A New Algorithm For Attribute Reduction Applied To Product Configuration

Ma Junyan[1, a]

[1]School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China.

[a]majunyanbyr@163.com

Keywords: Rough set,attribute reduction,non-compatible core,non-positive region,product configuration

Abstract. Attribute reduction is important to the mining of product configuration rules, the accuracy of attribute reduction will directly affect the subsequent mining of configuration rules.In this paper,the non-positive region in incompatible configuration decisions table is analyzed.A new attribute reduction algorithm is proposed,which is applied to attribute reduction in incompatible information decision table.At last ,the new algorithm is verified in product configuration instance decision table and the reduction effect is good.

1. Introduction

The reduction of attributes is necessary to discovery the configuration rules in decisions table. The accuracy of attribute reduction affects the configuration rules directly. Rough set is an effective tool to attribute reduction. Many researchers have studied the importantance of rough set in attribute reduction,and proposed numbers of valid reduction algorithm. However, the minimum attribute reduction is proved to be NP-Hard problem, it is generally use heuristics to identify optimal or sub-optimal reduction.

Reduction divided into two categories: 1. Reduction of algebraic theory, this kind of attribute reduction just only guarantee the indistinguishability of compatible data not changed, the relevant content references [1,2,3]; 2. Reduction of information theory, this reduction ensures the indistinguishability of all data does not changed. Thus, the definition of reduction in the information theory is complete, the relevant content references [4,5,6].

In this paper, a new reduction algorithm is proposed. This new heuristic algorithm is based on information theory,the initial set of algorithm is incompatible core [7] This algorithm can be applied to attribute reduction of inconsistent decision table. Experimental results shows that the algorithm is superior to the method based on conditional entropy.

2. Basic Concepts

The basic concepts of Rough sets, such as reduction, core, and the upper and lower approximation of the set, please refer to the literature. In this paper, some needed symbol is recommended[8],in addition, some new concepts are proposed.

Definition 1.(decision table DT (decision table) we formally known as quadruple $DT = (U, C \cup D, V, f)$is a decision table, which U: $U = \{x1, x2, \ldots, xn\}$ is non-empty finite set of objects, called domain; $C \cup D$: $C = \{\alpha | \alpha \subset C\}$is conditional attribute set.Each $\alpha_j \subset C(1 \leq j \leq m)$is a simple attribute of C; $D = \{d \in D\}$ is decision attribute set, and $C \cap D = \emptyset, C \neq \emptyset, D \neq \emptyset$; V: $V = \cup V\alpha(\forall\alpha \in C \cup D\}$ is the information function of decision table, f_a is the information function of attributes a. When the IND (C) □IND (D), the decision table is compatible, otherwise, it is incompatible, where IND(C), IND(D) denote condition equivalence class and decision equivalence class.

Definition 2. $C, D \subset A$ are two attribute sets,$X \in U/D, POS_p(Q) - \cup_{X \in U/Q} P_-(X)$ is the positive region of U/D.The non-positive region and positive region of Q are complementary set, the non-positive region marked as $U_2 = U - POS_p(Q)$,the positive region marked as $U_1 = POS_C(D)$.

Definition 3.(dependence of knowledge), $K = (U, S), \forall P, Q \in IND(K)$ is a knowledge base,then

$$\gamma_P(Q) = k = \frac{|POS_P(Q)|}{|U|} = \frac{|\cup_{X \in U/Q} \underline{P}_-(X)|}{|U|} \quad (1)$$

is the dependence between P and Q.

Definition 4 (importance of attribute), $DT = (U, C \cup D, V, f), \forall \beta \in C, \forall \alpha \in C - B$,then

$$sig(\alpha, B; D) = \gamma_{IND(B \cup \{\alpha\})}(D) - \gamma_{IND(B)}(D) = \frac{|POS_{B \cup \{\alpha\}}(D) - POS_B(D)|}{|U|} \quad (2)$$

3. Property in non-positive region

Definition 5. $S = (U, C, D)$,in the decision table, $U - POS_C(D)$ is a non positive region,$U \quad POS_C(D)$ is the boundary region of D relative to C.

Based on the above definitions 3 and 5, a new definition is proposed. The dependence of knowledge marked $\delta_P(Q)$.

Definition 6.(dependence of knowledge based on boundary region)

$$\delta_P(Q) = k^- = \frac{|U - POS_P(Q)|}{|U|} = \frac{|U - \cup_{X \in \frac{U}{D}} \underline{P}_-(X)|}{|U|} \quad (3)$$

$|U|$: global region,$|U - \cup_{X \in \frac{U}{D}} \underline{P}_-(X)|$: the classification based on P cannot accurately divide into Q

Definition 7 (the importance of attribute based on boundary region)

$$sig^-(\alpha, B; D) = |\delta_B(Q) - \delta_{B \cup \{\alpha\}}(D)| \quad (4)$$

Definition 8 [7] In an inconsistent decision table, if a condition attribute is removed, a combination of the partition block containing the incompatible elements is considered, then the attribute is the core of incompatible region,named incompatible core.

$$m_{ij} = \begin{cases} \{a \in C \wedge a(x_i) \neq a(x_j)\} \\ D(x_i) \neq D(x_j) \wedge (x_i \in U_2 \wedge x_j \in U_2) \\ \emptyset \quad ,others \end{cases} \quad (5)$$

$U_2 = U - POS_C(D)$.m_{ij} is a single property, which is incompatible core.

4. The new algorithm

Input: decision table $DT = (U, C \cup D, V, f)$.

Output: attribute reduction based on information theory.

Steps

Step 1:Calculate the positive region:$U_1 = POS_C(D)$,non-positive region: $U_2 = U - POS_C(D)$;

Step2:According to the formula (5),calculate $CORE_D(C)$,$B = CORE_D(C)$,if $U - POS_B(D) = U - POS_C(D)$, turn to step 5;

Step 3: According to the formula (4),calculate the importance of each condition attribute, select the most important attribute α_m (If there are multiple properties at the same time meet the maximum, then choose one to make categories minimul based on $\frac{U}{R} \cup \{\alpha_m\}$), $R = R \cup \{\alpha_m\}$;

Step 4: If $U - POS_B(D) \neq U - POS_C(D)$, go to Step 3. Otherwise, go to step 5;

Step 5: Output RED_C (D), the algorithm ends;

5. Case analysis

In this section, the new algorithm is applied to the product configuration process:attribute reduction. The integrity of the proposed reduction algorithm is verified. Table 1 is the material configuration information of single-stage centrifugal pumps. Single-stage centrifugal pumps can be divided into six main modules[9],expressed as {impeller, pump, body suspension, bearings, mechanical seal rotor},corresponding to condition attributes {a, b, c, d, e, f}, the attribute value 0 equivalent to cast iron,1 equivalent to cast steel. Decision attribute is the corrosion resistance of pump,

corresponding to {D}, the property value 0 equivalent to poor corrosion resistance, 1 equivalent to better corrosion resistance. (numx) in Table 1 means the quantity of object in equivalent class.

Table 1.Configuration information of Centrifugal pump

U	a	b	c	d	e	f	D
Z_1	0	0	1	0	1	0	0(10x)
Z_2	1	1	1	1	1	0	1(40x)
Z_3	1	0	1	1	0	0	1(10x)
Z_4	0	0	1	0	0	1	1(30x)
Z_5	0	0	1	0	1	1	1(50x)
Z_6	0	0	1	0	1	1	0(10x)
Z_7	1	0	1	0	1	1	0(10x)
Z_8	1	0	1	0	1	1	1(50x)

Incompatible categories are $\{Z_8, Z_6\}$ and $\{Z_7, Z_8\}$,

$$\frac{|\{Z_8, Z_6\}|}{|U|} = \frac{|\{Z_7, Z_8\}|}{|U|} = 2/7$$

They accounted for the same percentage of global data. Using Hu matrix algorithm, Information Entropy algorithm and the new algorithm to reduce the attributes in decision table 1 respectively,comparative results are shown in Table 2.

calculation process: Step1: $U_1 = \{Z_1, Z_2, Z_3, Z_4\}$, $U_2 = \{Z_5, Z_6, Z_7, Z_8\}$;

Step2: $B = CORE_D(C) = \{a\}$; $U - POS_B(D) \neq U - POS_C(D)$

Step3: $\delta_P(Q) = 1$; $sig(b, B; D) = 4/24$; $sig(c, B; D) = 0$; $sig(d, B; D) = 5/24$; $sig(e, B; D) = 8/24$; $sig(f, B; D) = 6/24$; so $\{a_m\} = \{e\}$;

Step4: when $B = \{a, e, f\}$, $U - POS_B(D) = U - POS_C(D)$;

Step5: Output $RED_C(D) = \{a, e, f\}$;

Table 2 Reduction results

	Hu matrix	Information Entropy	new algorithms
Reduction results	{aef} {bef} {def}	{aef} {bef} {def}	{aef}

$U/\{a, e, f\} = \{\{Z_1\}, \{Z_2\}, \{Z_3\}, \{Z_4\}, \{Z_5, Z_6\}, \{Z_7, Z_8\}\} = U/C$;

$U/\{b, e, f\} = \{\{Z_1\}, \{Z_2\}, \{Z_3\}, \{Z_4\}, \{Z_5, Z_6, Z_7, Z_8\}\} \neq U/C$;

$U/\{d, e, f\} = \{\{Z_1\}, \{Z_2\}, \{Z_3\}, \{Z_4\}, \{Z_5, Z_6, Z_7, Z_8\}\} \neq U/C$;

Hu's algorithm and Information Entropy algorithm:Although the result of Hu's algorithm contains all suitable reduction results, we can not distinguish which one is the complete reduction from all reduction results.

Information Entropy algorithm: Information Entropy has neglected the occurrence of a merger in the division of some of the incompatible parts, and the ratio of which is not changed.

New algorithms: By the reduction {aef} ,it is easy to known that impeller, seal and rotor are important to the overall corrosion resistance. The result of reduction really matches the fact. The new algorithm can ensure global data indiscernibility relation is changeless, the reduction results with the original decision table contains information on exactly the same.In conclusion,the reduction result of new algorithm is best.

6. Conclusion

In this paper, a new algorithm is proposed,the initial set is non compatible core, the heuristic condition is based on non positive region.The final reduction can ensure that the indistinguishability of all the data in the domain are kept unchanged. This reduction algorithm is suitable for the process of attribute reduction in product configuration, and the effect is good.

References

[1]Xiaohua Hu,Nick Cercone.Learning in Relational Databases:A RoughSet Approach [J] .Computational Intelligence, 1995, 11 (2): 323-337.

[2]Dongyi Ye ,Zhaojiong Chen.A new discernibility matrix and the computation of a core [J] - Acta Electronica Sinica,2002,30(07):1086-1088.

[3]Guoshun Huang,Yunsheng Liu. Improvement of Discernibility Matrix and the Computation of a Core [J].Journal of Fudan University (Natural Science),2004,43(05): 865-868,873.

[4]Guoyin Wang, Yu Hong,Dachun Yang. Decision Table Reduction Based on Conditional Information Entropy[J].CINESE J. C0MPUTERS,2002,25(7):759-766.

[5]Guoyin Wang. Calculation Methods for Core Attributes of Decision Table [J].CHINESE JOURNAL OF COMPUTERS, 2003,26(5): 611-615.

[6]Xuefeng Wei,Li Sun. ALgorithm of Discernibility Matrix Based on Distribution Reduction [J]. Science Technology and Engineering,2009,9(18):5373-5378.

[7]Fengjuan Chen. Methods for calculating core attributes of inconsistent decisiontable [J].COMPUTER ENGINEERIING AND DESIGN,2012,33(3):1187-1191.

[8]Duoqian Miao,Guodao Li. Rough Sets Theory, Algorithms and Applications[M]. Tsinghua University Press, 2008.

[9]Yibin Li.Hua Huang,Yi Liu,Xiaorui Chen. Design of Centrifugal Pump Based on Reconfigurable Module Method [A],The fourth national conference on water conservancy machinery and systems[C],2011,282-285.