An Incremental Updating Algorithm of Computation of a Core Based on Array

Xiaoyan Wang, Yuanxia Shen, Xiao Zheng

School of Computer Science & Technology, Anhui University of Technology, Ma'anshan 243032, China

Keywords: rough set, attribute reduction, discernibility matrix, core, array, incremental, updating

Abstract. For the cases of core attribute updating in dynamic decision table, using array elements to count the number of attribute combinations only containing a single attribute in discernibility matrix, this paper introduces an incremental updating algorithm of computation of a core based on array. When objects are added to the decision table, the old core attribute set can be updated effectively by the changes of the values of array elements which represent core attribute .The new core attribute set can be obtained quickly. Finally, the example is given to verify the effectiveness of the algorithm.

Introduction

The rough set theory introduced by Pawlak (1982) has often proved to be an excellent mathematical tool in dealing with vague,ambiguous,imperfect data[1]. Attribute reduction is one of the research focuses in the rough set theory.In keeping classification ability of the information system unchanged,it can induce decision or classification rules through attribute reduction.In attribute reduction algorithms, Obtaining an attribute reduction set usually begins from the computation of core. Therefore, to compute core attribute set is a crucial step. Literatures [2-6] studied calculation method of core attribute. But these studies are only suitable for static decision tables. Literature [7] proposed an incremental updating algorithm for computing core based on the improved discernibility matrix. When the discernibility matrix is updated,it only needs to insert a row and a column into it or delete a row and modify the corresponding column, which can effectively improve the updating efficiency of a core in the algorithm. Literature [8] put forward the incremental updating algorithm for computing core based on the discernibility matrix proposed by Professor Ye Dongyi. Literature [9] presented an incremental updating algorithm based on information entropy ,using binary discernibility matrix. Incremental updating algorithms of computation of a Core proposed by literatures [7-9] all depended on the discernibility matrix. Literature [10] presented an incremental updating algorithms of computation of a Core based on conflicts.The algorithm didn't depend on the discernibility matrix,but needed reprocessing of the information system to get the new decision table and compute the core attribute. It often exists the same attribute combination in discernibility matrix, and for a large decision table, it will have big space complexity to store discernibility matrix. Because the binary code from the subscript of an array element can reflect the comparison for condition attributes of two objects , we can use array elements to represent attribute combination in discernibility matrix. The values of the array elements can represent the frequency occurrence of a attribute combination in discernibility matrix. This paper focuses on the dynamic changes of the core attributes, thus, it only needs to take the changes of array elements representing the core attribute into consideration. After the detailed analysis of the various cases of added objects, this paper puts forward an incremental updating algorithm for computing core in decision table based on array. This algorithm can be used to conduct batch processing to added objects. When updating the values of array elements which represent core attribute, the changes of these values can be used to tell the added or reduced elements in core attribute set. In this way, the original core attribute set gets dynamic updating and consequently the new core attribute set is formed quickly. The algorithm put forward in this paper is proved to be efficient and feasible both by theoretical analysis and illustration of an example.

Related theories

Here are some theories related to the paper in rough set.

Definition 1[11] An information system is expressed as $S=<U,R,V,f>$. Here, U is a collection of objects, also known as the domain and R is the attribute collection, $V=\bigcup_{r\in R} V_r$, V_r expresses the attribute domain and $f:U\times R\rightarrow V$ is an information function which specifies attribute values of every object in U . That is, for $x\in U$, $r\in R$, it exists $f(x,r)\in V_r$. If the set of attributes R can be divided into the condition attribute and decision attribute set, that is, $R=C\bigcup D$, $C\bigcap D=\phi$, $D\neq\phi$, then the information system is called decision system or decision table.

Definition 2 Let $S=(U,C\cup D), R=C\cup D$,where C is the set of condition attributes and D is the set of decision attributes. We call $POS_C(D)=\bigcup_{X\in U/D}\underline{C}X$ as a positive region of the partition U/D with respect to C.

Definition 3 Let $S=(U,R)$, R is an attribute set and $R=C\cup D$,for $\forall c\in C$,if $POS_{R-\{c\}}(D)=POS_R(D)$,c is unnecessary in R, Otherwise, c is necessary in R. The set of all essential attributes about C is called the core of C marking $CORE(C)$.

Definition 4 Let $S=(U,C\cup D)$,if each condition class is contained in the decision class,the decision table is consistent,or else it is inconsistent.

Definition 5[7] For the information system IS,The discernibility matrix is defined as follows.

$$m_{ij}=\begin{cases}\{a\in C:f(x_i,a)\neq f(x_j,a)\} & x_i\in U_1,x_j\in U_1,\text{when } f(x_i,D)\neq f(x_j,D)\\ \{a\in C:f(x_i,a)\neq f(x_j,a)\} & x_i\in U_1,x_j\in U_2^{'}\\ \Phi & \text{Other situations}\end{cases}$$

$$U_1=\bigcup_{i=1}^{k}\underline{C}\varphi_i,\ \ U_2=U-U_1,\ \ U_2^{'}=delrep(U_2)$$

The literature [7] presented an improved discernibility matrix which is applicable to any decision tabel. The space complexity and computational complexity of the improved discernibility matrix are significantly lower than that of HU method. Using the improved discernibility matrix, literature [7] also proposed theorem 1 for judging core attributes.

Theorem 1[7] For the information system IS,if IDM(C, M) $=\{m_{ij}\mid m_{ij}\in M$ and m_{ij} is single attribute$\}$,then $IDM(C,M)=CORE(C)$, in other words,the attribute belongs to $CORE(C)$ iff m_{ij} is a single attribute.

Incremental updating algorithm of core

Using the array element to represent the condition attribute combination; first, each condition attribute is distributed the appeared position in all condition attribute combinations in discernibility matrix. If there is some condition attribute to appear in condition attribute combination, we can mark 1 on corresponding position in attribute combination, otherwise mark 0. We can get a set of binary code and then convert it to a decimal number. Next,we use the decimal number as subscript of array element. The condition attribute combination can be expressed with the array element.

Theorem 2 To a decision table containing n condition attributes, In the discernibility matrix, the binary code corresponding to a condition attribute combination is converted to decimal. The decimal value is $2^m(0\leq m<n)$,which means that the array element representing the condition attribute combination is core attribute.

Proof: In the discernibility matrix, the binary code corresponding to a condition attribute combination is converted to decimal. The decimal value is $2^m(0 \le m < n)$,so there is the only 1 in the binary code . That is, the condition attribute combination corresponding to the binary code only containing single attribute. According theorem 1, we called the attribute as core attribute.

There being n condition attributes in a decision table, We can define a array which contains the number of elements to be n. In the discernibility matrix, the binary code corresponding to a condition attribute combination is converted to decimal. The decimal value is $2^m(0 \le m < n)$.We can use the array element which subscript is m to represent the corresponding core attribution, and use the array element to count the number of appearance of the corresponding core attribute in discernibility matrix.

To a decision table containing n condition attributes,Now we define an array $A[n-1]$,which contains n elements.the lower bound of array subscript is 0 and the high bound is $n-1$. Based on the establishment of discernibility matrix given in Definition 5, the assignment algorithm for the array elements representing core attribute can be described as follows:

```
Input: $U_1 = POS_C(D)$, $U_2 = U - U_1$, $U_2^{'} = delrep(U_2)$
Output: assigned array elements $A[n-1](0 \le i < n)$
BEGIN
 (1)for $(i = 0; i < n; i++)$
        $A[i] = 0$;
 (2)for each $x \in U_1$ do
 (3){ for each $y \in U_1 \cup U_2^{'}$ do
  ｛Compare $n$ condition attributes in sequence, if the corresponding condition attribute is same,we mark 0 ,otherwise mark 1. If the binary code is found to contain two 1, stop comparison and go to (3);or if we get the decimal value corresponding to the binary code being $2^i$,then $A[i] = A[i] + 1$，go to (3);｝
 (4)Return the assigned array elements;
END
```

To a decision table containing n condition attributes, if the value of each array element is 0 based on the assignment algorithm for the array elements representing core attribute,then there's not the core attribute in the decision table.

The influences of adding object x upon core attribute set are as follows:

(1)If the object x repeats with an object in U_1, then the values of the array elements remain unchanged, and so does the core attribute set.

(2) If the object x is repeatitive or inconsistent with an object in U_2, then the values of the array elements remain unchanged, and so does the core attribute set.

(3)If the object x is inconsistent with the object y in U_1 ,then the objects in U_1 are on the decrease,but the objects in U_2 increase. It is necessary to delete the row that the object y belongs to and change the column that the object y belongs to. It is reduced that condition attribute combinations from comparing the object y with the original inconsistent objects in the decision table ,but It is added that condition attribute combinations from comparing the object y with the original consistent objects whose decision attribute is the same as the object y in the decision table . Update core attribute set according to the changes of the values of array elements.

(4)If the new consistent class produces after adding the object x , then the objects in the U_1 increase. It is necessary to add the row and column that the object x belongs to in discernibility matrix.After adding the row and column that the object x belongs to, if the value of new array element being 0 changes to higher than 0, then the new core attribute set is obtained by adding the core attribute corresponding to the array element in the original core attribute set.

Based on the above analysis, the following propositions about core attribute update are tenable.

Corollary 1 If the value of original element being greater than 0 is still higher than 0 and no new array elements are assigned when objects are added to the decision table, then the original core attribute set will not change.

Corollary 2 If the value of some array element being higher than 0 changes to 0 when objects are added to the decision table,then the new core attribute set is obtained by deleting the core attribute corresponding to the array element from the original core attribute set.

Corollary 3 If the value of new array element being 0 changes to higher than 0, the new core attribute set is obtained by adding the core attribute corresponding to the array element in the original core attribute set.

Core updating algorithm based on array is described as follows:

Input: (1) $U_1 = POS_C(D)$, $U_2 = U - U_1$, $U_2^{'} = delrep(U_2)$,the assigned array elements, $CORE(C)$;

(2) The set D including added objects.

Output: The new core attribute set $NCORE(C)$.

BEGIN

for each $x \in D$ do

{//The values of array elements are updated based on the four cases.

(1)If there is the object y in U_1 and for $\forall c \in C$, $c(x) = c(y) \wedge d(x) = d(y)$ then the values of the array elements remain unchanged, and so does the core attribute set.

Else

(2) If there is the object y in $U_2^{'}$ and for $\forall c \in C$, $c(x) = c(y)$ then the values of the array elements remain unchanged, and so does the core attribute set.

Else

(3)If there is the object y in U_1 and for $\forall c \in C$, $c(x) = c(y) \wedge d(x) \neq d(y)$ then the row including the object y is deleted and the column including the object y is altered in discernibility matrix . The corresponding values of array elements need to be updated. $U_1 = U_1 - \{y\}$, $U^{'}_{2} = U^{'}_{2} \bigcup \{x\}$.

(4) If the new consistent class produces after adding the object x, then the objects in the U_1 increase. It is necessary to add the row and column that the object x belongs to in discernibility matrix. $U_1 = U_1 \cup \{x\}$ and $U_2^{'}$ remains unchanged. The corresponding values of array elements need to be updated.

After the values of array elements are updated, the new core attribute set $NCORE(C)$ is obtained based on Corollary 1,2,3.

Return $NCORE(C)$;

END

The instance analysis

Example Table 1 is a decision table, where D is condition attribute set, $C = \{a, b, c, e\}$, and D is decision attribute set, $D = \{d\}$.

Table 1 the decision table

	a	b	c	e	d
x_1	1	0	1	1	2
x_2	1	0	1	1	1
x_3	1	1	1	1	3
x_4	0	1	0	0	1
x_5	0	1	1	1	2

Table 2 the discernibility matrix of the table 1

	x_3	x_4	x_5	x_1
x_3	Φ	ace	a	b
x_4	ace	Φ	ce	$abce$
x_5	a	ce	Φ	ab

The discernibility matrix of the decision table can be established as shown in Table 2.There is four condition attributes in the decision table, so the array $A[3]$ is defined .In the binary code of each attribute combination,let it be comparison value of condition attributes a,b,c,e that come from two different objects from left to right. The array element $A[0]$ counts the number of the condition attribute combinations which only contain e and the array element $A[1]$ counts the number of the condition attribute combinations which only contain c and the array element $A[2]$ counts the number of the condition attribute combinations which only contain b and the array element $A[3]$ counts the number of the condition attribute combinations which only contain a. According to array elements assignment method,assignments for the four array elements are as follows: $A[0]=0$， $A[1]=0$， $A[2]=1$， $A[3]=2$. The core attribute set of the decision table is $CORE(C)=\{a,b\}$.

For table 1, updating cases of the core attribute set are explained by adding different object x.

(1) If the object x is $\{0,1,0,0,1\}$, then x repeats with x_4 in U_1.The values of the array elements remain unchanged, and so does the core attribute set.

(2) If the object x is $\{1,0,1,1,1\}$, then x repeats with x_2 in U_2.The values of the array elements remain unchanged, and so does the core attribute set.

(3) If the object x is $\{1,0,1,1,3\}$, then x is inconsistent with x_1 and x_2 in U_2.The values of the array elements remain unchanged, and so does the core attribute set.

(4) If the object x is $\{1,1,1,1,1\}$, then x is inconsistent with x_3 in U_1. The discernibility matrix is updated as shown in Table 3. The value of $A[2]$ has changed from 1 to 0 and the value of $A[3]$ has changed from 2 to 1.Other array elements remain unchanged.so $NCORE(C)=\{a\}$.

(5) If the object x is $\{1,0,0,1,1\}$, then x is consistent with the objects in $U_1 \cup U_2^{'}$. The discernibility matrix is updated as shown in Table 4. The value of $A[1]$ has changed from 0 to 1 and other array elements remain unchanged.so $NCORE(C)=\{a,b,c\}$.

Table 3 updated discernibility matrix

	x_4	x_5	x_1	x_3
x_4	Φ	ce	$abce$	ace
x_5	ce	Φ	ab	a

Table 4 updated discernibility matrix

	x_3	x_4	x_5	x	x_1
x_3	Φ	ace	a	bc	b
x_4	ace	Φ	ce	Φ	$abce$
x_5	a	ce	Φ	abc	ab
x	bc	Φ	abc	Φ	c

Conclusions

In the paper, it is studied the incremental updating algorithm of core attribute in relevant literatures and using array elements to represent core attribute. It is proposed an incremental updating algorithm of core attribute set based on array. It is not necessary to store the discernibility matrix in the new algorithm. When objects of the decision table are added, the new core attribute set is obtained by the updating of the values of array elements. In this algorithm, space complexity is small and realization is simple. It is suitable for any decision table, and can realize batch processing.

Acknowledgements

This paper is supported by National Natural Science Foundation for the youth(61300059)and Provincial Key Project of Natural Science Research for Anhui Colleges of China (KJ2012Z024, KJ2012Z031) .

References

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

[2] Hu X H, Cercone N. Learning in relational databases: A rough set approach. *Computational Intelligence*, 1995,11(2): 323-337.

[3] Ye Dongyi, Chen Zhaojiong. A new discernibility matrix and the computation of a core. *Acta electronica Sinica*, 2002, 30 (7): 1086-1088

[4] Wang Guoying. Calculation methods for core attributes of decision table. *Chinese Journal of Computers*, 2003, 26 (5): 611-615

[5] Zhao Jun, Wang Guoyin , WU Zhongfu, etal. An efficient approach to compute the feature core. *Mini micro systems*, 2003,24 (11): 1950-1953

[6] Yang Ming, Sun Zhihui. Improvement of Discernibility Matrix and the Computation of a Core. *Journal of Fndan University(Natural Science)*, 2004, 43 (5): 865-868

[7] Yang Ming. An incremental updating algorithm of the computation of a core based on the improved discernibility matrix. *Chinese Journal of Computers*,2006,29(3):407-413

[8] Shen Dongsheng, Huang Zheyu. A An incremental updating algorithm of the computation of a core based on the discernibility matrix. *Journal of Zhangzhou Teachers College (Nat. Sci)* , 2006 (4): 37-40

[9] Qian Wenbin,Yang Bingru,Xu Zhangyan, etal. Efficient incremental updating algorithm for core attribute based on information entropy. *PR & AI*, 2013,26 (1): 42-49

[10] Ge Hao, Li Longshu, Yang Chuanjian. Incremental updating algorithm of the computation of core based on the collision . *Control and Decision*,2011,26 (7): 984-990

[11]Wang Guoyin .Rough theory and knowledge acquisition[M].Xi’an:Xi'an Jiaotong University Press,2001:118-119