An Optimal Algorithm for Attribute Reduction Based on Discernibility Matrix

Jiang Yu, Du Bin, Wang Peng, Liu Yintian
College of Software Engineering,Chengdu University of Information Technology, Chengdu 610225,China;
e-mail jiangyu@cuit.edu.cn

Abstract: Many attribute reduction algorithms based on the discernibility matrix with polynomial complexity have been developed in rough set theory. However, those algorithms are all incomplete with respect to the smallest or smaller reduction. Investigating the attribute reduction based on the attribute frequency function $p(a_k)$,which is defined as the number of occurrences of attribute a_k in discernibility matrix, demonstrates that there is a fatal disadvantage of the traditional definition of the attribute frequency function $p(a_k)$: A_i ($A_i \in As, \exists A_j \in As, A_j \subseteq A_i$,where $i \neq j$) increases the number of occurrences of the condition attributes in A_i , which enhance the significance of those attributes. However, A_i is redundancy. Redefine the attribute frequency function $p(a_k)$ based on discernibility matrix. A optimal algorithm for attribute reduction in decision table based on discernibility matrix is introduced. This algorithm using the iteration is to select the indispensable condition attribute according to the attribute frequency function $p(a_k)$, and adds the indispensable condition attribute into the reduction R. In every iteration, this algorithm not only can guarantee that it chooses the condition attribute a_k as reduction attribute that must appear the most times in discernibility set but can guarantee that it will choose the condition attribute a_p that also appears the most times in discernibility set in next iteration. Therefore, this algorithm can ensure that the solution it finds out is an optimal reduction. The time complexity of it in the worst case is analysed and the proof of its completeness is given.

Keywords: rough set; discernibility matrix;attribute significance; optimal reduction;complete algorithm

1 Introduction

Rough set theory proposed by Pawlak[1] is a new mathematical approach to deal with vague and uncertain information. In recent years,it has successfully been applied to fields as data mining、pattern recognition and process control[2].In this theory,attribute reduction is one of the most important parts,which can remove the redundancy and incompatibility attributes so that we can obtain the key information and make the decision rule.

Generally speaking,the reduction of the decision table is not unique,and people hope to find out a smallest reduction.However,S.K.M Wong and W.Ziarko proved that finding out a smallest reduction is NP-hard[3].Now,there are many algorithms having the multinomial time complexity.But,they are incomplete for smaller or smallest reduction.

In this paper, we analyze the disadvantage of the traditional definition of the attribute frequency function $p(a_k)$,which is defined as the number of occurrences of the condition attribute a_k in discernibility set As. we redefine the significance of attribute based on discernibility matrix. A optimal algorithm for attribute reduction in decision table based on discernibility matrix is introduced. This algorithm using the iteration is to select the indispensable condition attribute. In every iteration, this algorithm not only can guarantee that it chooses the condition attribute a_k as reduction attribute that must appear the most times in discernibility set but can guarantee that it will choose the condition attribute a_p that also appears the most times in discernibility set in next iteration. Therefore, this algorithm can ensure that the solution it finds out is probable a smaller reduction.

Supported by the National Natural Science Foundation of China (Grant No. 60702075).

2 Basis Concept

Decision table S=(U,C,D,V,f), Here U is the universe of objects,while C and D are the condition attribute set and the decision attribute set respectively.With every attribute $r \in C \cup D$,the set of values V is associated.Each attribute has a determinant function $f:U\times(C \cup D)\rightarrow V$.U、C、D and V all are non-empty and finite set,and $C \cap D=\varnothing$.Table 1 is a decision table.

Table 1 An example of decision table

U	a	b	c	d
x1	2	2	0	1
x2	1	2	0	0
x3	1	2	0	1
x4	0	0	0	0
x5	1	0	1	0
x6	2	0	1	1

Given |U|=n，discernibility matrix of the decision table is a matrix that contains n×n elements.Each element of the discernibility matrix, or DM for short, is defined as $a^*(x,y)=\{a\in C\mid a(x)\neq a(y)\wedge \mathrm{w}(x,y)\}$.

Where $\forall x,y\in U$, $w(x,y)$ is defined as follows[4]:

$x\in pos_C(D)\wedge y\notin pos_C(D)$

Or $x\notin pos_C(D)\wedge y\in pos_C(D)$

Or $x,y\in pos_C(D)\wedge (x,y)\notin ind(D)$

Table 2 The discernibility matrix corresponding to ta-

	1	2	3	4	5	6
1	∅	{a}	{a}	{a,b}	{a,b,c}	∅
2	{a}	∅	∅	{a,b}	{b,c}	{a,b,c}
3	{a}	∅	∅	{a,b}	{b,c}	{a,b,c}
4	{a,b}	{a,b}	{a,b}	∅	∅	{a,c}
5	{a,b,c}	{b,c}	{b,c}	∅	∅	{a}
6	∅	{a,b,c}	{a,b,c}	{a,c}	{a}	∅

Definition 1 Discernibility function is defined as: $\triangle^*=\bigwedge_{x,y\in U}\bigvee a^*(x,y)$.

Definition 2 discernibility set As of decision table S is the set including all non-empty elements of DM,where DM is the discernibility matrix of the decision table S=(U,C,D,V,f).

According to above definitions,if As={{a,d},{a,b,c},{b,c,d}},then $\triangle^*(As)=(a\vee d)\wedge(a\vee b\vee c)\wedge(b\vee c\vee d)=ab\vee ac\vee ad\vee bd\vee cd$. that is ,{a,b}、{a,c}、{a,d}、{b,d} and {c,d} are the reductions of the decision table S.

3 Attribute Significance Analysis based on discernibility matrix

At present，Many attribute reduction algorithms are based on discernibility matrix.In [5], let $p(a_k)$ be the attribute frequency function of attribute a_k in discernibility matrix, which is defined as the number of occurrences of attribute a_k. $p(a_k)$ is regarded as attribute significance of the attribute a_k .However,investigating the attribute reduction based on the attribute frequency function of attribute a_k in discernibility matrix demonstrates that these strategies are all incomplete with respect to the smallest reduction.Why? Here is an example.

Example 1 Given As ={{a,b,d},{a,b,d},{a,d},{b,d},{c,d},{a,b,c},{a,b,c},{a,c},{b,c}}. 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 [5], 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 smallest reduction {c,d}. It shows that the definition of the attribute frequency function $p(a_k)$ in As is not complete with respect to the smallest reduction. Because the discernibility set As ={{a,b,d},{a,b,d},{a,d},{b,d},{c,d},{a,b,c},{a,b,c},{a,c},{b,c}} and the discernibility set As' ={{a,b,d},{a,d},{b,d},{c,d},{a,b,c},{a,c},{b,c}} have the same attribute reduction sets. Therefore, the copies of {a,b,d} and {a,b,c} in As are redundancy, but these copys increase the number of occurrences of the attribute a、b、c and d, which enhance the significance of the attribute a、b、c and d.

The attribute frequency function $p(a_k)$ defined as the number of occurrences of a_k in As' is complete with respect to the smallest reduction? The answer is negative.

In As' ={{a,b,d},{a,d},{b,d},{c,d},{a,b,c},{a,c},{b,c}}, Now, p(a)=p(b)=p(c)=p(d)=4.The significance of attributes is ordered as a=b=c=d. According to the reduction strategy based on attribute frequency function illustrated in [5], if a or b is selected as the reduction attribute. The final reduction must be {a,b,c} or {a,b,d} rather than the smallest reduction {c,d}. It also shows that the definition of the attribute frequency function $p(a_k)$ in As' is not complete with respect to the smallest reduction. Why? According to the absorption laws of "∧", if $A_i\subseteq A_j$, then $\triangle^*(\{A_i,A_j\})=\triangle^*(\{A_i\})$, therefore, {a,b,d} and {a,b,c} in As' are redundancy, but they increase the number of occurrences of the attribute a、b、c and d, which enhance the significance of the attribute a、b、c and d. In other words, the discernibility set As'={{a,b,d},{a,d},{b,d},{c,d},{a,b,c},{a,c},{b,c}} and the discernibility set As''={{a,d},{b,d},{c,d},{a,c},{b,c}} have the same attribute reduction sets.

In a word, According to analysis on the definition about attribute frequency function, we know that there are a large amount of elements in decernibility set As, which enhance the significance of attributes, but they are redundancy. Now, we will redefine the attribute frequency function $p(a_k)$.

Definition 3 Function $f_{\not\subset}(As)$ can be described as follows：Given $As=\{A_1,A_2,\ldots,A_m\}$， $As'=As$

Function $f_{\not\subset}(As)$

```
{ for(i=1;i<=m;i++)
    for(j=1;j<=m;j++)
      {
        if((i≠j) && (Ai⊆Aj))  As' = As' − Aj ;
      }
  return  As' ;
}
```


978-1-935068-01-3 © 2009 SciRes.

Obviously, $\triangle^*$ (As)=$\triangle^*$ ($f_{\not\subset}(As)$),and inexistence $As'' \subset f_{\not\subset}(As)$,$\triangle^*$ (As)=$\triangle^*$ (As'').

Definition 4 Let $p(a_k)$ be the attribute frequency function in $f_{\not\subset}(As)$ .It is defined as the number of occurrences of attribute ' a_k ' in $f_{\not\subset}(As)$.

4 Algorithm and Complexity Analysis

Definition 5 operation $\partial_B(As)$ can be definition as: $\forall$ a_k $\in$B , find p(a_k) in As.$\exists$ a_p $\in$B, $\forall$ a_k $\in$B, $p(a_k) \le p(a_p)$, let $\partial_B(As) = p(a_p)$. Especially, if B=∅ or As=∅, then $\partial_B(As) = 0$.

Based on the significance of attribute, a method of attribute reduction is presented. The reduction process is as follows:

Algorithm 1: an Optical Reduction Algorithm Based on the Attribute Frequency Function or ORABAFF for short.

Input: Discernibility set As .

Output:A optical attribute reduction set R.

(1) Construct the attribute reduction R,R⇐∅;empty set B、E、F、G、P、T; integer I;

(2) Construct the dicernibility set A;A⇐ $f_{\not\subset}(As)$;

(3) If A=∅,then end algorithm and output R;

(4) $\forall$ a_k $\in$C − (R∪T),find $p(a_k)$ in A,let $B \Leftarrow \{a_k \mid \forall a_p \in \mathrm{C}-(\mathrm{R}\cup\mathrm{T}), \mathrm{p}(a_p) \le \mathrm{p}(a_k)\}$, and E⇐B;

4.1 a_k $\in$E, Let G⇐{D|D∈A∧D∩{ a_k }=∅}; I⇐$\partial_{B-\{\alpha_\kappa\}}(G)$; F⇐{ a_k }; E⇐E−{ a_k }

4.2 if E=∅ then goto(5);

4.3 a_k $\in$E, Let G⇐{D|D∈A∧D∩{ a_k }=∅};

4.4 if I< $\partial_{B-\{\alpha_\kappa\}}(G)$ then I⇐ $\partial_{B-\{\alpha_\kappa\}}(G)$ and F⇐{ a_k };

4.5 E⇐E−{ a_k };goto 4.2;

(5) R⇐R∪F, E⇐{D'| D'=D − F , D∈A∧D∩ F≠∅}, G⇐{D|D∈A∧D∩ F =∅};

5.1 E_i∈E, let I⇐$\partial_{E_i}(G)$ and P⇐ E_i; E⇐E−{E_i};

5.2 If E=∅ then goto(6);

5.3 E_i∈E, if I>$\partial_{E_i}(G)$,then I⇐$\partial_{E_i}(G)$ and P⇐ E_i;

5.4 E⇐E−{E_i}; goto 5.2

(6) T⇐T∪P,

$A \Leftarrow \{D' \mid D' = D - T, D \in A \wedge D \cap R = \emptyset\}$,

A⇐ $f_{\not\subset}(A)$, goto(3);

According to definition of decision table, U and C all are finite set.Given |U|=n,|C|=m,here m and n are positive integer.In the worst case, $|As| = {(n^2-n)}/{2} = N$ .The time complexity of computing $f_{\not\subset}(As)$ is O(mN^2). The time complexity of computing $\partial_B(As)$ and finding $p(a_k)$ in A are O(mN) ,The time complexity of computing $A \Leftarrow \{D' \mid D' = D - P, D \in A \wedge D \cap R = \emptyset\}$ is O(mN) too.The times of the algorithm running is at most m.Therefore,The time complexity of the algorithm in the worst case is as follows:

$$O(m(mN^2 + mN + m*mN + mN + m*mN + mN))$$
$$= O(m^2N(N + 3 + 2m))$$
$$= O\left|m^2 \frac{n^2-n}{2}(\frac{n^2-n}{2} + 3 + 2m)\right|$$
$$= O(m^2n^4 + m^3n^2)$$

5 Algorthim Proof

5.1 Algorithm Completeness Proof

By Pawlak, if P is the reduction of the condition attribute set C,here P⊆C,then P must satisfy two conditions: ① $pos_p(D)=pos_C(D)$;② ∀a∈P,$pos_{p-\{a\}}(D) \neq pos_C(D)$.P is called a Pawlak reduction, otherwise a reduction. If an algorithm ensures that the solution it finds out is a Pawlak reduction for any decision table, this algorithm is completeness.In [6],they redefine the completeness of the algorithm for Pawlak reduction by decernibility matrix.It can be described as follows:

(1) if ∀A∈As,A∩R≠∅;

(2) if ∀a∈R,∃A∈As,A∩R−{a}=∅;

Therefore,The algorithm is complete if and only if the reduction it finds out must satisfy above two conditions.

According to algorithm 1,the change sequence of decernibility set A is $A_0,\ldots,A_i,\ldots,A_k$. and A_0= $f_{\not\subset}(As)$, A_0⊆As, A_k, k≤|C|, where C is the condition attribute set. If i≤j then ∀ A' ∈A_j,∃ A'' ∈A_i, $A' \subseteq A''$ .The change sequence of reduction set R is $R_0,\ldots,R_i,\ldots,R_k$, and R_0=∅, R_k = R, $R_0 \subset \ldots \subset R_i \subset \ldots \subset R_k$,$|R_{i+1}|=|R_i|+1$. $T_0,\ldots,T_i,\ldots,T_k$, which are the change sequence of set T, and $T_0 \subseteq \ldots \subseteq T_i \subseteq \ldots \subseteq T_k$.

Proof:(1)Assume ∃ A' ∈As, A' ∩ R=∅ , which means that ∃ A'' ∈ A_0, $A'' \subseteq A'$, A'' ∩R=∅, and A'' ⊆T. that is , in some iteration (T_i−T_{i-1})⊆ A'' and (T_i−T_{i-1})∈A_i .However, according to algorithm 1, in this iteration $(T_i - T_{i-1}) \cup (R_i - R_{i-1}) \in A_i$. According to definition 3 about function $f_{\not\subset}()$,if $\forall B_i$、$B_j \in A_i$ and $i \neq j$, then $B_i \not\subset B_j$ and $B_j \not\subset B_i$. Therefore, $(T_i - T_{i-1}) \in A_i$ and $(T_i - T_{i-1}) \cup (R_i - R_{i-1}) \in A_i$ can not be satisfied simultaneously. That is, the assumption is not right.In other words, if $\forall A' \in As$, $A' \cap R \neq \varnothing$.

(2) According to algorithm 1, we know that $(R_i - R_{i-1}) \cup (T_i - T_{i-1}) \in A_{i-1}$,that is, $\exists T' \subseteq T_{i-1}$, $(R_i - R_{i-1}) \cup (T_i - T_{i-1}) \cup T' \in As$. Because of $R \cap T = \varnothing$, $(T_i - T_{i-1}) \subseteq T$ and $T' \subseteq T$, so $R \cap ((T_i - T_{i-1}) \cup T') = \varnothing$,in other words, $(R - (R_i - R_{i-1})) \cap ((R_i - R_{i-1}) \cup (T_i - T_{i-1}) \cup T') = \varnothing$.that is, if $\forall a \in R$, $\exists A \in As, A \cap R - \{a\} = \varnothing$.

In a word, the reduction R that algorithm 1 finds out satisfies above two conditions.Therefore, the algorithm is completeness.

5.2 Optimal Algorithm Analysis

Let $As = \{A_1, A_2, ..., A_m\}$.If an algorithm can guarantee that it removes more elements from decernibility set As In every iteration, the times of its iteration is much fewer. That is, the solution it finds out is a smallest or smaller reduction. Therefore, this algorithm is a smallest or smaller algorithm.

According to algorithm 1,in every iteration, it chooses condition attribute a_k appearing the most times in As as reduction attribute, and it can guarantee it will choose condition attribute a_p that appears the most times in As in next iteration. That is, in every iteration, it not only can guarantee that it remove the most elements from decernibility set, but can ensure that it will remove the most elements from decernibility set in next iteration. Therefore, algorithm 1 can ensure that the solution it finds out is a smallest or smaller reduction.

6 Discussion and Conclusion

In [5,7,8],the reduction algorithms are incomplete for Pawlak reduction. Therefore, those algorithms are incomplete for smallest or smaller reduction. In [6],it proposes a reduction algorithm based on the ordered attributes or RA-Order for short, and proves that this algorithm is complete for Pawlak reduction and can find out a unique Pawlak reduction when the order S is fixed. In [9,10], it also proposes a algorithm which is complete for Pawlak reduction.Though above two algorithms can find out a Pawlak reduction, they can not guarantee that the reduction they find out is a smallest or smaller reduction. In [11], Jelonek algorithm can find out a smaller or smallest reduction.However,it is incomplete for Pawlak reduction, that is, it is incomplete for smaller and smallest reduction. In this paper, a complete algorithm for attribute reduction in rough set theory based on the significance of attribute is introduced. This algorithm not only can find out a Pawlak reduction,but also can guarantee that the reduction set it find out is probable a smallest or smaller reduction set. And it has polynomial complexity, which in the worst case is $O(m^2n^4 + m^3n^2))$. At present, one of the most important applications of reduction is data mining.And finding out a smallest reduction is more important in data mining.The algorithm of this paper can solve this problem nicely.

Table 3 The result of experiment in UCI database

Database	number of condition attribute	number of core	result of Jelonek algorithm	result of algorithm1
Ballons Adult+stretch	4	2	2	2
Ballons Small-yellow+ Adult-stretch	4	4	4	4
Lung cancer	56	0	5	4
voting	16	7	9	9
zoo	16	2	5	4
Backup soybean	35	2	10	9
Agaricus-lepiota	22	0	5	4

References

[1] Pawlak Z.Rough sets[J].International Journal of Computer and Information Science,1982,11(5)：341-356.

[2] Slowinski R.Intelligent decision support——handbook of applications and advances of the rough sets thorey[M].London：Kluwer Academic Publishers,1992.

[3] Wong S.K.M,Ziarko W..on optional decision rules in decision tables[J].Bulletin of polish Academy of Sciences,1985,33(11/12)：693-696.

[4] A.Skowron,C.Rauszer. The discernibility matrices and functions in information systems[C].In:R.Slowinski ed. Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory, Dordrecht: Kluwer Academic Publishers, 1992：331-362

[5] WANG Jue,WANG Ren,MIAO Dou-Qian,et al.Data Enriching Based on Rough Set Theory[J].Chinese Journal of Computers,1998,21(5)：393-400.

[6] WANG Jue,WANG Ju, Reduction Algorithm Based on Disernibility Matrix:The Ordered Attributes Method, J.Comput.Sci.&Technol,Nov.2001(16)：489-504.

[7] Hu X H,Cercone N.Learning in relational databases：a Rough Set approach[J]. Computational Intelligence,1995,11(2)：333-338.

[8] MIAO Duo-qian,HU Cui-rong.A houristic algorithm for reduction of knowledge[J].Journal of Computer Research and Development,1999,36(6)：681-684.

[9] WANG Bing,CHEN Shan-ben.A Complete algorithm for Attribute Reduction Based on Disernibility Matrix[J].Journal of ShangHai JiaoTongUniversity,2004,38(1)：43-46.

[10] Jiang Yu, Wang Xie, Ye Zhen. Attribute Reduction Algorithm of Rough Sets Based on Discernibility Matrix,2008,20(14)：3717-3725.

[11] J. Jelonek et al.Rough set reduction of attributes and their domains for neural networks[J].Computational Intelligence, 1995, 11(2): 339-347.

978-1-935068-01-3 © 2009 SciRes.