Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010

A HEURISTIC METHOD TO ATTRIBUTE REDUCTION FOR CONCEPT LATTICE

JUNHONG WANG, JIYE LIANG, YUHUA QIAN

School of Computer and Information Technology, Shanxi University, Taiyuan, 030006, Shanxi, China
Key Laboratory of Computational Intelligence and Chinese Information Processing of
Ministry of Education, Taiyuan, 030006, China
E-MAIL: wjhwjh@sxu.edu.cn, ljy@sxu.edu.cn, jinchengqyh@sxu.edu.cn

Abstract

Attribute reduction plays an important role in data mining based on concept lattice theory, which makes knowledge discovery from data easier and knowledge representation simpler. However, many existing methods often employ a discernibility matrix to calculate a set of complete reducts and are time-consuming. To solve this problem, in this paper, based on an expanded concept lattice, called closed label lattice, a heuristic algorithm to attribute reduction is proposed. The proposed algorithm and an illustrative example show that the algorithm can effectively obtain an attribute reduct from a formal context.

Keywords:

Concept lattice; closed label lattice; attribute reduction; heuristic information

1. Introduction

Concept lattice theory was proposed by Wille[1] in 1982, which is a kind of important mathematical tool for conceptual data analysis and knowledge processing. It provides a theoretical framework for the discovery and design of concept hierarchies from relational information systems. Most of the researches on concept lattice focus on such topics as: construction of concept lattice[2,3], acquisition of rules[4,5], extended model of concept lattice and relationship with rough set[6-10]. To date, concept lattice has been applied to information retrieval, digital library, software engineering and other aspects[11,12].

In the concept lattice theory, the data for analysis are described by formal context (U, A, I), which consists of universe U, attributes set A, and relation $I \in U \times A$. Based on the formal context, we can construct some formal concepts and the set of all the above formal concepts forms a concept lattice. The concepts are constituted by two parts: extension, which consists of all objects belonging to the concept, and intension, which comprises all attributes shared by the objects. The concept lattice reflects the relationship of generalization and specialization among concepts. It is an intuitive and effective way to represent, discover and design knowledge structures.

Attribute reduction plays an important role in knowledge representation and data mining. Many types of attribute reduction have been proposed in rough set theory[13-16]. Comparing with the studies on attribute reduction in rough set theory, there is less effort investigated on the issue in formal concept analysis. The attribute reduction in concept lattice is to find the minimal sets of attributes, which can determine a concept lattice isomorphic to the one determined by all attributes while the object set remains unchanged. It makes the discovery of implicit knowledge in data easier and knowledge representation simpler, and extends the theory of concept lattice. The theory and methods of reduction in concept lattice have been studied by some authors[17-22] recently. In [17], the Concepts of granular consistent set and granular reduct in the formal context were introduced, and then discernibility matrices and Boolean functions were respectively employed to determine granular consistent sets and calculate granular reducts in formal contexts. In [19], Based on a discernibility matrix, an approach to attribute reduction in concept lattice was presented, and the characteristics of core attribute were analyzed. In terms of rough set theory, the reduction of concept lattices in the sense of lattice isomorphism was studied and two kinds of reduction methods for concept lattices in formal contexts are proposed in [20]. In [21], relations of attribute reduction between object and property oriented formal concept lattices were discussed.

978-1-4244-6527-9/10/$26.00 ©2010 IEEE

Most of these existing methods study attribute reduction of concept lattice by rough set theory and employ discernibility matrix to calculate a complete reduct. The purpose of the paper is to study the method of attribute reduction from the set of intension of concepts and to give a heuristic algorithm for obtaining one reduct.

This paper is organized as follows. Basic definitions of concept lattice and closed label lattice are recalled in Section 2. In Section 3, the basic theory of attribute reduction in concept lattice are introduced. In Section 4, a new heuristic approach to attribute reduction in concept lattice is proposed. Finally, some conclusions are given in Section 5.

2. Preliminaries

In this section, we review some basic concepts of concept lattice and closed label lattice[1,6,23].

Definition 1. A formal context is a triplet (U, A, I), where $U = \{x_1, x_2, \ldots, x_n\}$ is a non-empty, finite set of objects called the universe of discourse, $A = \{a_1, a_2, \ldots, a_m\}$ is a non-empty, finite set of attributes, and $I \subseteq U \times A$ is a binary relation between U and A, where $(x, a) \in I$ means that object x has attribute a.

In this paper, we assume that the binary relation I is regular, that is, it satisfies the following conditions: for any $(x, a) \in U \times A$,

(1) there exist $a_1, a_2 \in A$ such that $(x, a_1) \in I$ and $(x, a_2) \notin I$,

(2) there exist $x_1, x_2 \in U$ such that $(x_1, a) \in I$ and $(x_2, a) \notin I$.

For $X \subseteq U$ and $B \subseteq A$, we define

$$\begin{aligned} X^* &= \{a \in A : \forall x \in X, (x, a) \in I\}, \\ B' &= \{x \in U : \forall a \in B, (x, a) \in I\}. \end{aligned} \quad (1)$$

X^* is the maximal set of attributes shared by all objects in X. Similarly, B' is the maximal set of objects that have all attributes in B. For $x \in U$ and $a \in A$, we denote $x^* = \{x\}^*$ and $a' = \{a\}'$. Thus x^* is the set of attributes possessed by x, and a' is the set of objects having attribute a.

Definition 2. Let (U, A, I) be a formal context. A pair (X, B), with $X \subseteq U$ and $B \subseteq A$, is called a formal concept of the context (U, A, I) if $X^* = B$ and $B' = X$. The set of objects X and the set of attributes B are respectively called the extension and the intension of the formal concept (X, B).

In the paper, for a formal concept C, the extension noted as extension(C) and the intension noted as intension(C).

The set of all formal concepts forms a complete lattice called a concept lattice and is denoted by $L(U, A, I)$. The meet and join of the lattice are given by:

Table 1. A formal context

U	a	b	c	d	e
1	0	1	0	1	0
2	1	0	1	0	1
3	1	1	0	0	1
4	0	1	1	1	0
5	1	0	0	0	1

$$\begin{aligned} (X_1, B_1) \wedge (X_2, B_2) &= (X_1 \cap X_2, (B_1 \cup B_2)'^*), \\ (X_1, B_1) \vee (X_2, B_2) &= ((X_1 \cup X_2)^{*\prime}, B_1 \cap B_2). \end{aligned} \quad (2)$$

The corresponding partial order relation $\leq$ in the concept lattice $L(U, A, I)$ is given as follows: for $(X_1, B_1), (X_2, B_2) \in L(U, A, I)$,

$$(X_1, B_1) \leq (X_2, B_2) \Longleftrightarrow X_1 \subseteq X_2 \Longleftrightarrow B_1 \supseteq B_2. \quad (3)$$

A formal context can be represented by a table the rows of which are headed by the object names and the columns headed by the attribute names. A value 1 in row x and column a means that the object x has the attribute a.

In this paper,we use the table in [17] for example.

Example 1. Table 1 depicts an example of formal context $F = (U, A, I)$, where $U = \{1, 2, 3, 4, 5\}$, $A = \{a, b, c, d, e\}$, and for each $(x_i, a_j) \in U \times A$, $(x_i, a_j) \in I$ iff object x_i has value 1 in attribute a_j, i.e., $a_j(x_i) = 1$. Fig. 1 is the Hasse diagram of the concept lattice derived from Table 1.

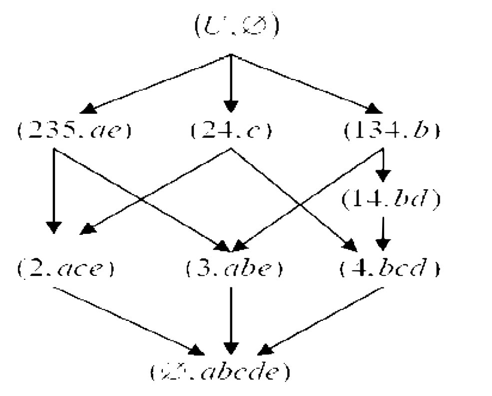

Figure 1. $L(U, \{a, b, c, d, e\}, I)$

In [6],we introduced a new lattice structure called closed label lattice. In the new structure, all concepts are depicted with its intension reduction. In the following, we briefly introduce some concepts of closed label lattice.

let (U, A, I) be a formal context, $Y \subseteq A$, then Y is a closed itemset iff $Y'^{*} = Y$. closed itemset is the maximal set of attributes that shared by some object set. Thus, (1) Every intension of concept in concept lattice is a closed itemset. (2) The set of intension for all concepts in concept lattice equals to the set of closed itemset generated from formal context.

Definition 3. Let (U, A, I) be a formal context. The closed label of C , expressed as $Closedlabel(C)$. For $\forall Y \in Closedlabel(C)$, satisfied:

(1) $Y'^{*} \neq Y$

(2) $Y'^{*} = intension(C)$

(3) $\forall Z \subset Y, Z'^{*} \subset Y'^{*}$

Definition 4. Let (U, A, I) be a formal context. C called a closed label concept, expressed as $(extension(C), Closedlabel(C), intension(C))$.

The set of all closed label concepts forms a closed label lattice and is denoted by $CL(U, A, I)$.

Example 2. The closed label lattice of Table 1 is shown in Fig. 2.

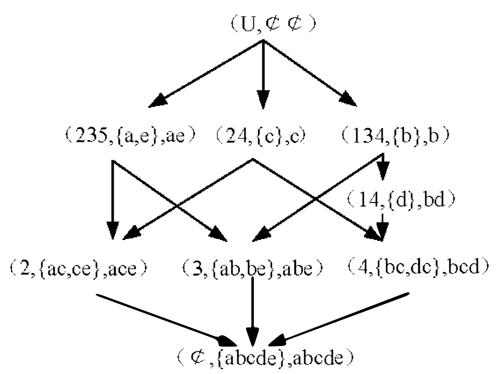

Figure 2. $CL(U, \{a, b, c, d, e\}, I)$

3. Theory of attribute reduction for concept lattices

In section 2, we have briefly reviewed concept lattices for attribute representation in formal contexts and the structure of closed label lattice. In order to obtain the concept lattice with relatively less attributes, attribute reduction has been one of the important issues in concept lattice theory. In this section, we cite the related definitions of attribute reduction in concept lattices[18,19].

Definition 5. Let (U, A, I) be a formal context and $L(U, A, I)$ the associated concept lattice. $L_U(U, A, I)$ is the set of extension of all concepts:

$L_U(U, A, I) = \{X | (X, B) \in L(U, A, I)\}$.

For two concept lattices $L(U, A_1, I_1)$, $L(U, A_2, I_2)$, if $L_U(U, A_1, I_1) = L_U(U, A_2, I_2)$, then $L(U, A_1, I_1)$ is equivalent to $L(U, A_2, I_2)$, and noted as

$$L(U, A_1, I_1) =_U L(U, A_2, I_2) \qquad (4)$$

if $L(U, A_1, I_1) =_U L(U, A_2, I_2)$, obviously, one has that $L(U, A_1, I_1)$ is isomorphic to $L(U, A_2, I_2)$, i.e, $L(U, A_1, I_1) \cong L(U, A_2, I_2)$.

Definition 6. Let $L(U, A_1, I_1)$, $L(U, A_2, I_2)$ be two concept lattice. If for arbitrary concept $(X, B) \in L(U, A_2, I_2)$, there exists $(X', B') \in L(U, A_1, I_1)$ such that $X' = X$ then we write $L(U, A_1, I_1) \leq L(U, A_2, I_2)$.

If $L(U, A_1, I_1) \leq L(U, A_2, I_2)$, thus $L_U(U, A_1, I_1) \supseteq L_U(U, A_2, I_2)$. Furthermore, if we also have $L(U, A_2, I_2) \leq L(U, A_1, I_1)$. then we write $L(U, A_1, I_1) =_U L(U, A_2, I_2)$.

Definition 7. Let (U, A, I) be a formal context. If there exists $D \subseteq A$, such that $L(U, D, I_D) =_U L(U, A, I)$, then D is called a consistent set of (U, A, I). Furthermore, if $L(U, D - \{d\}, I_{D-\{d\}}) \neq_U L(U, A, I)$, for all $d \in D$, then D is called an attribute reduct of U, A, I. The intersection of all the reducts of (U, A, I) is called the core of (U, A, I).

Due to the above discussion, the essential of the approach to attribute reduction is to find the minimal attribute set which can determine a concept lattice isomorphic to the one determined by all attributes.

4. A heuristic method of attribute reduction for concept lattice

From the construction of closed label lattice, we can see that the set of closed labels of every concept are intension reduction for the concept. Therefore, from the point of closed labels, attribute reduction of concept lattice can be obtained easier. In this section, we introduce a heuristic method of attribute reduction for concept lattice based on closed label lattice.

Up to the definition of closed label lattice, we have the following property.

Property 1. Let $F = (U, A, I)$ be a formal context, $L(U, A, I)$ the associated concept lattice and $CL(U, A, I)$ the association closed label lattice. Then $L(U, A, I) \cong CL(U, A, I)$.

Proof. Since by Definition 4 we know that $L_U(U, A, I) = CL_U(U, A, I)$. So, $L(U, A, I) \cong CL(U, A, I)$.

Based on the Property 1, the set of attribute reduct of $CL(U, A, I)$ equals to the set of attribute reduct of $L(U, A, I)$. Q.E.D.

The following theorem shows that the consistent sets of $CL(U, A, I)$ can be determined by closed label of all the concept in $CL(U, A, I)$.

Theorem 1. Let $F = (U, A, I)$ be a formal context and $CL(U, A, I)$ the associated closed label lattice. An attribute subset $D \subseteq A$ is referred to as a consistent set of $CL(U, A, I)$, iff for every concept $C \in CL(U, A, I)$, there exists $Y \in Closedlabel(C)$ such that $Y \subseteq D$.

Proof. ($\Longrightarrow$). If $D \subseteq A$ is referred to as a consistent set of $CL(U, A, I)$, then for every concept $C \in CL(U, A, I)$, there exists concept $C' \in CL(U, D, I_D)$, such that $extension(C) = extension(C')$. Since, $extension(C) = \{Y'|Y \in Closedlabel(C)\}$, we conclude that $extension(C) = extension(C')$ if and only if there exists $Y \in Closedlabel(C)$ such that $Y \subseteq D$.

($\Longleftarrow$). If for every concept $C \in CL(U, A, I)$,there exists $Y \in Closedlabel(C)$ such that $Y \subseteq D$. Thus for every concept $C \in CL(U, A, I)$, there exists a concept $C' \in CL(U, D, I_D)$, such that $extension(C) = extension(C')$. Therefore, $D \subseteq A$ is a consistent set of $CL(U, A, I)$. Q.E.D.

Let $\{D_i|D_i, i \in \tau\}$(τ is a index set) be all reduct of $CL(U, A, I)$, then an element of core satisfied: $b \in \cap_{i\in\tau} D_i$. By using the closed label of concepts, it is easy to obtain the core of $CL(U, A, I)$ as follows:

Theorem 2. Let $F = (U, A, I)$ be a formal context and $CL(U, A, I)$ the associated closed label lattice. An attribute $a \in A$ is an element of the core of $CL(U, A, I)$ iff there exists $C \in CL(U, A, I)$ such that $Closedlabel(C) = \{a\}$.

Proof. ($\Longrightarrow$). If $a \in A$ is an element of the core of $CL(U, A, I)$, then $a \in \cap_{i\in\tau} D_i$. Thus, there exists $C \in CL(U, A, I)$ such that $\{a\} \in Closedlabel(C)$. By the property of closed label, $\{a\} = Closedlabel(C)$.

($\Longleftarrow$). Assume that there exists $C \in CL(U, A, I)$ such that $Closedlabel(C) = \{a\}$, then a included in all the reduct, i.e. $a \in \cap_{i\in\tau} D_i$. Therefore, $\{a\}$ is an element of the core of $CL(U, A, I)$. Q.E.D.

Example 3. From Fig.2,

$Closedlabel((24, \{c\}), c) = c$

$Closedlabel((134, \{b\}), b) = b$

$Closedlabel((14, \{d\}), bd) = d$

So, c, b, d are the element of core. Thus, the core of $CL(U, A, I)$ and $L(U, A, I)$ is $\{c, b, d\}$.

In the following, a heuristic method based on closed label lattice is proposed. In a heuristic method of attributes reduction, there are two key issues, one is core's calculation, the other is measuring of significance of attributes. By Theorem 2, we can directly obtain the element of core by traversing the concepts in $CL(U, A, I)$. we calculate the significance of each of attributes by $|a'|$, $a \in A$, where $|a'|$ denotes the number of objects that shared by a.

Algorithm 1: a heuristic method of attribute reduction for concept lattice

Input: A formal context $F = (U, A, I)$, where $U = \{u_1, u_2, \cdots, u_{|U|}\}$, $A = \{a_1, a_2, \cdots, a_{|A|}\}$, and associated closed label lattice $CL(U, A, I)$

Output: a reduct R

01 Initialize $Core = \emptyset$
02 For every concept $C \in CL(U, A, I)$ begin
03 if $Closedlabel(C)$ has only one element Y and
04 if $Card(Y) = 1$
05 then $Core = Core \cup Y$
06 End
07 Initialize $R = Core$
08 For every concept $C \in CL(U, A, I)$ begin
09 Initialize $r = false$
10 for every element $Y \in Closedlabel(C)$ begin
11 If $Y \subseteq R$
12 Then $r = true$,break
13 End
14 If $r = false$, break
15 End
16 If $r = true$,turn to step 19
17 For $a \in A - R$, calculate $max(|a'|)$
18 Let $R = R \cup a$,turn to step 08
19 Output R
20 End

As follows, we analyze the complexity for computing a reduct by Algorithm 1. We suppose $|C|$ is the number of concepts in closed label lattice. (1) The time complexity for computing $Core$ is $O(|C|)$. (2) The time complexity for computing $|a'|$ is $O(|U||A|)$. (3) The time complexity for testing whether R is a reduct is $O(|C|)$. So, the time complexity of this algorithm is $O(|C| + |A|(|U||A| + |C|)) = O(|U||A|^2 + |A||C|)$.

Example 4. By Example 3, the core of $CL(U, A, I)$ is $\{b, c, d\}$. However, $\{b, c, d\}$ is not a reduct. In fact, for concept $(235, \{a, e\}, ae)$, both a and e are not included in the attribute set $\{b, c, d\}$.

For the attribute a, e in $A - \{b, c, d\}$, we have $|a'|$=3, $|e'|$=3. We choose a added in $\{b, c, d\}$. For every concept, there exists an element of closed label include in $\{a, b, c, d\}$. Therefore, $\{b, c, d\}$ is a reduct.

5. Conclusions

To improve the limitation of time-consuming of existing attribute reductions in concept lattice theory, in the present study, a heuristic algorithm of attribute reduction has been proposed using an expanded concept lattice called closed label lattice. The algorithm has been provided a new point of view for studying attribute reduction in the concept lattice theory. From the proposed algorithm and an illustrative example, we affirm that the algorithm can effectively obtain an attribute reduct from a formal context. In the future, we will develop the corresponding version in formal decision context.

Acknowledgements:

This work was supported by the National Natural Science Foundation of China (Nos. 60773133, 70971080, 60903110), the Natural Science Foundation of Shanxi Province, China (Nos. 2008011038, 2010021017-3, 2009021017-1).

References

[1] Wille. R., "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," Ordered Sets, I. Rival, Ed., Dordrecht-Boston: Reidel, pp. 445–470, 1982.

[2] Godin. R. "Incremental concept formation algorithm based on Galois (concept) lattices", Computational Intelligence, Vol 11, No. 2, pp. 246-267, 1995.

[3] Ho. T. B., "An approach to concept formation based on formal concept analysis", IEICE Transaction on Information and Systems, Vol E78-D, No. 5, pp. 553-559, 1995.

[4] Godin. R., Missaoui. R., "An incremental concept formation approach for learning from databases", Theoretical Computer Science, Special Issue on Formal Metthods in Databases and Software Engineering, Vol 133, pp. 387-419, 1994.

[5] Ho. T. B., "Discovering and using knowledge from unsupervised data", Decision Support System, Elsevier Science, Vol 21, No. 1, pp. 27-41, 1997.

[6] Liang. J. Y., Wang J. H., "A new lattice structure and methor for extracting association rules based on concept lattcice", International Journal of computer Science and Network Security, Vol 6, No. 11, pp. 107-114, 2006

[7] Yao. Y. Y., "A Comparative Study of Formal Concept Analysis and Rough Set Theory in Data Analysis," Proceedings of Rough Sets and Current Trends in Computing 2004, LNAI, Vol 3066, Berlin: Springer, pp. 59–68, 2004.

[8] Yao. Y. Y., Chen. Y. H., "Rough Set Approximations in Formal Concept Analysis," Transactions on Rough Sets, Vol V, LNCS, Vol 4100, pp. 285–305, 2006.

[9] Shao. M. W., Liu. M., Zhang. W. X., "Set Approximations in Fuzzy Formal Concept Analysis", Fuzzy Sets and Systems, Vol 158, No. 23, pp. 2627–2640, 2007.

[10] Hu. K., Sui. Y., Lu. Y., Wang. J., Shi. C., "Concept Approximation in Concept Lattice, "Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, Vol 2035, pp. 167–173, 2001.

[11] Jiang. G. Q., Chute. C. G., "Auditing the Semantic Completeness of SNOMED CT Using Formal Concept Analysis", Journal of the American Medical Informatics Association, Vol 16, No. 1, pp. 89-102, 2009.

[12] Tonella. P., "Using a concept lattice of decomposition slices for program understanding and impact analysis", IEEE Transactions on Software Engineering, Vol 29, No. 6, pp. 495-509, 2003

[13] Qian. Y. H., Liang. J. Y., Pedrycz. W., Dang. C. Y., "Positive approximation: An accelerator for attribute reduction in rough set theory", Artificial Intelligence, Vol 174, No. 9-10, pp. 597-618, 2010.

[14] Wu. W. Z., Zhang. M., Li. H. Z., Mi. J. S., "Knowledge reduction in random information systems via dempster-shafer theory of evedence", Information Sciences, Vol 174, pp. 143-164, 2005.

[15] Qian. Y. H., Liang. J. Y., Dang. C. Y., "Incomplete multigranulation rough set", IEEE Trasactions on Systems, Man and Cybernetics-Part A, Vol 40, No. 2, pp. 420-431, 2010

[16] Qian. Y. H., Liang. J. Y., Dang. C. Y., "Knowledge structure, knowledge granulation and knowledge distance in a knowledge base", International Journal of Approximate Reasoning, Vol 50, No. 1, pp. 174-188, 2009.

[17] Wu. W. Z., Leung. Y., Mi. J. S., "Granular Computing and Knowledge Reduction in Formal Contexts", IEEE Transactions on Knowledge and Data Engineering, Vol 21, No. 10, pp. 1461-1474, 2009

[18] Zhang. W. X., Wei. L., Qi. J. J., "Attribute Reduction Theory and Approach to Concept Lattice," Science in China: Ser. F Information Sciences, Vol 48, No. 6, 713–726, 2005.

[19] Zhang. W. X., Wei. L., and Qi. J. J., "Attribute Reduction in Concept Lattice based on discernbility matrix," in: Proceedings of RSFDGRC 2005, Lecture Notes in Artificial Intelligence, pp. 157-165, 2005

[20] Liu. M., Shao. M. W., Zhang. W. X., Wu. C., "Reduction Method for Concept Lattices Based on Rough Set Theory and Its Application," Computers and Mathematics with Applications, Vol 53, No. 9, pp. 1390–1410, 2007.

[21] Wang. X., Zhang. W. X., "Relations of attribute reduction between object and property oriented concept lattices, Knowledge-Based Systems", Vol 21, pp. 398-403, 2008

[22] Wang. H., Zhang. W. X., "Knowledge reduction in consistant decision formal context", International Journal of Information Technology, Vol 11, No. 5, pp. 322-331, 2005.

[23] Ganter. B., Wille. R., "Formal Concept Analysis", Mathematical Foundations, Berlin: Springer, 1999.