue
                    Add new rough concept Hn
                    Add edge Hn->H
                    Adjust the edge between Hn and the concept which the layout is higher than Hn
                    if intSet == fxSet:    End to handle with object and go next

get_Hn_Concept(A, intSet, fxintSet, β1, β2)
Input: Object Set, New Intent Set, Object Set corresponding to New Intent Set, Two Roughness Threshold
Output: All Rough Concept satisfied the roughness threshold β1 and β2
        for mex in Power Set of Object Set A:
            Obtain the attribute set mexs corresponding to mex
            Co = 1- |the intersection of mexs and intSet|/max{|mexs|, |intSet|}
            if Co > β1:
                continue
            Ca = 1- |the intersection of mex and fxintSet|/max{|mex|, |fxintSe|}
            if Ca > β2:
                continue
            Produce A New Candidate Concept(mex, intSet)
```

The rough concept lattice corresponding to the formal context in Tab 1 is shown in Fig 2. From the Fig 2, we can see that the concepts: 3#({2, 3, 5}, {a, b}) and 4#({2, 3, 5}, {b, d}), 6#({1, 2, 5}, {a, b}) and 8#({1, 2, 5}, {a, d}), 7#({2, 4, 5},{a, d}) and 9#({2, 4, 5}, {a, b}) is impossible to generate simultaneously. The concept 3# and 4# are corresponding to the concept 13#({2, 3, 5}, {b}) in the Fig 1, while concepts 6#, 7#, 8# and 9# are in new layout between 2#({1, 2, 4, 5}, {a}) and 12#({2, 5}, {a, b, d}.

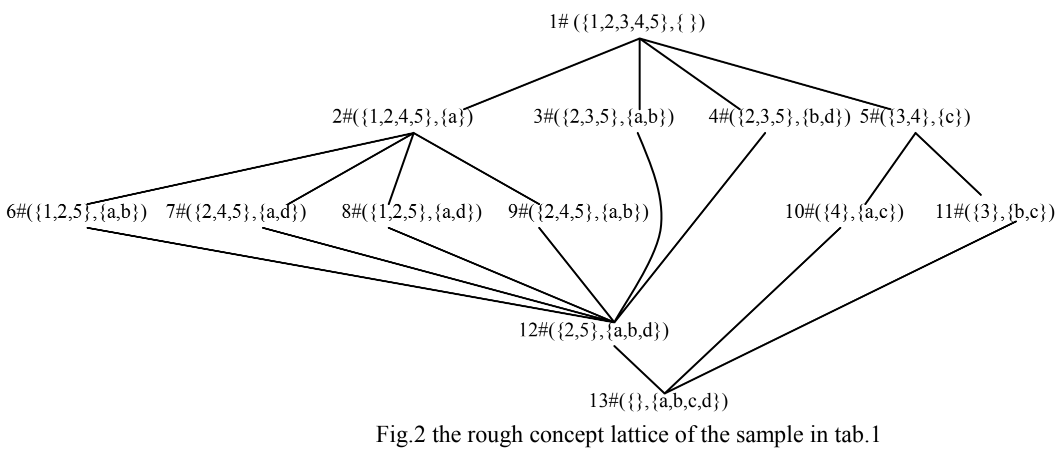


Fig.2 the rough concept lattice of the sample in tab.1

4 Algorithm Analysis & Evaluation

In order to observe the relations of two roughness threshold β_1 and β_2, we do some test in P4 2.8 GHz PC and using Python 2.5.2 in Ubuntu8.04. Because the time complexity of computing power set is super high and one set has n elements will has 2^n subsets, it costs a lot to compute Co and Ca. Here, we adopt two rules: (1) Calculate the power set Beforehand and save the result to local file be invoked when program is running, rather than real-time generating the power generating sets; (2)When calculate Co and Ca, the current set dissatisfy two threshold, then abandon the calculation of its parent set. We generate the random formal context and the number of attribute is 20, the rate between object and attribute is 0.3, the number of object is 20, the movement of the number of rough concept when β_1 and β_2 augment gradually, the result is shown in Fig 3:

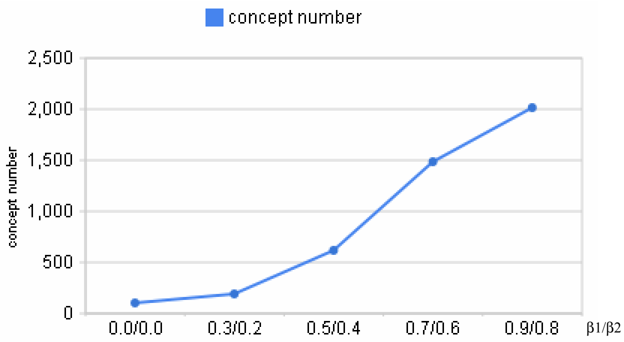


Fig 3 Test Result When β_1, β_2 varies

When β_1=0.0, β_2 augment gradually and β_2=0.0, β_1 augment gradually, the result is shown in Fig 4:

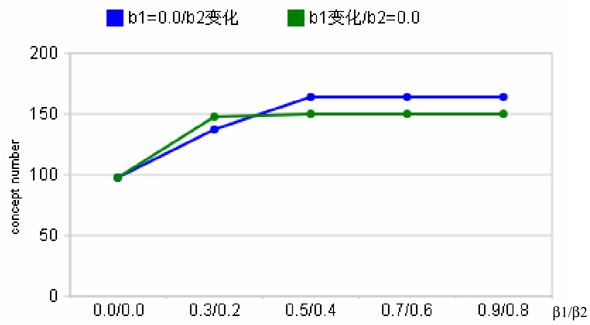


Fig 4 Test Result When β_1, β_2 varies separately

From the fig 3, we can see that the bigger β_1 and β_2 is, the more the rough concept is and the more rough the lattice is. If we concern the information just about objects or attributes, we could only adjust the relevant roughness threshold and thus, it reduce the rough concept number and enlarge the detail of information.

The time complexity hoik because of the increment of the formal context. One Method to solve this problem is that partition the whole formal context into many small formal contexts, then constructs the small rough lattice distributed, finally merge them together. The detailed method is similarly to the paper[9].

5 Conclusion

This paper combines the rough set theory and the concept lattice and put forward the variable precision rough concept lattice model. We change the mapping relation between the object set and the attribute set from the traditional Galois connection into the approximate mapping relation and also introduce two parameters β_1 and β_2 to incarnate the roughness between the extent of concepts and the intent of concepts. As a result of two roughness and the approximate mapping of object set and attribute set, the concept becomes rough then form the rough concepts satisfied two thresholds. We can obtain more interesting information than the traditional concept lattice. By adjusting two roughness β_1 and β_2, we can get the rough concept lattice with different roughness which significantly enrich the expression of concept lattice.

Researching on the nature of rough concept lattice, clustering method of rough concept lattice and the application of the variable precision rough concept lattice model in data mining and so on, is the next work.

References

[1] Krohn U, Davies NJ, Weeks, R.Concept lattices for knowledge management[J]. BT Technology Journal, 1999, 17(4):108-116.

[2] Kuzentsov SO. Machine learning on the basis of formal concept analysis[J]. Automation and Remote Control, 2001, 62(10):1543-1564

[3]Carpineto C, Romano G. Information retrieval through hybrid navigation of lattice represenatations[J]. International Journal of Human computer Studies, 1996, 45:553-578

[4] Godin R,MiLi H,Mineau GW,Missaoui R, Arfi A , ChauT-T. Design of class hierarchies based on concept (Galois) lattices[J]. Theory and Application of Object Systems, 1998, 4(2):117-134

[5] Pawlak Z. Rough set[J]. International of Computer and Information Science, 1982, 11: 341-356

[6] Yao Y Y. A comparative study of formal analysis and rough set theory in data analysis[C]. Proceedings of Rough Sets and Current Trends in Computing. Springer, 2004:59 - 68

[7] Liang Ji-ye. Research on the Method of Intelligent Data Analysis Based on Rough Set and Concept Lattice[D], doctoral dissertation ,Institute of Computing Technology, The Chinese Academy of Sciences,2004

[8] Ziarko, W. Variable precision rough set model [J]. Journal of Computer and System Sciences, 1993,46, 39–59.

[9] Yun Li, ZT Liu, L Chen, XH Xu, W Chen; Horizontal Union Algorithm of Multiple Concept Lattices [J]; Acta Electronica Sinica; 2004.