Formal Concept Analysis Based on Rough Set Theory and a Construction Algorithm of Rough Concept Lattice

Haifeng Yang

School of Computer Science and Technology, TaiYuan University of Science and Technology, TaiYuan, People's Republic of China
yhftxy7537750@163.com

Abstract. FCA(Formal Concept Analysis), which is accurate and complete in knowledge representation, is an effective tool for data analysis and knowledge discovery. A new lattice structure named RCL (Rough Concept Lattice) is presented. Using the approximation method of rough sets, we described the extent as approximation extent so that it can deal with uncertainty knowledge. In the end, a construction algorithm CARCL is provided based on it.

Keywords: Formal Concept Analysis, Rough Set, Intent, Approximation, Extent.

1 Introduction

Formal concept analysis, which was presented by Wille R in 1982, is an effective tool for data analysis and knowledge discovery [1].Every node is a formal concept, which was composed of two elements: Intent and Extent. Intent is description of concept, and extent is objects set, the elements of which have all the attributes of intent. The process of Construction is actually concept clustering. In addition, the relation of specialization and generalization can be shown through Hash map vividly and sententiously. It can be used in information searching, digital library, knowledge discovery, etc [2]. Now, many researches are centered in construction algorithm of concept lattice and its improvement [3] [4], and extension of concept lattice using other theories, such as rough set、 fuzzy theory, etc[5-7].

Concept lattice is accurate and complete in knowledge representation, and intent of node describes the common attribute of objects. For an Extent of normal concept lattice, the relation among the attributes of the intent is "and". So, if an object has no only an attribute value of the intent, it has to be got out of the extent set. It is an advantage of FCA, but it is limited in some areas too. If an object has one, another, or some of attributes of the intent, this uncertain data can't be shown, or probably hard to be shown, in FCA. Further more, we have to traverse all the nodes of FCA to get this uncertain knowledge, which will pay out huge cost. For example, a people may have one, another or some of symptoms of one certain illness, but a doctor must not debar this people from the illness just because of the information being uncertainty. So, expressing this uncertain knowledge in concept lattice will be valuable.

First, a new lattice structure, named RCL, is presented based on decision context in this paper. Next, using the approximation method of rough sets, it describes the extent afresh. Last, a Construction Algorithm of Rough Concept Lattice (CARCL) is provided based on it.

H. Deng et al. (Eds.): AICI 2011, CCIS 237, pp. 239–244, 2011.

2 Rough Concept Lattice

Definition 1. For two random sets A, B, the CADAP (Condition Attribute –Decision Attribute Product) of set A and B, is defined by:

$$A*B = \{(a, b) \mid a \in P(A), b = \text{“b1b2b3…bn”}, bi \in B\ (i \in [1, n])\} \quad (1)$$

Definition 2. Suppose a triple S= {O, C∪D, V1∪ V2} is a decision table where O= {x1, x2,…, xn}, C= {c1, c2, …, cm}, D= {d1, d2, …, dr}, V1 and V2 are finite and nonempty sets, the elements of which are called objects, condition attributes, decision attributes, condition attribute values and decision attribute values, and C∩D=Φ. The relationships of objects and attributes are described by a binary relation R between O and C*D, which is a subset of Cartesian product O×C*D. For a pair of elements x∈ O and a ∈C*D, if (x, a) ∈R, we say that the object x has the attribute a, or the attribute a is possessed by the object x. So the triple Ks= (O, C*D, R) is called Decision Context which was educed from the decision table S.

Table I is a decision context.

Table 1. A simple decision context

C∪D \ O	1	2	3
Fever(c_1)	√	√	√
Cough(c_2)		√	√
Headache(c_3)	√		√
Lack of Power(c_4)	√		
In appetence (c_5)		√	
Conclusion(D)	Grippe (d_1)	Chicken pox (d_2)	Measles (d_3)

Let Ks= (O, C*D, R) be a decision context. Then there is only one partial ordering set respond to it and this partial ordering set can create a structure of normal concept lattice.

Definition 3. For a decision context Ks= (O, C*D, R), operator f and g can be defined by:

$$\forall x \in O, f(x) = \{y \mid \forall y \in C*D, xRy\} \quad (2)$$

$$\forall y \in C*D, g(y) = \{x \mid \forall x \in O, xRy\} \quad (3)$$

Which is to say that f is a mapping from object x to all attributes possessed by x, and g is a mapping from attribute y and all objects which has y.

Definition 4 Let Ks= (O, C*D, R) be a decision context. For an attribute set Y ⊆ C*D, we can define the set M and N:

$$M = \{x \mid \forall x \in O, f(x) \cap Y \neq \Phi\} \quad (4)$$

$$N = \{x \mid \forall x \in O, f(x) \cap Y = Y\} \quad (5)$$

A triple H (M, N, Y) can create any node of concept lattice L, and L is called RCL (Rough Concept Lattice) educed from the decision context Ks.

For a node H (M, N, Y), the element Y is called Intent, which is description of concept; M is called Upper Approximation Extent; and N is called Lower Approximation Extent. The elements of M are objects which have one, another, or some of attributes of Y probably; and the elements of N are objects which have all the attributes of Y certainly.

Definition 5. Let an ordering set H (M, N, Y) be a node of RCL. It is complete on relation R, and it has the properties:

$$M=\{x \in O \mid \exists y \in Y, xRy\},\ N=\{x \in O \mid \forall y \in Y, xRy\} \tag{6}$$

$$Y=\{y \in C^*D \mid \forall x \in M, \exists y \in C^*D, xRy\} \tag{7}$$

Definition 6. For two rough concepts (M1, N1, Y1) and (M2, N2, Y2), (M1, N1, Y1) is a sub-concept of (M2, N2, Y2), written (M1, N1, Y1)≤ (M2, N2, Y2), and (M2, N2, Y2) is a super-concept of (M1, N1, Y1), if and only if Y1⊂Y2. If there is no concept (M, N, Y) that is a sub-concept of (M2, N2, Y2) and is a super-concept of (M1, N1, Y1), (M1, N1, Y1) is an immediate sub-concept of (M2, N2, Y2), and (M2, N2, Y2) is an immediate super-concept of (M1, N1, Y1).

Obviously, if (M1, N1, Y1) is a sub-concept of (M2, N2, Y2), M1 is the subset of M2, and N2 is a subset of N1.

Definition 7. For two nodes of RCL: r and t. If r is the super-concept of all other nodes of the RCL, we call r the Root-Node; and if t is the sub-concept of all other nodes of the RCL, we call t the Twig-Node.

3 A Construction Algorithm

3.1 The Main Steps of Construction Method

There are three steps in constructing a RCL. First, to get all intents from decision context, we will turn different decision attribute values into one string, and get the ordering sets, the elements of which is one element of condition attributes' power set and a decision string; Next, to get all Upper Approximation Extents and Lower Approximation Extents, we will traverse all records of decision context. M is consisted of all objects possessed one or some of attributes of an appointed intent, and N is consisted of all objects possessed all attributes of an appointed intent. The process of construction is to find immediate sub-concept nodes and immediate super-concept nodes for every new node. We can suppose the node as sextet (M, N, Y, Parent, Children, no).

3.2 The Description of CARCL (the Construction Algorithm of Rough Concept lattice)

```
INPUT:   Decision Table S (O, C∪D, V1∪V2)
OUTPUT: Rough Concept Lattice L
Function1: Get-Intent(S)
BEGIN
     FOR every record in S
       {FOR every v in current record {d=d & v}
        Des←d}
     Get power set of condition attributes set C: P(C) except Φ
     FOR every d in Des
       {Create Node h and set h->M, h->N and h->Y Φ
        FOR every element P in P(C) {h->Y =P ∪{d}; H←h}}
ENDS
Function2: Get-M-N(S, H)
BEGIN
     FOR every h in H
       {FOR every Record in S
           {IF the object possessed all the attributes of h->Y THEN {h->N←o}
           FOR every element d in attributes set of object 'o'
              {IF d∈h->Y THEN {h->M←o, EXIT FOR}}}
     FOR every node h in H {IF h->M = Φ THEN delete this node}
ENDS
Function3: Build-RL(S, H)
BEGIN
    L={(O•Φ, C∪Des, Φ, {1}, 0), (Φ, Φ, Φ, {0}, Φ, 1)}
    FOR every h in H
      {h->no=Count
      FOR every node n in L
          IF h is an immediate sub-concept of n
         THEN {h->Parent ←n->no; n->Children←h->no;
                FOR every immediate sub-concept node n' of n
                     {IF n'->Y ⊆ h->Y
                      THEN {h->Children←n'->no; n'->Parent←h->no;
                             delete n'->no from n->Children}}}
      Count++; L←h}}
ENDS
```

3.3 Algorithm Analysis and Example

We can turn multi-decision-attributes values into one string, get power sets of condition attributes, and initialize all nodes with intents, in Function1. Its time complexity is O $(2^n \times n \times n)$. Upper Approximation Extent M and Lower Approximation Extent N for every node will be initialized in Function2. It's time complexity is O $(n \times 2^n \times |h.Y|$. And we can build RCL in Function3. It is complete, but inefficient.

Figure1 is a part of RCL based on a decision context which is expressed through table1. #1 and #2 are Root-Node and Twig-Node respectively. For M, the relation among the intent of other nodes is "or", and for N, the relation among the intent is "and". For example, M of #9 is {1, 2, 3}, which express that its objects may have fever or cough and may have Chicken pox possibly, N of #9 is {2}, which express that its objects have fever and cough and have Chicken pox certainly.

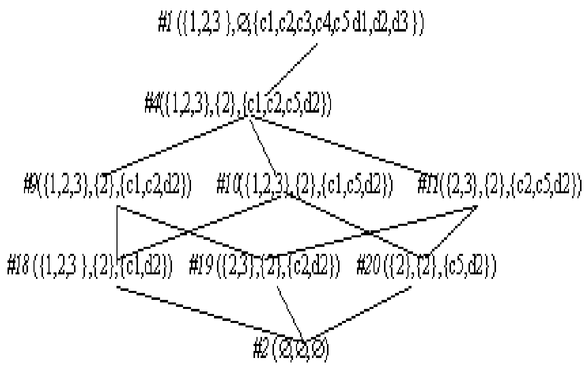

Fig. 1. An Example of a Rough Concept Lattice

4 Conclusions

In this paper, we described extent as approximation extent by using upper (lower) approximation method of Rough Set, so that it can deal with uncertainty knowledge. Based on a decision context, the correctness of CARCL is verified. Next research will be extraction based on RCL and the character and application of RCL.

References

1. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. M. Rival I Ordered Sets, 415–470 (1982)
2. Hu, K., Lu, Y., Shi, C.: Advances in concept lattice and its application. Tsinghua Univ. (Sci. & Tech.) 40(9), 77–81 (2000)
3. Godin, R.: Incremental concept formation algorithm based on Galois(concept) lattices. Computational Intelligence 11(2), 246–267 (1995)

4. Xie, Z.-P., Liu, Z.-T.: A Fast Incremental Algorithm for Building Concept Lattice. J. Chinese J.Computers 25(5), 490–496 (2002)
5. Qu, K.-S., Zhai, Y.-H., Liang, J.-Y., Li, D.-Y.: Representation and Extension of Rough Set Theory Based on Formal Concept Analysis. Journal of Software 18(9), 2174–2182 (2007)
6. Qiang, Y., Liu, Z.-t., Lin, W., Shi, B.-s., Li, Y.: Research on Fuzzy Concept Lattice in Knowledge Discovery and a Construction Algorithm. Acta Electronica Sinica 33(2), 350–353 (2005)
7. Qiu, W.-g., Zhang, G.-d.: Formal Concept Lattice Based on Composition Rough Operator. Computer Engineering 34(24), 16–18 (2008)