Reduction in Decision Table Based on Pair-Wise Complementarity of Condition Attributes

Xueen Wang
Institute of Integrated Automation
Xi'an Jiaotong university
Xi'an, Shaanxi China.
wangxueen@gmail.com

Chongzhao Han
Institute of Integrated Automation
Xi'an Jiaotong university
Xi'an, Shaanxi China.
czhan@mail.xjtu.edu.cn

Deqiang Han
Institute of Integrated Automation
Xi'an Jiaotong university
Xi'an, Shaanxi China.
deqhan@gmail.com

Abstract – The reduction of attributes is a critical problem in the rough set theory. Finding the minimal reduct is turned out to be a NP-hard problem. Many heuristic algorithms, which use the significance of the condition attribute with reference to the decision attributes as the indication for attribute selection, have been proposed in this area. In this paper the pair-wise complementarity of condition attributes is defined based on conditional information entropy and employed as a heuristic in the attribute reduction process. Finally, a heuristic algorithm of reduction is proposed and tested on the UCI machine learning repository. It can be verified by the experimental results that the proposed algorithm is feasible and effective.

Keywords: relative reduction, conditional information entropy, rough set, decision table, condition attribute.

1 Introduction

Rough set theory was proposed by Professor Zdzislaw Pawlak in the 1982 [1]. It is a powerful mathematic tool for data analysis, dependency analysis and rule mining in information system with vague and uncertain knowledge. With the rapid growth of rough set theory, it seems that the rough set is fundamentally important in artificial intelligence and cognitive science, especially in research areas such as machine learning, intelligent system, inductive reasoning, pattern recognition, mereology, knowledge discovery, decision analysis and expert system, et al. [2]

The rough set theory is based on the indiscernsibility relation, which expresses the ability to discern some objects by using available information. To find the hidden knowledge for classification analysis, which is often expressed in form of decision rules, in an information system is a core problem in rough set theory. According to Occam Principle, the most ideal decision rules should be the simplest ones [3]. Generally there is a lot of redundant information in a information system. To get the simplest rules, the minimal subset of attributes keeps the same indiscernsibility relation with the set of all attributes should be found at first, which is the task of attributes reduction. So, reduction plays a fundamental role in analysis of information system using rough set theory. Unfortunately, searching for the optimal reduct is proved to be a NP-hard problem [4]. Therefore, many heuristic algorithms have been proposed. One important type of reduction algorithm is based on the discernibility matrix, which is introduced in [5]. In [6] the frequency of attributes absent is acquired from the discernibility matrix, and it is used as a heuristic in reduction. The second important type is based on the Pawlak's measure of attribute significance from the viewpoint of algebra. The third important type of reduction algorithm is based on the theory of information entropy. In [7], a comparative study on the quantitative relationship between some basic concepts of rough set theory from the viewpoints of information and algebra is given. It has been proved that the two viewpoints are equivalent in a consistent information system. The attribute significance denoted by information entropy is used as heuristic information in reduction algorithm [8][9][10]. Otherwise, many methods based on stochastic search algorithm, which can find the optimal solution, have been proposed, such as GA-based reduction in [11] and PSO-based method in [12]. In the recent years, the reduction in the inconsistent and incompletely information system is a hot research topic in this area [13][14].

In most of the heuristic reduction algorithms in decision table, the significance of the condition attribute with reference to the decision attributes is measured by the classification ability. The mutual information of each condition attributes pair is not taken into consideration. From the viewpoint of the information fusion, the fusion performance will be better if more complementarity (difference) information between two attributes (features) is obtained. Therefore, the pair-wise complementarity of condition attributes is used for selection of attribute in reduction process in this paper.

The rest of the paper is organized as follows. The basic notions of rough set theory are briefly presented in Section 2. A heuristic reduction algorithm based on pair-wise complementarity of condition attributes is proposed in section 3. In section 4, experimental results on UCI repository is given. The results demonstrate the feasibility and effectiveness of the proposed algorithm.

2 Basic notions related to rough set theory

The basic definitions and notions related to rough set theory are shown in this section. Information system is defined by Pawlak for data representation [15]. A decision table is an information system with a set of decision attributes.
A decision table can be given as:

$$S=(U,C,D,V,f) \tag{1}$$

Where U is the universe of discourse, a non-empty finite set of objects; C is a finite set of condition attributes; D is a finite set of decision attributes; $V=\bigcup_{a\in C\bigcup D}V_a$, where V_a is a set of the domain for an attribute $a\in C\bigcup D$; $f:U\times(C\bigcup D)\rightarrow V$ is the total decision function such that $f(x,a)=v_a\in V_a$ for $x\in U$, $a\in C\bigcup D$. A decision table is consistent if all object pairs that have the same condition values on C also have the same decision value on D , otherwise, it is inconsistent if there exists at least one object pair that have the same condition values on C but different decision values on D [13]. For an information system $S=(U,C,V,f)$, if V contains null value for at least one attribute, then S is an incomplete information system, otherwise it is complete [16].
For any subset $P\subseteq C\bigcup D$, an equivalence relation (also called indiscernibility relation) can be defined as:

$$IND(P)=\{(x,y)\in U^2\big|f(x,a)=f(y,a),\forall a\in P\} \tag{2}$$

A partition of U can be got based on $IND(P)$ denoted by $U/IND(P)$, or simply U/P . Every block of the partition is an equivalence class which can be denoted as $[x]_P=\{y\in U|(x,y)\in IND(P)\}$.
For any every $X\subseteq U$, the lower approximation and upper approximation of X with respect to B can be get by:

$$\begin{aligned}P_*(X)&=\left\{x\in U\mid[x]_p\subseteq X\right\}\\P^*(X)&=\left\{x\in U\mid[x]_p\cap X\neq\varnothing\right\}\end{aligned} \tag{3}$$

The object in the lower approximation can be certain classified as X using P . The object in the upper approximation can be possibly classified as P using B. The pair of $\left(P_*(X),P^*(X)\right)$ is referenced as the Pawlak's rough set of X with respect to P .
A P -Positive region of D is a set of all objects from the universe U that can be classified with certain to one class of U/P employing attributes from P .

$$POS_P(D)=\bigcup\nolimits_{X\in U/D}P_*(X) \tag{4}$$

A dependency of the D on P is defined as:

$$\gamma_P(D)=\frac{|POS_P(D)|}{|U|} \tag{5}$$

Where $|A|$ is the cardinality of set A . An attribute $a\in P$ is indispensable in P , if $\gamma_P(D)\neq\gamma_{P-\{a\}}(D)$; otherwise a is dispensable in P with respect to D .
Based on those definitions a reduct set of condition attributes C with respect to decision attributes D can be defined as follow, a set of attributes $R\subseteq C$ is called the relative reduct of C , if $POS_R(D)=POS_C(D)$ and all element of R is indispensible. In one word, a relative reduct can be considered as a minimum subset of condition attributes that preserves classification ability of the set of all condition attributes [13]. The intersection of all reducts is called the core, the elements of which are those attributes that cannot be eliminated.

3 Reduction algorithm based on the pair-wise complementarity of condition attributes

The important problem in heuristic algorithms is how to measure the attribute significance, which is hard to find the optimal solution. In this paper, the significance is denoted by the pair-wise complementarity of attributes, which is defined based on the conditional information entropy. Firstly, some definitions of information entropy are presented as follows.
Given a decision table $S=(U,C,D,V,f)$, for $X\subseteq C\bigcup D$ and $Y\subseteq C\bigcup D$, the probabilistic space of $U/IND(X)$ and $U/IND(Y)$ in the σ algebra are:

$$[U/X:P]=\begin{bmatrix}X_1 & X_2 & \cdots & X_n\\ p(X_1) & p(X_2) & \cdots & p(X_n)\end{bmatrix} \tag{6}$$

$$[U/Y:P]=\begin{bmatrix}Y_1 & Y_2 & \cdots & Y_m\\ p(Y_1) & p(Y_2) & \cdots & p(Y_m)\end{bmatrix} \tag{7}$$

Where $p(X_i)=\frac{|X_i|}{|U|}$, for $i=1,\cdots,n$; $p(Y_j)=\frac{|Y_j|}{|U|}$, for $j=1,\cdots,m$.
The information entropy of X can be defined as:

$$H(X)=\sum_{i=1}^{n}p(X_i)\log(p(X_i)) \tag{8}$$

The conditional information entropy of X with reference to Y can be defined as:

$$H(Y/X) = \sum_{i=1}^{n} p(X_i) \sum_{j=1}^{m} p(Y_j/X_i) \log\left(p(Y_j/X_i)\right) \tag{9}$$

The conditional information entropy has already been used to measure the attribute significance in [8] [9]. In [8] a bottom-up algorithm MIBARK was proposed, which is based on the mutual information between condition attributes and decision attributes. The mutual information can be calculated as $I(C,D) = H(D) - H(D/C)$. The change between $I(B,D)$ and $I(\{a\} \cup B, D)$ for $a \in C$ denotes the significance of the attribute a with reference to D . In [9], the conditional information entropy was directly used to measure the significance of the attribute $a \in C$ with reference to D by $H(D/\{a\})$. Two algorithms, CEBARKCC and CEBARKNC are proposed in paper [9]. Essentially, there is no difference between the first algorithm CEBARKCC and the algorithm MIBARK. The second one is an up-bottom algorithm, which delete the attribute with the minimal significance one by one if the conditional information entropy of residual attributes with reference to D is equal to $H(D/C)$.

The conditional information entropy $H(\{b\}/\{a\})$ denotes the importance of a to b . The smaller the entropy is, a is more important to b , in other words, there are more common elements between $U/\{a\}$ and $U/\{b\}$. Therefore the entropy is small, the difference between a and b is small, b will get few information from a . In other words, a have less complementarity to b . Then the conditional information entropy is employed to denote the complementarity of each pair of attributes. Let $a \in C$ and $b \in C$, the complementarity of attribute a to b can de defined as:

$$\kappa(a,b) = H(\{a\}/\{b\})(1 - H(D/\{a\})) \tag{10}$$

Where $H(D/\{a\})$ is the conditional information entropy of attribute a with reference to D , which denotes the classification ability of attribute a . The complementarity of a to b is proportional to the classification ability of a and the classification ability of a is inversely proportional to $H(D/\{a\})$. The complementarity of a to b is more as the value of $\kappa(a,b)$ is bigger. Obviously the complementarity between a and b is asymmetry, i.e. $\kappa(a,b) \neq \kappa(b,a)$. So a complementarity matrix of each pair of condition attributes is needed.

Based on Eq. 10, a complementarity of an attribute to a set of attributes can be defined and it will be used to measure the attribute significance in the proposed reduction algorithm. The attribute significance is proportional to the complementarity. Let $a \in C$ and $B \subseteq C$, the complementarity of attribute a to B can de defined as:

$$\kappa'(a,B) = (1 - H(D/\{a\})) \sum_{b \in B} \kappa(a,b) \tag{11}$$

Based on the definitions above, a bottom-up algorithm was proposed. Let B is the reduct and C_O is the core of the decision table. The algorithm starts with the core, i.e. $B = C_O$, which is the same with most heuristic bottom-up algorithms. Then the attribute with the maximum complementarity to the attribute set B will be added into the set one by one until $H(D/B) = H(D/C)$.

The flow of the proposed algorithm is shown as follow:

Input: a decision table $S = (U, C, D, V, f)$

Output: a relative reduct of the decision table

Step 1: computer the conditional information entropy of each attribute with reference to the set of decision attributes $H(D/\{a_i\})$ and $H(D/C)$.

Step 2: computer the core C_O and the candidate attribute set $A = C - C_O$ for selection.

Step 3: computer the complementarity matrix of each pair of attributes

Step 4: if $|C_O| \neq 0$ { let $B = C_O$ }.

else {let B equals empty set, add the attribute with the minimal conditional information entropy $H(D/\{a_i\})$ into B }.

if $|B| \neq 0$, jump to step 6.

Step 5: select an attribute with the minimal complementarity $\kappa'(a,B)$ for $a \in A$. Delete a from A .

Step 5.1: calculate the conditional information entropy $H(D/B \cup \{a\})$

Step 5.2: if $H(D/B \cup \{a\}) \neq H(D/B)$ {add a into B }, else { jump back to step5}

Step 6: calculate the conditional information entropy $H(D/B)$.

if $H(D/B) \neq H(D/C)$ { jump to step 5 }

else {the program is **End** and B is the relative reduct}

Based on the flow of algorithm, the complexity of the algorithm can be analyzed. The time complexity is composed of three parts in the proposed algorithm, time for calculating core attributes, time for non-core attributes and time for complementarity matrix, which are $t_c = O(|C||U|^2)$, $t_{non-core} = O(|C||U|^2)$ and $t_{c\_m} = O(|C||C-1||U|^2)$. The algorithm CEBARKCC and MIBARK, which are both based on information entropy, have the same time complexity [9].

Table 1. A decision table

U	C				D
	Outlook(a_1)	Temperature(a_2)	Humidity(a_3)	Windy(a_4)	
1	Sunny	Hot	High	False	N
2	Sunny	Hot	High	True	N
3	Overcast	Hot	High	False	P
4	Rain	Mild	High	False	P
5	Rain	Cool	Normal	False	P
6	Rain	Cool	Normal	True	N
7	Overcast	Cool	Normal	True	P
8	Sunny	Mild	High	False	N
9	Sunny	Cool	Normal	False	P
10	Rain	Mild	Normal	False	P
11	Sunny	Mild	Normal	True	P
12	Overcast	Mild	High	True	P
13	Overcast	Hot	Normal	False	P
14	Rain	Mild	High	True	N

The time complexity of CEBARKCC is composed of $t_c = O(|C||U|^2)$ and $t_{non-core} = O(|U|^3)$. The proposed method is not good at time performance because of calculation of the complementarity matrix. The same problem exits in the space complexity, a $O(|C||C-1|)$ memory should be needed for the complementarity matrix.
A decision table is shown in table 1. Take this table for example, the conditional information entropy of C with reference to D and the core are calculated firstly.
$H(D/C) = 0$; $C_O = \{a_1, a_4\}$;
$H(D/a_1) = 0.2088$;
$H(D/a_4) = 0.2743$;
$H(D/a_3) = 0.2373$;
$H(D/a_2) = 0.2686$.
The complementarity matrix is shown in table 2.

Table 2. Complementarity matrix of condition attributes pairs

$\kappa(row, col)$	a_1	a_2	a_3	a_4
a_1	0	0.4032	0.4686	0.4730
a_2	0.3970	0	0.3558	0.4568
a_3	0.2947	0.1882	0	0.3010
a_4	0.2948	0.2848	0.2966	0

Let $B = C_O$, calculate the Complementarity between the residual attributes and B .
$\kappa'(a_2, B) = (\kappa(a_2, \{a_1\}) + \kappa(a_2, \{a_4\}))(1 - H(D/a_2))$
=0.7245.
$\kappa'(a_3, B) = (\kappa(a_3, \{a_1\}) + \kappa(a_3, \{a_4\}))(1 - H(D/a_3))$
=0.6197
As the attribute a_2 has more complementarity to B , it is added into B . Now, $H(D/\{a_1, a_2, a_4\}) = H(D/C)$. Then the reduct is $\{a_1, a_2, a_4\}$.

4 Experimental Results

The reduction results of the proposed algorithm are presented in table 3. Nine data sets from UCI machine learning repository are chosen in the experiment [17]. The reduction results of CEBARKCC are also shown in table 3 for comparison. A discretization pre-process of continuous attributes has been done utilizing the software ROSETTA [18]. The inconsistent and incomplete instances have been deleted.
It can be concluded that there is no difference between the proposed algorithm and CEBARKCC, with the exception of “SPECT” and “Australian Credit”. Although they have different reducts, the cardinalities of the reducts are equal.
The time consumption of algorithms is not shown in the table. Some time consumption was recorded for an analysis. The algorithms are implemented in VC++ 6.0 and tested in personal computer with single 2.4GHZ processor and 384M memory. For “SPECT”, the time consumption of CEBARKCC is 803ms, while that of the proposed algorithm is 2967ms. There are 22 condition attributes in “SPECT”. The time consumption for calculating the complementarity matrix is 2215ms, which constitutes the most part of the time consumption of the proposed algorithm.

Table 3. The results of reduction experiment

Data Sets	Number of Attributes	Number of Instances	CEBARKCC	Algorithm in this paper
Australian Credit	14	345	{1, 3, 4, 5, 7, 8, 10, 13, 0}	{1, 3, 4, 5, 7, 8, 10, 13, 12}
Iris	4	75	{2, 3}	{2, 3}
SPECT	22	80	{3, 17, 19, 18, 7, 20, 21, 9, 0, 5, 2, 12}	{3, 17, 19, 0, 13, 7, 21, 12, 1, 4, 20, 8}
Glass	10	107	{0}	{0}
Pima Indian	9	384	{6, 4, 5, 7, 3}	{6, 4, 5, 7, 3}
HSV	11	122	{3, 10, 5, 8, 6, 9, 2, 4, 1}	{3, 10, 5, 8, 6, 9, 2, 4, 1}
Heart Disease	30	569	{14, 11}	{14, 11}
BUPA	6	346	{2, 1, 5, 4, 3, 0}	{2, 1, 5, 4, 3, 0}
MONK(1)	6	124	{1, 5, 4, 2, 3}	{1, 5, 4, 2, 3}

For the data set "Iris", which has 4 condition attributes, both the time consumption for two algorithms is 15ms. The results in this section are in accordance with the analysis in section 3.

5 Conclusion

The reduction of attributes is a critical problem in the knowledge discovery and feature selection. A heuristic reduction algorithm based on the pair-wise complementarity of condition attributes is proposed and it is tested on the UCI repository. Its feasibility and effectiveness have been demonstrated by the experimental results. A fast calculation method for the complementarity matrix will be developed in the future work. Furthermore, the inconsistent and incomplete information system from the real world will be also considered.

Acknowledgements

This work is supported in part by the Grant for State Key Program for Basic Research of China (973) No.2007CB311006 and the Research Fund of Shaanxi Key Laboratory of Electronic Information System Integration (No.200910A).

References

[1] Z. Pawlak, "Rough sets," International Journal of Parallel Programming, Vol. 11, No. 5, pp. 341-356, Oct. 1982.

[2] Z. Pawlak, and A. Skowron, "Rudiment in Rough sets," Information Science, Vol. 177, No. 1, pp. 3-27, Jan. 2007.

[3] J. S. Mi, et al., "Approaches to knowledge reduction based on variable precision rough set model," Information Science, Vol. 159, No.3-4, pp. 255-272, Feb. 2004.

[4] S. K. M. Wong, and W. Ziarko, "On optimal decision rules in decision table," Bulletin of Polish Academy of Sciences, Vol. 33, pp. 693-696, 1985

[5] A. Skowron, and C. Rauszer, "The discernibility matrices and functions in information systems," in R. Slowinski(Ed.), Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory, Kluwer, Dordrecht, 1992.

[6] K. Y. Hu, et al., "Feature ranking in rough sets," AI Communications, Vol. 16, No. 1, pp. 41-50, Jan. 2003.

[7] G. Y. Wang, et al., "A Comparative Study of Algebra Viewpoint and Information Viewpoint in Attribute Reduction," Fundamenta Informaticae, Vol. 68, No. 3, pp. 289-301, Sep. 2005.

[8] D. Q. Miao, and J. Wang, "An information based algorithm for reduction of knowledge," Proceedings of the 1997 IEEE International Conference on Intelligent Processing Systems (ICIPS'97), Beijing, China, 28-31 October, 1997, Vol. 2, pp. 1155-1158.

[9] G. Y. Wang, et al., "Decision Table Reduction based on Conditional Information Entropy," Chinese Journal of Computer, Vol. 25, No. 7, pp. 759-766, 2002.

[10] J. Y. Liang, and Z. B. XU, "The Algorithm on knowledge reduction in incomplete information system," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10, No. 1, pp. 95-103, 2002.

[11] J. Wroblewski, "Finding minimal reducts using genetic algorithms," Proceedings of the Second Annual Join Conference on Information Sciences, Wrightsville Beach, September 28-October 1, 1995, pp.186–189.

[12] X. Y. Wang, et al., "Feature selection based on rough sets and particle swarm optimization," Pattern Recognition Letters, Vol. 28, No. 4, pp. 459-471, Mar. 2007.

[13] D. Q. Miao, et al., "Relative reducts in consistent and inconsistent decision table of the Pawlak rough set

model," Information Sciences, Vol. 179, No. 24, pp. 4140-4150, Dec. 2009.

[14] J. L. Yang, et al., "Attributes Reduction Based on Important Degree of Attributes in Incomplete Information System," Proceedings of the 18th International Conference on Fuzzy Systems, Korea, August, 2009, pp. 621-625.

[15] Z. Pawlak, "Rough sets: Theoretical aspects of reasoning about data," Kluwer Academic Publishers, Boston, 1991.

[16] M. Kryszkiewicz, "Rough set approach to incomplete information systems," Information Science, Vol. 112, No. 1-4, pp. 39-49, 1998.

[17] C. J. Merz, and P. Murphy, UCI repository of machine learning database, http://archive.ics.uci.edu/ml.

[18] A. Ohrn, and J. Komorowski, "ROSETTA-A rough set toolkit for analysis of data," Proceeding of Third International Joint Conference on Information Sciences (JCIS'97), Durham, NC, USA, March 1-5, 1997, Vol. 3, pp. 403-407.