An expanding application of heuristic reduction algorithm

Ning Xu, Jun Gao, Gang Shen
School of Computer Science and Information Engineering
Shanghai Institute of Technology
Shanghai, China

Yun Zhang
Faculty of Automation
Guangdong University of Technology
Guangzhou, China

Abstract—**Heuristic reduction algorithm has got an extensive application in decision table; based on rough sets theory and partition of set, the attribute reduction of information systems without decision attributes is discussed. And a heuristic rule and an algorithm are given out. Some example shows the algorithm is effective and is an instructive attempt.**

Keywords-soft computing; data mining; attribute reduction; rough sets; decision table

I. INTRODUCTION

The research of attribute reduction algorithms is one of hot spots in algorithm researches and it attracts many attentions. As a new mathematic tool, rough sets theory, proposed in last century by Poland mathematician Zdzistaw Pawlak [1]~[2], is famous for dealing with imprecise and uncertain data; also the attribute reduction of a dataset is one of its main research areas [3]~[5]. The research of rough sets thinks if the class of a dataset is unchanged, then the redundant attributes can be reduced. Based on the theory, many attribute reduction algorithms [5]~[10] were put forwarded and the research promote the development of algorithm and rough sets applications greatly.

Decision table is a main research object, its attribute reduction has been researched extensively [4~6], especially in attribute significance heuristic reduction algorithms [7]. Based on these researches, the reduction of decision tables without decision attribute was come up, because it also is a kind of datasets and is widely used, such as dimensionality reducing of sparse matrix or data pre-processing, or classifying, etc. The datasets have redundant attributes when meet some classifying and the datasets need simplifying. How to find the redundant attributes and reduce the attributes? From rough sets reduction theory and set theory, the paper discusses the problem and gives out a heuristic rule and reduction algorithm. An example shows the algorithm's validity.

II. ROUGH SETS AND SET THEORY

Rough sets calls a dataset an information system S, $S=\{U, A, V, f\}$, U is a nonempty limited collection of data objects, $|U|=n$, A is a nonempty limited collection of the object attributes, V is the values collection of the attributes, f is a map: $U\times A\rightarrow V$. Hear $A=C\cup D$, and C is condition attribute set and D is decision attribute set., The information system is called decision table as $D\neq\emptyset$ [3].

When $D=\emptyset$, the information systems are called decision table without decision attribute.

Relevant rough sets reduction theory and set theory are depicted together bellow [3].

Definition 1. Information system $S=\{U, A, V, f\}$, $A=C\cup D$, $C\cap D=\emptyset$. If $c\in C$, U/c is the collection of all equivalent classes of c on U, called quotient set: $U/c=\{[x]_c, x\in U\}$.

Definition 2. $R\subseteq C$, $R\neq\emptyset$, then $\cap R$ is a equivalence relation, and is called indiscernibility relation on R, noted as ind(R); ind(R)$=\{(x, y)\in U\times U \mid f(x, c)=f(y, c), c\in R\}$, in quotient set: $U/\text{ind}(R)=\{[x]_{\text{ind}(R)} \mid x\in U\}$.

Definition 3. $c\in R$, if ind(R-$\{c\}$)=ind(R), then c is unnecessary in R, or can be reduced, otherwise c is necessary in R. If every $c\in R$ is necessary, then R is called independent.

Definition 4. If $R\subseteq C$, R is independent, and ind(R)=ind(C), call R is a reduct of C, write as red(C).

Definition 5. The intersection of all reducts of C is called core of C, and noted as core(C), core(C) = $\cap$red(C).

According to set theory, an equivalence relation R on a set U also is a partition π over U.

Definition 6. Suppose two partitions of set U are $\pi_1=\{A_1, A_2, \ldots, A_n\}$, $\pi_2=\{B_1, B_2, \ldots, B_m\}$. If $\forall A_i\subseteq B_j$, j=1 or 2 or … or m, then call π_1 is a subdivision of π_2, and described as $\pi_1\preccurlyeq\pi_2$. If $\exists A_i\subset B_j$, j=1 or 2 or … or m, then call π_1 is a proper subdivision of π_2, and described as: $\pi_1\prec\pi_2$.

Definition 7. Suppose π_1 and π_2 are two partitions of set U, then $\{A_i\cap B_j \mid A_i\cap B_j\neq\emptyset, i=1, 2, \ldots, n, j=1, 2, \ldots, m\}$ is called as *product partition* of π_1 and π_2, described as: $\pi_1\bullet\pi_2$.

If R_1 and R_2 are two equivalence relations of set U, π_1 and π_2 are two partitions corresponding to R_1 and R_2, then product partition $\pi_1\bullet\pi_2$ is the partition of equivalence relation $R_1\cap R_2$.

If partition π_1 is corresponding to condition attribute $c_1\in R$, $R\subseteq C$, π_2 is corresponding to condition attribute $c_2\in R$, …, π_i is corresponding to condition attribute $c_i\in R$, …, then ind(R) will be corresponding to $\prod\pi=\pi_1\bullet\pi_2\bullet\pi_3\bullet\ldots$. The reduction problem can be analyzed by partition.

The research is supported by National Natural Science Foundation of China (NSFC, No.U0735003), Science Foundation of Shanghai Municipal Education Commission (No. 06 0Z 021), Shanghai Institute of Technology Foundation (No.YJ2008-07, A22/1020M090059)

978-1-61284-459-6/11/$26.00 ©2011 IEEE

If $R_1 \subseteq C$, $R_2 \subseteq R_1$, π_1 and π_2 are two partitions corresponding to R_1 and R_2, it can be proved that $\pi_1 \preccurlyeq \pi_2$, which means R_1 is finer relation than R_2, or R_2 is a coarser relation then R_1.

III. HEURISTIC RULE

The decision table without decision attribute, $D=\varnothing$, $S=\{U, C, V, f\}$, its indiscernibility relation on C is $\mathrm{ind}(C)=\cap C$. $\cap C$ means a quotient set $U/\mathrm{ind}(C)$, which is corresponding to the finest equivalent relation over U or the great dividing by C (because the same objects can be reduced from S, and the attribute reduction will not be affected). That is say $|\mathrm{ind}(C)| = |U| = n$, one object x ($x\in U$) is an equivalence class.

From rough sets reduction theory, the problem of attribute reduction on decision table without decision attribute is: find $R\subseteq C$, make:

$$J=\min|R| \tag{1a}$$

$$U/\mathrm{ind}(R)=U/\mathrm{ind}(C) \tag{1b}$$

Or: $|\mathrm{ind}(R)| = n$

In the heuristic reduction algorithm research, the attribute significance heuristic reduction algorithm is studied and applied most widely. For the algorithm easy to implement, the key of the algorithm is set up its heuristic rule.

Suppose R is the attribute subset selected, $\mathrm{core}(C)\subseteq R$, and $\mathrm{ind}(R)\neq\mathrm{ind}(C)$. The next step is selecting $c\in C-R$, it should subdivide R with most possible. If R and c have partitions on U: π_R and π_c they are:

$$\pi_R=U/\mathrm{ind}(R)=\{X_1, X_2, \ldots, X_r\}$$

$$\pi_c= U/c=\{Y_1, Y_2, \ldots, Y_t\}.$$

From Definition 7, their product partition π (includes void):

$$\pi=\pi_R \bullet \pi_c=U/\mathrm{ind}(R\cap c)=\begin{Bmatrix} E_{11} & E_{12} \cdots E_{1t} \\ E_{21} & E_{22} \cdots E_{2t} \\ \vdots & \vdots \quad \vdots \\ E_{r1} & E_{r2} \cdots E_{rt} \end{Bmatrix} \tag{2}$$

Among (2): $E_{ij}=X_i\cap Y_j$, $i=1, 2, \ldots, r$, $j=1, 2, \ldots, t$,

and it meets: $X_i=\bigcup_{j=1}^{t} E_{ij}$, $\quad Y_j=\bigcup_{i=1}^{r} E_{ij}$.

Increasing equivalence relation to R, the relation of R will be finer further. The more of elements (not void) in formula (2) is, the finer of relation R will be. So the condition attribute c, which will be selected from C-R next, should have the most elements (not void) in E matrix.

About information system $S=\{U, C, V, f\}$, first finds $\mathrm{core}(C)$, then $R\leftarrow\mathrm{core}(C)$. Based on R, selects c ($c\in C-R$), and computes $U/\mathrm{ind}(R)=\{X_1,X_2,\ldots, X_r\}$ and $U/c=\{Y_1,Y_2,\ldots, Y_t\}$. By formula (2) gets E matrix, then simplifies elements in E as bellow equation to get an $E1$ matrix firstly. Then counts not empty elements in $E1$according to the next equations and get $E3$:

$$E_{ij}=X_i\cap Y_j, \implies E1_{ij}=\begin{cases} 0, & E_{ij}=\varnothing \\ 1, & E_{ij}\neq\varnothing \end{cases}.$$

$$\implies E2_i=\sum_{j=1}^{t} E1_{ij},$$

$$\implies E3_i=\begin{cases} 0, & E2_i=1 \\ E2_i, & E2_i>1 \end{cases}.$$

$E3_i$ shows the mount of subdividing R by c. The mount is the more the better, it is a classified feature differences between c and R, and it can be called difference degree between c and R by partition on U. So an attribute significance heuristic rule can be given out.

Definition 8. The attribute significance of $c\in C-R$ is called maximal difference degree with R as:

$$\mathrm{sig}\,(c)=\sum_{i=1}^{r} E3_i \tag{3}$$

Or using function expression, the significance is:

$$\mathrm{sig}\,(c)=\sum_{i=1}^{r} g\Big(\sum_{j=1}^{t} f(E_{ij})\Big) \tag{3a}$$

Here: $$\boldsymbol{f}(E_{ij})=\begin{cases} 0, & E_{ij}=\varnothing \\ 1, & E_{ij}\neq\varnothing \end{cases}, \tag{3b}$$

$$\boldsymbol{g(*)}=\begin{cases} 0, & \sum f(E_{ij})=1 \\ \sum f(E_{ij}) & \end{cases}. \tag{3c}$$

IV. ALGORITHM AND EXAMPLE

After the definitions above, a reduction algorithm HRA-mdd (Heuristic Reduction Algorithm of Significance of Attribute, Max-difference-degree) about decision table without decision attribute could be expressed:

input $S=\{U, C, V, f\}$, $C=\{c_1, c_2,\ldots,c_m\}$
output $\mathrm{red}(C)$.

Step 1 computing $\mathrm{ind}(C)$, $\mathrm{core}(C)$ and $R=\mathrm{core}(C)$

Step 2 $C'=C-R$ IF $C'=\varnothing$ THEN $\mathrm{red}(C)=R$, STOP. ELSE

Step 3 IF $R=\varnothing$ THEN $R=\{c_p \mid \mathrm{rank}(c_p)=\max(\mathrm{rank}(c_i), c_i\in C')\}$, computing $\mathrm{ind}(R)$

Step 4, $U=U-x(x\in[x]_R=\{x\})$

Step 5 computing E for every c_i, and $\mathrm{sig}(c_i)$, $c_i\in C'$;

Step 6 $c_{max}=\{c|\mathrm{sig}(c)=\max(\mathrm{sig}(c_i), c_i\in C')\}$, $R=R\cup c_{max}$, computing $\mathrm{ind}(R)$

Step 7 IF $\mathrm{ind}(R)=\mathrm{ind}(C)$ THEN $\mathrm{red}(C)=R$ STOP. ELSE

Sept 8 $C'=C'-c_{max}$, RETURN Step 4

The x ($x \in [x]_R=\{x\}$) is reduced from U for it needn't divided. This algorithm can get minimum reduction of information system without decision attributes quickly.

Now, below is a dataset CTR (Car Test Results) without decision attribute. According to the algorithm HRA-mdd the computing results will be:

TABLE I. CTR DATASET WITHOUT DECISION ATTRIBUTES

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
1	0	1	1	1	1	1	1	1	0
2	0	1	0	1	1	0	1	0	0
3	0	1	0	1	1	1	1	0	0
4	0	0	1	1	1	1	1	0	1
5	0	1	0	1	1	0	0	0	0
6	0	1	0	0	1	0	0	1	2
7	0	1	0	1	1	0	1	0	2
8	1	0	0	0	0	1	2	0	1
9	0	0	0	0	0	1	2	0	0
10	0	0	0	0	0	1	0	1	0
11	1	0	0	1	0	1	2	0	1
12	1	0	0	1	1	0	0	0	0
13	0	0	0	0	1	0	0	0	0
14	1	0	1	1	0	1	1	0	0
15	1	0	0	0	0	0	2	0	0
16	0	0	1	1	1	0	1	0	0
17	0	1	0	1	1	0	1	1	0
18	0	0	0	1	1	0	1	1	0
19	1	0	0	1	0	1	0	0	0
20	0	0	0	1	0	1	0	0	0

$U/\mathrm{ind}(C)=\{\{1\},\{2\},\{3\},\{4\},\ldots, \{19\},\{20\}\}$,

$\mathrm{core}(C)=\{c_1, c_2, c_4\}$,

In the first round:

$U/\mathrm{ind}(R)= \{\{1,3,5,7,17\},\{4,16,18,20\},\{8,15\},\{9,10,13\},\{11,12,14,19\}\}=\{X_1, X_2, X_3, X_4, X_5\}$;

$C'= \{c_3, c_5, c_6, c_7, c_8, c_9\}$, their $\mathrm{sig}(c_i)$ is computed, the result and the order is:

c_3 c_5 c_6 c_7 c_8 c_9

5 6 1 2 4 3

Selecting c_6 and $R=R\cup c_6$, but $\mathrm{ind}(R)\neq\mathrm{ind}(C)$. In the second round, the computing result is:

c_3 c_5 c_7 c_8 c_9

2 5 1 3 4

After selecting c_7, $R=R\cup c_7$, but $\mathrm{ind}(R)\neq\mathrm{ind}(C)$. In the third round, the sort order is:

c_3 c_5 c_8 c_9

1 3 2 3

$R=R\cup c_3$, $R=\{c_1, c_2, c_4, c_3, c_6, c_7\}$ and $\mathrm{ind}(R)=\mathrm{ind}(C)$, so $\mathrm{red}(C)=R$. There are three attributes unnecessary in CTR dataset classification. In the last round, $\mathrm{sig}(c_3)=\mathrm{sig}(c_8)$, if $R=\{c_1, c_2, c_4, c_6, c_7, c_8\}$, $\mathrm{ind}(R)=\mathrm{ind}(C)$, so this is an other reduction result got by the algorithm. The example shows the algorithm is effective.

Comparing with attribute reduction of decision table, with decision attributes, the reduction of decision table without decision attributes needs more attributes to divide partition because the classification is the great dividing.

The time complexity of the algorithm is from two parts. One is the system indiscernibility relation $\mathrm{ind}(C)$, and the second is the attribute significance $\mathrm{sig}(c)$. Because both parts include the intersection computing of equivalence relation, the time complexity is $O(|U|^2)$. The complexity of the algorithm would be $O(|C||U|^2)$ in the worst when computes all condition attributes.

V. CONCLUSION

Rough sets and heuristic reduction algorithm are the foundation for attribute reduction. The discussion, from equivalence relation to set partition, gives out a new attribute significance heuristic rule and detail reduction algorithm for information systems without decision attributes. An example shows the effectiveness and efficiency of the algorithm. This only is an instructive attempt, will to solve the problem of attribute reduction of information systems without decision attributes. Rough sets theory get a new using and will be widespread further.

REFERENCES

[1] Z. Pawlak, Rough Sets, International Journal of Computer and Information Sciences, Vol.11, No.5, 1982, pp.341-356

[2] Z. Pawlak, Rough sets and their applications, Microcomputer Applications, Vol.13, No.2, 1994. pp.71-75.

[3] Wenxiu Zhang, Weizhi Wu, Jiye Liang, Theory and method of rough sets, First Press, Beijing, Sciences Press, 2001.7.

[4] Hu, Xiaohua, Knowledge discovery in database: An Attribute-oriented rough set approach (Rules, Decision Matrices), PhD dissertation, The University of Regina (Canada), 1995

[5] Guoyin Wang, Yiyu Yao, A survey on Rough set theory and applications, Chinese Journal of Computers, Vol.7 2009, pp.1229-1246

[6] Feng Hu, Guoyin Wang, Quick reduction algorithm Based on Attribute Order, Chinese Journal of Computers, Vol.30, No.8, 2007, pp.1429-1435

[7] Chunyan Deng, Yuejin Lv, Jinhai Li, Efficient attribute reduction algorithm on decision table, Computer Engineering and Applications, 2009,45(4): pp.152-155

[8] Weihua Xu, Xiaoyan Zhang, Jianmin Zhong, Wenxiu Zhang, Heuristic algorithm for attribute reduction in ordered information systems, Computer Engineering, Vol.36, No.17, 2010, pp.69-71

[9] Zengtai Gong, Yongping Guo, Zhanhong Shi, Attribute reduction of generalized information systems, Computer Engineering and Applications. Vol.46, No.23, 2010, pp.34-37

[10] Dongyi Ye, Zhaojiong Chen, Shenglan Ma, A New Fitness Function for Solving Minimum Attribute Reduction Problem, Lecture Notes in Computer Science, 2010, Volume 6401, Rough Set and Knowledge Technology, pp.118-125

[11] Ning Xu, The theory and technique research of attribute reduction in data mining based on rough sets, PhD dissertation, Guangdong University of Technology, 2005

[12] Ning Xu, Yun Zhang, Xiao Wei, An attribute reduction algorithm based on classified knowledge, Proccedings of 2010 IRAST International Congress on Computer Applications and Computational Science (CACS 2010), pp.999-1002