Analysis on Attribute Reduction Strategies of Rough Set

Wang Jue (王 珏) and Miao Duoqian (苗夺谦)

Institute of Automation, Chinese Academy of Sciences, Beijing 100080, P.R. China

Received May 29, 1997; revised September 16, 1997.

Abstract

Several strategies for the minimal attribute reduction with polynomial time complexity ($O(n^k)$) have been developed in rough set theory. Are they complete? While investigating the attribute reduction strategy based on the discernibility matrix (DM), a counterexample is constructed theoretically, which demonstrates that these strategies are all incomplete with respect to the minimal reduction.

Keywords: Rough set, minimal attribute reduction.

1 Introduction

As a basic problem in rough set (RS) theory[1], the attribute reduction is related to such applications as knowledge discovery in databases (KDD). Attribute reduction (referred to as reduction in this paper) is to remove superfluous attributes from information systems (information tables or decision tables) while preserving the consistency of classifications. The reduction with the minimal number of attributes for a given database is called the minimal reduction. If there exists a strategy which can find the minimal reduction for any given database, this strategy is said to be complete for the minimal reduction.

From RS theory, it is obvious that there exists the minimal reduction for databases which are represented as information systems, and it is already proved that the problem of finding the minimal reduction without strategies is NP-hard[2]. Now, several strategies based on different considerations with time complexity of $O(n^k)$ (n is the number of attributes in information systems) have been developed[3–5], and they can obtain the minimal reduction in plenty of experiments. Therefore, they are surmised to be complete for the minimal reduction.

When we investigate the reduction strategy based on the discernibility matrix (DM)[6], a new definition of the attribute significance is presented, which is based on the number of each attribute symbol appearing in DM, called attribute frequency function, and its time complexity is $O(n^2)$. We have attempted to prove the completeness of this strategy with respect to the minimal reduction but failed to do so. However, a counterexample is constructed theoretically, which demonstrates that all strategies listed in this paper are incomplete.

2 Attribute Significance

Actually, the reduction strategy depends on the definition of the attribute significance. Some definitions on the attribute significance reported presently are listed below:

Research supported by '863' Program and NNSF of China.

Let $S = \langle U, A, V, f \rangle$ be an information system, where U is the universe, A is the set of attributes on $U, V_n \in V$ is the value set of the attribute a in A, and f is an information function. C and D are the sets of condition and decision attributes, respectively[1].

(1) Define the attribute significance according to the roughness of attributes.

In [3], the significance of attribute a is defined as:

$$SGF(a, R, D) = K(R \cup \{a\}, D) - K(R, D) \tag{1}$$

where $K(R, D) = \text{card}(POS_R(D))/\text{card}(POS_C(D))$, $POS_R(D)$ is the positive region of R with respect to D, $K(\emptyset, D) = 0$[1], and card(E) is the cardinality of set E.

In [4], the significance of attribute a is defined as:

$$SGF(a, R, D) = \gamma_{R\cup\{a\}} - \gamma_R \tag{2}$$

where $\gamma_R = \text{card}(POS_R(D))/\text{card}(U)$, and if $R = \emptyset$, then $\gamma_R = 0$.

For information tables, $D = C = A$ will hold if there are not the same instances in database.

(2) Define the attribute significance according to information entropy[5].

Let $H(a/R)$ be the condition entropy of the attribute a for the attribute set R. The attribute significance of information table and decision table can be respectively defined as:

$$SGF(a, R) = H(a/R), \quad \text{and} \quad SGF(a, R, D) = H(D/R) - H(D/R \cup \{a\})$$

(3) Define the attribute significance according to attribute frequency function in DM.

Let M be DM of an information system S, which removes all the elements that have non-empty intersection with the set of core attributes R, and $R \subset A$. $p(a)$ is the number of appearing times of attribute a in M, called attribute frequency function[2]. Then for any attribute $a \in A - R$, the significance $SGF(a, R)$ is defined as:

$$SGF(a, R) = p(a)$$

3 Binary Tree Representation of DM

By the definition of DM, each term of DM is a subset of the attribute set of information systems. Whether the attribute a is in a term or not is represented as a or $\sim a$, thus a term in DM can be represented as a branch in a complete binary tree with respect to the set of attributes in information systems. For example, the branch of the leaf marked as K_3 in the tree below represents the term cd.

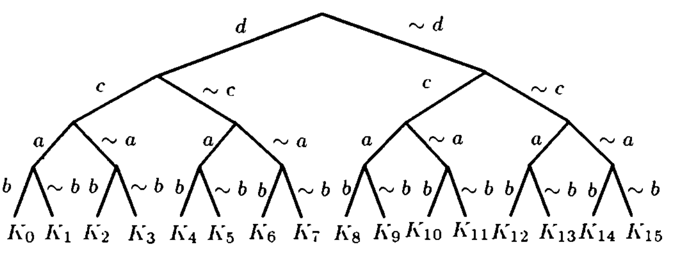

Fig.1. Binary tree representing DM.

[1] Our discussions involve only the cases in which the number of classification is greater than 1 for decision tables.

[2] For decision tables, all elements on the same equivalence class are removed from M.

The marks on the leaves represent the number of the terms appearing in DM, and the core attributes are not considered because the set of core attributes is certain for a given information system. Thus, $K_7 = K_{11} = K_{13} = K_{14} = 0$, and $K_{15} = 0$ in the above tree.

With such a tree, the above reduction strategy based on DM is thus explained: if an attribute d has the maximum $p(d)$, then we construct a tree on which d is put nearest to the root. To cut the left half of the tree is to delete all the terms with the attribute d from DM, and d is a reduction attribute. Using the attribute frequency function, cut the branches successively till all branches in the tree are cut, and the reduction set R will be acquired.

4 Analysis on Reduction—A Counterexample

According to the binary tree representation of DM and the explanation of the above reduction strategy on binary tree, if a reduction with fewer attributes than the number of attributes in an information system is obtained, it means all leaves of some subtrees in the right half of the tree are marked 0. The minimal reduction will then be a tree thus constructed: the right half in it has the greatest number of subtrees whose leaves are marked 0. For example, consider the set with 4 attributes $\{a, b, c, d\}$, the corresponding binary tree is shown as Fig.1. In the tree, $K_7, K_{11}, K_{13}, K_{14}$ are the marks of core attributes, K_{15} is a null term and K_0 is a full term (that is, term *abcd* in the tree in Fig.1). All these terms are not considered.

If $K_{12} = 0$, that is, the term *ab* doesn't exist in the DM, and $K_3 = K_5 = K_6 = K_9 = K_{10} = 1$, then the minimal reduction must be $\{c, d\}$ according to the above binary tree explanation of the reduction. If the above reduction strategy is complete, d or c should have the greatest attribute frequency in M, then the following inequalities should hold.

$$\begin{aligned} p(d) > p(c), &\quad \text{i.e.,} \quad K_4 + K_5 + K_6 > K_8 + K_9 + K_{10} \\ p(d) > p(a), &\quad \text{i.e.,} \quad K_2 + K_3 + K_6 > K_8 + K_9 + K_{12} \\ p(d) > p(b), &\quad \text{i.e.,} \quad K_1 + K_3 + K_5 > K_8 + K_{10} + K_{12} \end{aligned}$$

According to the assignments to K_i, we have $K_4 > K_8$, $K_2 + 1 > K_8$, $K_1 + 1 > K_8$.

Obviously, assuming $K_1 = K_2 = 0$ and $K_4 = K_8 = 2$, all the above inequalities will not hold, but the minimal reduction is still $\{c, d\}$.

By the above analysis, a set of terms in DM can be constructed as below (since the terms like *abcd* don't affect the order of the attribute significance, DM doesn't include them though they exist in the DM).

$$\{abd, abd, ad, bd, cd, abc, abc, ac, bc\}$$

Now, $p(a) = p(b) = 6$, and $p(c) = p(d) = 5$. The significance of attributes is ordered as $a = b > c = d$. According to the reduction strategy based on attribute frequency function illustrated in Section 2, a or b must be selected as the reduction attribute. Now the final reduction must be $\{a, b, c\}$ or $\{a, b, d\}$ rather than the minimal reduction $\{c, d\}$. It shows that the reduction strategy is not complete with respect to the minimal reduction. By the above set of terms in DM the following information table can be constructed:

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	1	2	3	4	5	6	7	7	9	9	11	12	13	14	15	16	17	17
b	1	2	3	4	5	5	7	8	9	9	11	12	13	14	15	15	17	18
c	1	1	3	3	5	5	7	7	9	10	11	12	13	14	15	16	17	18
d	1	2	3	4	5	6	7	8	9	10	11	11	13	13	15	15	17	17

Such counterexample is constructed according to the reduction strategy based on the attribute frequency in DM. Interestingly, it can also be taken as a counterexample for all reduction strategies listed in Section 2.

5 Analysis on the Completeness of the Other Two Kinds of Strategies

For above counterexample, $U = \{1, 2, 3, \ldots, 18\}$, $A = \{a, b, c, d\}$, the set of the core $C_0 = \emptyset$. Since this counterexample has no decision attribute and there are not the same instances, it corresponds to a decision table where $C = D = A$. The significance of each attribute in the strategy (1) (see Section 2) can be computed as below[3]. Since

$$U/D = \{\{1\}, \{2\}, \{3\}, \ldots, \{18\}\}$$
$$U/a = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7, 8\}, \{9, 10\}, \{11\}, \{12\}, \{13\}, \{14\}, \{15\}, \{16\}, \{17, 18\}\}$$

we have

$$POS_a(D) = \{1, 2, 3, 4, 5, 6, 11, 12, 13, 14, 15, 16\}$$

and

$$C = D,\ POS_C(D) = U$$

Then

$$SGF(a, C_0, D) = \text{card}(POS_a(D))/\text{card}\ (POS_C(D)) = 12/18$$

Similarly,

$$SGF(b, C_0, D) = 12/18$$
$$SGF(c, C_0, D) = SGF(d, C_0, D) = 10/18.$$

The significance of attributes is ordered as: $a = b > c = d$. Therefore, in its first step, the strategy based on roughness must select a or b as the reduction attribute, thus the final reduction must be $\{a, b, c\}$ or $\{a, b, d\}$, rather than the minimal reduction $\{c, d\}$.

In strategy (2) (see Section 2), the attribute significance is thus computed:

$$H(a/C_0) = H(a) = -\left[\frac{12}{18}\log\frac{1}{18} + \frac{6}{18}\log\frac{2}{18}\right] = 1.1549$$

and $H(b) = H(a)$;

$$H(c/C_0) = H(c) = -\left[\frac{10}{18}\log\frac{1}{18} + \frac{8}{18}\log\frac{2}{18}\right] = 1.1215$$

and $H(d) = H(c)$.

The significance of attributes is ordered as: $a = b > c = d$. Therefore the strategy based on information entropy will also select a or b as the reduction attribute in its first step, and the final reduction is still $\{a, c, d\}$ or $\{b, c, d\}$, rather than the minimal reduction $\{c, d\}$.

It manifests that the above two kinds of strategies are also incomplete with respect to the minimal reduction.

[3] Strategy (2) will obtain the same results as (1) for the counterexample.

6 Conclusion

In this paper several reduction strategies with polynomial time complexity are analyzed and a counterexample is theoretically constructed, which demonstrates that all the strategies listed in this paper are incomplete with respect to the minimal reduction. The conjecture that these reduction strategies are complete with respect to the minimal reduction is therefore proved to be not true. To find the strategies of the minimal reduction with time complexity of $O(n^k)$ is still an open problem. However, the approach to constructing counterexample, which is presented in the paper, can serve as a tool which further explores the properties of reduction strategies.

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

[2] Wong S K M, Ziarko W. On optimal rules in decision tables. *Bulletin of the Polish Academy of Sciences, Mathematics,* 1985, 33: 693–696.

[3] Hu X H, Cercone N. Learning in relational databases: A rough set approach. *Int. J. of Computational Intelligence,* 1995, 11: 323–338.

[4] Jelonek J *et al.* Rough set reduction of attributes and their domains for neural networks. *Int. J. of Computational Intelligence*, 1995, 11: 339–347.

[5] Miao D Q, Wang J. An information based algorithm for reduction of knowledge. Technical Report, Institute of Automation, Chinese Academy of Sciences, March, 1996.

[6] Skowron A, Rauszer C. The discernibility matrices and functions in information systems. Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory, 1992, 331-362.

Wang Jue is a Professor in Institute of Aotomation, Chinese Academy of Sciences. His research interests include knowledge representation, expert system, ANN, GA, muti-agent system, machine learning and data mining.

Miao Duoqian received his Ph.D. degree from Institute of Automation, Chinese Academy of Sciences in 1997. He works in Dept. of math., Shanxi University. His research interests are rough set theory and machine learning.