Discernibility Matrix Based Algorithm for Reduction of Attributes

Ruizhi Wang, Duoqian Miao, Guirong Hu
Department of Computer Science and Technology, Tongji University
Shanghai, 201804, P.R.China
wrz977@sohu.com, Miaoduoqian@163.com

Abstract

In rough set theory, it has been proved that finding the minimal reduct of information systems or decision tables is a NP-complete problem. Therefore, it is hard to obtain the set of the most concise rules by existing algorithms for reduction of knowledge. In this paper, the method of finding sub-optimal reduct based on discernibility matrix is proposed. In general, our method is better than existing methods with respect to the minimal reduct. However, we find that existing minimal reduct searching algorithms are incomplete for reduction of attributes in information systems or decision tables. Through analysis, we present a conjecture about the completeness of the minimal reduct algorithm.

1. Introduction

Rough set theory introduced by Prof. Z.Pawlak in 1982 is a new mathematical tool to reason about vagueness and uncertainty. It also provides techniques to reduce knowledge in databases, by which the irrelevant or superfluous knowledge (attributes) can be eliminated according to the learning task without losing essential information about the original data in the databases. As a result of the reduction of knowledge, a set of concise and meaningful rules is produced.

It is well known that an information system or a decision table may usually have more than one reduct. This means the set of rules derived from reduction of knowledge is not unique. In practice, it is always hoped to obtain the set of the most concise rules. Therefore, people have been attempting to find the minimal reduct of information system or decision table, which means that the number of attributes contained in the reduct is minimal. Unfortunately, it has been proven that finding the minimal reduct of an information system or a decision table is a NP-complete problem[2]. So, it is hard to obtain the set of the most concise rules by existing algorithms for reduction of attributes.

In this paper, using discernibility matrix introduced by Skowron and Rauszer, we propose a heuristic algorithm for reduction of attributes. By comparing with existing algorithms, we show that the proposed algorithm is more efficient than existing ones for reduction of consistent decision tables. With respect to the minimal reduct, we find that our method as well as existing algorithms is still incomplete. In section 4.4, we make a conjecture about the completeness of the minimal reduct algorithms.

2. Preliminaries

2.1 Information System and Decision Table

Starting point of rough set theory is a set of data about some objects of interests. Data are usually organized in a form of a table called information system or decision table.

An information system is formally defined as a quart-tuple, denoted by $IS = \langle U, A, V, f \rangle$, where $U = \{x_1, x_2, \cdots, x_n\}$ is a finite, nonempty set of objects called the universe; $A = \{a_1, a_2, \cdots, a_m\}$ is a finite, nonempty set of attributes; $V = \bigcup_{a \in A} V_a$ is the domain of A, V_a is the domain of a; $f : U \times A \to V$ is a mapping from $U \times A$ to V called the information function[1].

Information systems with distinguished decision and condition attributes are called decision tables. A decision table is denoted by $DT = \langle U, C \cup D, V, f \rangle$, where U, V and f are the same as defined above; $A = C \cup D$, C is a set of condition attributes, D is a set of decision attributes, and $C \cap D = \phi$. Each row of a decision table determines a decision rule, which specifies decisions (actions) that should be taken when conditions pointed out by condition attributes are satisfied.

Proceedings of the 2006 IEEE/WIC/ACM International Conference on
Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)
0-7695-2749-3/06 $20.00 © 2006 IEEE

2.2 Discernibility Matrix

Let $IS = \langle U, A, V, f \rangle$ be an information system. By $M(IS)$ we denote an $n \times n$ matrix (c_{ij}), called the discernibility matrix of *IS* [3], such that

$$c_{ij} = \{a \in A : a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, 2, \cdots, n.$$

Let $DT = \langle U, C \cup D, V, f \rangle$ be a decision table. By $M(DT)$ we denote an $n \times n$ matrix (c_{ij}), called the discernibility matrix of *DT*, such that

$$c_{ij} = \begin{cases} \phi, & f_D(x_i) = f_D(x_j); \\ \{a \in A : a(x_i) \neq a(x_j)\}, & f_D(x_i) \neq f_D(x_j) \end{cases}$$

Since $M(IS)$ and $M(DT)$ are symmetric and $c_{ii} = \phi$ for $i = 1, 2, \cdots, n$, we represent $M(IS)$ and $M(DT)$ only by elements in the lower triangle of $M(IS)$ and $M(DT)$, respectively, i.e. the c_{ij} 's with $1 \leq j < i \leq n$.

Using discernibility matrix, Skowron and Rauszer have proven several properties and constructed efficient algorithms related to information systems and decision tables, e.g. dependencies, core, reduct.

The set of all indispensable attributes in *A* is called the core of *IS*, denoted by $CORE(A)$. $CORE(A)$ can be characterized by $M(IS)$ in the following way:

$$CORE(A) = \{a \in A : c_{ij} = \{a\} \quad for\ some\ i, j\}$$

A set $B \subseteq A$ is called a reduct in *A* if *B* is independent in *A* and $IND(B) = IND(A)$. The set of all reducts in *A* is denoted by $RED(A)$.

Theorem 1. Let $IS = \langle U, A, V, f \rangle$ be an information system and let $\phi \neq B \subseteq A$. The following conditions are equivalent[3]:

(1) *B* contains a reduct from $RED(A)$;

(2)For all *i* and *j,* such that

$$c_{ij} \neq \phi \quad and \quad 1 \leq j < i \leq n, \; c_{ij} \cap B \neq \phi.$$

For decision tables, the similar conclusion still holds.

3. A Heuristic Algorithm for Reduction of Knowledge

3.1 Significance of Attributes

For simplifying discussion, let *M* be the matrix of information system *IS* or decision table *DT*, and $R \subseteq A\ or\ R \subseteq C$. By M_R, we denote the matrix from *M* eliminating all the elements whose intersection with set R is nonempty. $p_{M_R}(a)$ represents the times that attribute a appears in matrix M_R. When $R = \phi$, we know that $M_R = M$.

Given $R \subseteq A\ or\ R \subseteq C$, for any attribute $a \in A - R\ (or\ C - R)$, the significance of attribute *a*, denoted by $SGF(a, R, A)\ (or\ SGF(a, R, C))$, can be defined as follows:

$$SGF(a, R, A) = p_{M_R}(a).$$

In general, the bigger $SGF(a, R, A)$ $(or\ SGF(a, R, C))$ is, the more important attribute *a* in *A* (or *C*) is.

3.2 Algorithm for Reduction of Knowledge

Input: Let *M* be the discenibility matrix of an information system *IS* (or decision table *DT*).

Output: One reduct of *IS* (or *DT*).

Step1: Set $R = \phi$.

Step2: Compute the core of *IS* (or *DT*) from *M*, denoted by *Co*.

Step3: $R \Leftarrow Co$.

Step4: Construct matrix M_R from *M*.

Step5: For any $a \in A - R\ (or\ C - R)$, compute $SGF(a, R, A)\ (or\ SGF(a, R, C))$ from M_R. Let

$$SGF(a', R, A) = \max_{a \in A - R} \{SGF(a, R, A)\}.$$

Step6: $R \Leftarrow R \cup \{a'\}$.

Step7: The algorithm stops and outputs *R*, when $M_R = \phi_{n \times n}$. Otherwise, go to Step 4.

4. An Illustrative Example

4.1 Several Existing Algorithms for Reduction of Knowledge

There are three kinds of algorithms for reduction of knowledge in existing papers. They are all similar to the proposed algorithm in this paper, but the significance of attributes.

(1) **Algorithm 1** [5]

$$SGF(a, R, A) = k(R \cup \{a\}, A) - k(R, A)$$

where $k(R, A) = \dfrac{cardPOS_R(A)}{cardPOS_A(A)}$.

(2) **Algorithm 2** [5]

$$SGF(a, R, A) = \gamma_{R \cup \{a\}}(A) - \gamma_R(A),$$

where $\gamma_R(A) = \dfrac{cardPOS_R(A)}{cardU}$.

(3) **Algorithm 3** [5]

Proceedings of the 2006 IEEE/WIC/ACM International Conference on
Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)
0-7695-2749-3/06 $20.00 © 2006 IEEE

$SGF(a,R,A)=H(a|R)$,

where $H(a|R)$ represents the conditional entropy attribute a with respect to R.

Conducting simple computation, we have proven that these two algorithms (1) and (2) are equivalent.

4.2 An Example

As shown in table 1, the information system $IS=\langle U,A,V,f\rangle$ is given in this example, where $U=\{1,2,\cdots,18\}$, $A=\{a,b,c,d\}$. We know that $Co=\phi$ and the minimal reduct $R=\{c,d\}$. However, using algorithm (1), (2), (3) and the algorithm proposed in this paper, we get the same reduct of the information system— $R=\{a,b,c\}$ or $R=\{a,b,d\}$. They are not the minimal reduct.

4.3 Refinements of the proposed Algorithm

The refined algorithm is:

Input: Let *M* be the discenibility matrix of an information system *IS* (or decision table *DT*).

Output: One reduct of *IS* (or *DT*).

Step1: Set $R=\phi$

Step2: Delete all duplicate elements from *M*, i.e. every element appears only one time in *M*.

Step3: Compute the core of *IS* (or *DT*) from *M*, denoted by *Co*.

Step4: $R\Leftarrow Co$.

Step5: Construct matrix M_R from *M*.

Step6: Classify elements in M_R in terms of their cardinal number, the class that its cardinal number equals *i* is denoted by L_i, $i=1,2,\cdots m$.

Step7: Let $i_0=\min\limits_{L_i\neq\phi}\{i\}$, for any $a\in A-R$ $(or\ C-R)$, compute $SGF(a,R,A)|L_{i_0}$ $(or\ SGF(a,R,C)|L_{i_0})$ from L_{i_0}. Let

$$SGF(a',R,A|L_{i_0})=\max_{a\in A-R}\left\{SGF(a,R,A)|L_{i_0}\right\}.$$

Step8: $R\Leftarrow R\cup\{a'\}$.

Step9: The algorithm stops and outputs *R*, when $M_R=\phi_{n\times n}$. Otherwise, go to Step5.

For this example, we have

1) The elements of discernibility matrix M of the information system are: ac(1), ad(1), bc(1), bd(1), cd(1), abc(2), abd(2), abcd(144), where the number in ( ) represents the times of the element appearing in M.

2) It is easy to see that $Co=\phi$.

3) Deleting all duplicate elements from *M*, and Classifying elements in M_R in terms of their cardinal number, we have

$L_1=\{\phi\}$, $L_2=\{ac,ad,bc,bd,cd\}$,

$L_3=\{abc,abd\}$, $L_4=\{abcd\}$.

4) $R=\phi$, compute the times of every element in L_2:

$SGF(a,R,A)|L_2=2$, $SGF(b,R,A)|L_2=2$, $SGF(c,R,A)|L_2=3$, $SGF(d,R,A)|L_2=3$.

5) Because $SGF(c,R,A)|L_2=SGF(d,R,A)|L_2=\max$, and they are equal in L_3 and L_4. So, we could select any one from attributes c and d, might as well c.

6) $R=\{c\}$, then $L_1=\{\phi\}$, $L_2=\{ad,bd\}$, $L_3=\{abd\}$, $L_4=\{\phi\}$.

7) In the next circle, $2=\min\limits_{L_i\neq\phi}\{i\}$, so we compute the times of every element in L_2. $SGF(a,R,A)|L_2=1$, $SGF(b,R,A)|L_2=1$, $SGF(c,R,A)|L_2=0$, $SGF(d,R,A)|L_2=2$.

8) Attribute *d* is selected, and $R=\{c,d\}$. In that time, $M_R=\phi_{n\times n}$. Therefore, the result is $R=\{c,d\}$. It is the minimal reduct.

4.4 A conjecture about the completeness of the minimal reduct algorithms

As to consistent decision tables, the discernibility matrix definition introduced by A.Skowron is correct, and the proposed algorithm is feasible. However, if the decision table is inconsistent, the definition has to be rewritten. YE Dong-yi et al.[6] proposed a new discernibility matrix definition for inconsistent decision table and presented a minimal reduct algorithms of attributes based on the new discernibility matrix. More detail could be found in [6]. Generally, the existing algorithms for reduction of attributes are incomplete with respect to the minimal reduct. Through analysis, we make a conjecture as follows:

Conjecture 1 If a given decision table is consistent and its core is nonempty, then the proposed algorithm is complete with respect to the minimal reduct.

Our future research will concentrate on testifying the conjecture. Moreover, we hope more researchers will join us to deal with the problem.

5. Conclusion

In this paper, the method of finding sub-optimal reduct based on discernibility matrix is proposed. Through an example, we show that the proposed algorithm outperforms other algorithms for reduction of attributes with respect to the minimal reduct generally. In addition, we find that our algorithm together with other algorithms fails to find the minimal reduct of information systems or decision tables. Finally, we make a conjecture about the completeness of the minimal reduct algorithms and will focus our research interests on testifying the conjecture in the future.

References

[1] Pawlak, Z., *Rough Sets - Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht ,1991.

[2] Ziarko, W.P.,"The Discovery, Analysis, and Representation of Data Dependencies in Databases", Piatesky-Shapiro, G. and Frawley, W.J. (eds.), *Knowledge Discovery in Databases*, AAAI Press/MIT Press, 1991, pp. 177-195.

[3] Skowron, A., Rauszer, C., "The discernibility matrices and functions in information systems", Sl/owi nski, R., editor, *Intelligent decision support: Handbook of applications and advances of rough set theory*, Kluwer Academic Publishers, Dordrecht, volume 11, 1992, pp. 331-362.

[4] Miao, D.Q., *Rough Sets and Its Application in Machine Learning*[D], Institute of Automation Chinese Academy of Sciences, Beijing, 1997.

[5] Miao, D.Q., Wang J., "Analysis on Attribute Reduction Strategies of Rough Set", *Chinese Journal of Computer Science and Technology*, 1998, 13(2): 189-192.

[6] Ye, D.Y., Chen, Z.J., "A New Discernibility Matrix and the Computation of a Core", *Chinese Journal of Electronics*, 2002, 30(7):1086-1088.

Table1. An Information System

U	A			
	a	B	c	d
1	1	1	1	1
2	2	2	1	2
3	3	3	3	3
4	4	4	3	4
5	5	5	5	5
6	6	5	5	6
7	7	7	7	7
8	7	8	7	8
9	9	9	9	9
10	9	9	10	10
11	11	11	11	11
12	12	12	12	11
13	13	13	13	13
14	14	14	14	13
15	15	15	15	15
16	16	15	16	15
17	17	17	17	17
18	17	18	18	17

Proceedings of the 2006 IEEE/WIC/ACM International Conference on
Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops)(WI-IATW'06)
0-7695-2749-3/06 $20.00 © 2006 IEEE