An Improved Attribute Reeducation Algorithm Based on Mutual Information for Non-Care Information System

Sumin Yang, Shaochong Feng, Jing Chen and Hongli Yuan

ABSTRACT

For an information system the attribute reduction is one key to choose some important indexes, but the core attribute is the foundation for the present heuristic attribute reduction algorithms of rough set theory based on discernibility matrix and mutual information theory. When we apply these algorithms on the non-core information system, there will be the following problems, such as too much calculation problem, excessive reduction, or insufficient reduction. According to the characteristics of non-core information system, we propose one new heuristic attribute reduction algorithm based on mutual information, in which the evaluation of attribute important degree depends on the increment of mutual information and information entropy. And the attribute with the largest attribute important degree is selected for the initial core attribute, by which the problem that the random selection attribute cause the computational complexity is solved. By the proposed algorithm we cannot only improve the efficiency of attribute reduction, but decrease the number of attribute reduction. Both the theoretic analysis and the simulation experiments verify that the proposed algorithm is validity.\footnote{Sumin Yang, Shaochong Feng, Jing Chen, Hongli Yuan, Equipment Simulation Training Center, Shijiazhuang, Campus of Army Engineering University of PLA, Shijiazhuang 050003, China}

INTRODUCTION

Rough set theory was first proposed by Poland mathematician Z. Pawlak in 1982, which can effectively analyse and deal with the imprecise, inconsistent, incomplete or other imperfect information so as to find out the implied knowledge and rules. The most significant difference compared with other uncertainty and
imprecise theory is that it can objectively describe or deal with uncertainty problem without providing any prior information of the processing data, such as the probability distribution of the statistics, membership function of the fuzzy theory.

Attribute reduction takes a crucial role in analysing the attribute the importance of different attributes in given data, whose purpose is to delete the unnecessary attribute under the condition of ensuring unchanged classification ability of information system. However, for an information system, Wong S. K. M and Ziarko [1] have proved that the attribute reduction is a non-deterministic polynomial problem, so most of researchers make use of the heuristic reduction algorithm so as to obtain the optimal or sub-optimal attribute reduction. But the researchers found that effective attribute reduction can be obtained if we can establish the relationship between knowledge and information, and study attribute reduction with the point of view of information entropy. Skowron[2] put forward one attribute reduction algorithm based on discernibility matrices, Articles [3] presented some improved attribute reduction algorithm based on discernibility matrices, which have lower computational complexity and storing capacity. On the basis of conditional information entropy, articles [4] studied the computation of a core and attribute reduction in distributed environment. Teng[5] presented a new reduced definition which integrates the complete and incomplete information systems into the corresponding reduced algorithm. Miao[6] proposed the knowledge reduction algorithm which is based on the mutual information between the conditional attributes and decision attributes; Articles [7-8] proposed rough sets attribute reduction algorithm based on mutual information, which can make use of heuristic information to reduce the search space, and can shorten the search time as far as possible, and can finally get an optimal or approximate optimal solution. But the superiority of the attribute reduction algorithm based on mutual information depends on the relative core attribute of decision table. In actual information system, there will be a lack of core attribute. When the core attribute are empty, we must compute mutual information every time when we add one attribute into reduction attributes, the computational complexity increases significantly. Therefore we propose a new attribute reduction algorithm with using the information entropy and mutual information increment, and set the attribute with the largest attribute importance and mutual information among all attribute as the core attribute, so the problem that the initial attribute selected randomly causes the computational complexity increasing is solved, and attribute reduction velocity can become faster than some other algorithms, and the number of attribute reduction is relatively small. The simulation experiments verify that the proposed algorithm can ensure the efficiency and accuracy of attribute reduction, but also can guarantee the quality of reduction attribute.
THE BASIC CONCEPTS OF ROUGH SET THEORY

Definition 1: $S = (U, A, V, f)$ is set to an information system. Among them, $U = \{u_1, u_2, \cdots, u_n\}$ is non-empty finite sets which is called the domain space, $A = \{a_1, a_2, \cdots, a_n\}$ is non-empty finite attribute set, which is called the attribute set, $V = \cup A$, $a \in A$, $V_a$ is attribute’s domain range, $f: U \times A \rightarrow V$ is the information function, when $x$ is $a$, $x$ has unique value in $V$. On the side, for sequence $C(c_1(x), c_2(x), \cdots, c_n(x))$ and sequence $D(d_1(x), d_2(x), \cdots, d_n(x))$, $A = C \cup D, C \cap D = \phi$, $S = (U, A, V, f)$ is called as decision table of the information system. $c_1(x), c_2(x), \cdots, c_n(x)$ is called as the condition attribute set.

Definition 2: For the given knowledge representation system $s = (U, A, V, f), \phi \neq B \subseteq A$, the in-discernable relationship of any attribute is as follows:

$$IND(B) = \{(x, y) \in U \times U : \forall a \in B(f(x, a) = f(y, a))\}$$ (1)

Definition 3: $U$ is a domain set, $P$ and $U$ is two equivalent relation of domain $U$ (knowledge), $U \backslash ind(P) = \{x_1, x_2, \cdots, x_n\}$, $U \backslash ind(Q) = \{y_1, y_2, \cdots, y_m\}$, then the probability distribution that $P$ and $Q$ effect on the $U$ is defined as follows:

$$\begin{pmatrix} x_1 \ x_2 \ \cdots \ x_n \\ p(x_1) \ p(x_2) \ \cdots \ p(x_n) \end{pmatrix} ; \begin{pmatrix} y_1 \ y_2 \ \cdots \ y_m \\ p(y_1) \ p(y_2) \ \cdots \ p(y_m) \end{pmatrix}$$ (2)

Among them, $p(x_i) = \frac{|U_i|}{|U|}, i = 1, 2, \cdots, n$, $p(y_j) = \frac{|U_j|}{|U|}, j = 1, 2, \cdots, m$; the symbol $|E|$ is the base of $E$.

Definition 4: According to the information theory, the information entropy of knowledge $P$ is $H(P) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$, the conditional entropy $H(Q|P)$ of the knowledge $P$ relative to $Q$ is:

$$H(Q|P) = -\sum_{j=1}^{m} p(x_j) \sum_{i=1}^{n} p(y_i|x_j) \log p(y_i|x_j)$$ (3)

The mutual information $I(P;Q)$ of the knowledge $P$ relative to $Q$ is:

$$I(P;Q) = H(P) - H(Q|P);$$ (4)

Definition 5: $U$ is a domain set, $P$ and $U$ is two equivalent relation of domain $U$ (knowledge), if $Ind(P) = Ind(Q)$, then $H(P) = H(Q)$.

Definition 6: The independent necessary and sufficient conditions of equivalent relation $P$ of domain $U$ is that there is $H(P) - \{R\} > 0$ for any $R \in P$. 

438
THE ATTRIBUTE IMPORTANCE BASED ON THE MUTUAL INFORMATION

In the process of decision, we pay attention which condition attribute is the most important for the last decision, so we must consider the mutual information between condition attribute and decision attribute. The article[7] proposed the method that obtain the attribute importance by the increasing amount of mutual information with adding one attribute. It is defined as follows:

$$SGF(a, R, Q) = I(Q; R \cup \{a\}) - I(Q; R) = H(Q|R) - H(Q|R \cup \{a\})$$ (5)

Based on the above formula(5), the chosen attributes are that there is more number in the domain, but from the information theory, it is to select the one which is chaotic, but the selected attributes are not useful for the decision.

In view of the above problems, the article [8] has made the improvement to the importance of attributes, which is defined as follows:

$$SGF(a, R, Q) = I(Q; R \cup \{a\}) - I(Q; R) / H(a) = (H(Q|R) - H(Q|R \cup \{a\})) / H(a)$$ (6)

The improved method not only considers the increment of mutual information after adding the attribute, but also considers its own information entropy. When the mutual information increment is equal, the smaller $H(a)$ is, the higher attribute importance degree is. But the above algorithm depends on the core attributes, when there is no core attribute, formula (6) become into:

$$SGF(a, R, D) = I(D) - I(D\mid a) / H(a) = I(a, D) / H(a)$$ (7)

In order to solve the problem, we improve formula (6), firstly the attribute importance of each condition attribute is calculated by formula (7), and we set the attribute with the maximal attribute important degree as the core attribute, the attribute important degree formula becomes:

$$SGF(a, R', Q) = I(Q; R' \cup \{a\}) - I(Q; R) / H(a) = (H(Q|R') - H(Q|R' \cup \{a\})) / H(a)$$ (8)

AN IMPROVED ATTRIBUTE REDUCTION ALGORITHM BASED ON MUTUAL INFORMATION

On the basis of the formula (8) in the section 3, the author proposes a new attribute reduction algorithm. Firstly the attribute importance of each condition attribute is calculated by formula (7), and we select the attribute with the maximal attribute importance for the core attribute, the evaluation of attribute important degree depends on two factors, one is the increment of mutual information the other is information entropy, the added attribute into the reduction sets is the one
that both the increment of mutual information and attribute important degree are the biggest. The specific algorithm description is as follows:

The Input: A compatible decision tablesystem, $C$ is condition attributes set, $D$ is the decision attribute, $\mathcal{V}$ and is the domain.

The Output: One attribute reduction sets;

(1) The mutual information $I(C; D)$ is calculated between condition attribute $C$ and decision attribute set $D$;

(2) All the attribute important is calculated by the formula (7), and set the attribute with maximal attribute importance degree for the core attribute $R^*$;

(3) Let $R = R^*$, the process is performed on the attribute set $R^* = C - R$, $C^* = C - R$ as follows:

① For each attribute $c_i \in C^*$, we calculate $I(\mathcal{Q}c_i) - I(\mathcal{Q}) / H(\mathcal{Q})$, and select the one that has maximum value $C_i$, if there is the same value for multiple attributes, we choose one which comes the earliest, then $R = R \cup \{c_i\}$, $C = C - R$.

② Then we judge whether $I(C; D)$ and $I(R; D)$ is equal, if they are the same, then the next step goes to (3), otherwise goes to ①.

(4) $R$ is a reduction result, and we output it.

THE SIMULATION EXPERIMENT

In order to verify the validity of the algorithm, we validate our algorithm by one non-core information system shown in Table 1:

From table 1, we can see that the information system has seven condition attributes $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$, ten experts give those evaluation results $\mathcal{U} = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, decision attributes are $\{d\}$, the value of set $C$ is set as $\mathcal{V} = \{1, 2, 3, 4\}$, the value of set $D$ is set as $\{1, 2, 3\}$, the in-discrimable relationship sets of attribute $D$ is $IND(D) = \{(x_i, x_2, x_3, x_4, x_5), \{x_3, x_6, x_8, x_{10}\}, \{x_i\}\}$, the below is the process in accordance with the algorithm of section 3:

<table>
<thead>
<tr>
<th>(U)</th>
<th>(c_1)</th>
<th>(c_2)</th>
<th>(c_3)</th>
<th>(c_4)</th>
<th>(c_5)</th>
<th>(c_6)</th>
<th>(c_7)</th>
<th>(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
<td>(X)</td>
</tr>
<tr>
<td>(c_1)</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(c_2)</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>(c_3)</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>(c_4)</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(c_5)</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>(c_6)</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>(c_7)</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>(D)</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

TABLE I. THE NON-CORE INFORMATION SYSTEM.
TABLE II. THE ATTRIBUTE MUTUAL INFORMATION AND THE ATTRIBUTE IMPORTANT DEGREE.

<table>
<thead>
<tr>
<th></th>
<th>c₁₁</th>
<th>c₂₁</th>
<th>c₃₁</th>
<th>c₄₁</th>
<th>c₅₁</th>
<th>c₆₁</th>
<th>c₇₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual Information increment</td>
<td>1.07548875</td>
<td>1.295461844</td>
<td>1.32192809</td>
<td>1.397416845</td>
<td>0.342609804</td>
<td>0.770950594</td>
<td>0.770950594</td>
</tr>
<tr>
<td>Attribute important degree</td>
<td>0.6107</td>
<td>0.6573</td>
<td>0.7159</td>
<td>0.4631</td>
<td>0.2525</td>
<td>0.4175</td>
<td>0.4175</td>
</tr>
</tbody>
</table>

TABLE III. THE ATTRIBUTE MUTUAL INFORMATION AND THE ATTRIBUTE IMPORTANT DEGREE.

<table>
<thead>
<tr>
<th></th>
<th>{c₃,c₅}</th>
<th>{c₃,c₆}</th>
<th>{c₃,c₇}</th>
<th>{c₃,c₈}</th>
<th>{c₅,c₆}</th>
<th>{c₅,c₇}</th>
<th>{c₅,c₈}</th>
<th>{c₆,c₇}</th>
<th>{c₆,c₈}</th>
<th>{c₇,c₈}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual information increment</td>
<td>2.3219</td>
<td>1.9142</td>
<td>2.6464</td>
<td>1.7066</td>
<td>2.9219</td>
<td>2.2879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute important degree</td>
<td>0.9207</td>
<td>0.9054</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.8645</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE IV. THE ATTRIBUTE MUTUAL INFORMATION AND THE ATTRIBUTE IMPORTANT DEGREE.

<table>
<thead>
<tr>
<th></th>
<th>{c₃,c₅}</th>
<th>{c₃,c₆}</th>
<th>{c₃,c₇}</th>
<th>{c₃,c₈}</th>
<th>{c₅,c₆}</th>
<th>{c₅,c₇}</th>
<th>{c₅,c₈}</th>
<th>{c₆,c₇}</th>
<th>{c₆,c₈}</th>
<th>{c₇,c₈}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutual information increment</td>
<td>3.3219</td>
<td>3.0219</td>
<td>3.1219</td>
<td>2.9219</td>
<td>2.9219</td>
<td>3.1219</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attribute important degree</td>
<td>1.0</td>
<td>0.968</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to formula (4), the mutual information $I(C; D)$ is calculated between condition attribute $C$ and decision attribute set $D$:

$I(C; D) = H(D) - H(D|C) = 3.3219$.

According to the formula (6), the attribute important degree of each attribute in $C$ sets is calculated, the results is listed in the table II, from the table we can see that the attribute $c₃$ has the most attribute important degree, so we set $c₃$ for core attribute, i.e. $R' = \{c₃\}$.

We Set $R = R'$, and perform the following operations on the attribute $R' = C - R$, $C' = C - R$.

Then we calculate the attribute important degree of the $\{c₃\}$ sets, the results is shown in Table III, it can be seen from the table III, $c₄$, $c₅$ and $c₆$ have the same attribute important degree, but $c₆$ has the maximal mutual information increment, so we add into the attributes reduction sets;

But $I(R; D) = 2.9219$, it is not equal to $I(C; D)$.

In accordance with the algorithm procedure of section 3, we must add other attribute into the reduction attribute sets, and calculate the attribute important degree of the $\{c₃\}$, the results is shown in Table IV. It can be seen from the table IV, $\{c₃,c₄\}$ has the maximal mutual information increment, so we add $c₄$ into the attributes reduction sets. Then $R = R \cup \{c₆,c₄\}$, $C' = C - R$. 

441
We can see that \( I(R; D) = 3.3219 \), which is the same as \( I(C; d) \), so we can terminate this algorithm. \( \hat{R} = \{c_1, c_2, c_6\} \) is one reduction attribute for the non-core information system.

But according to the attribute reduction algorithm proposed by article[8], the attribute reduction \( \{c_1, c_2, c_3, c_1, c_5, c_6\} \) is obtained by the algorithm because the non-core factor is not considered. With the simulation experiment results, we can see that the proposed algorithm can reduce the number of attribute reduction 57%, and decrease attribute reduction time about 42% while maintaining the information system classification ability unchanged. The experiment results validate the proposed algorithm has good attribute reduction quality and high efficiency for non-core information systems.

CONCLUSIONS

According to the characteristics of non-core information system, we propose a new heuristic attribute reduction algorithm based on rough set theory in this paper, which makes use of information entropy and mutual information increment to evaluate attribute important degree. The most important attribute is selected with both the maximal mutual information and the highest attribute important degree, so we can ensure that the added attribute is surely the most influence on decision attribute. The simulation experiment results also verify the correctness and validity of the proposed algorithm.

REFERENCES