Quantitative Analysis of the Difference between the Algebra View and Information View of Rough Set Theory

J.J. An, L. Chen, G.Y. Wang Y. Wu
Institute of Computer Science and Technology
Chongqing University of Posts and Telecommunications
Chongqing, 400065, P. R. China
crssc@cqupt.edu.cn wanggy@cqupt.edu.cn

Abstract
The attribute core of a decision table is often the start point and key of many decision information system reduction procedures based on rough set theory. The algebra view and information view are two main views and methods of rough set theory. In this paper, based on the problem of calculating the attribute core of a decision table, we will study the relationship between the algebra view and information view. Through simulation experiment, we find the quantitative difference of the attribute core of a decision table in both views of rough set theory. Especially, we find that their difference will be maximized in decision tables containing much inconsistent information.

1. Introduction
Rough set theory has been applied in such fields as machine learning, data mining, etc., since Professor Z. Pawlak developed it in 1982 [1, 2]. Reduction of decision table is one of the key problems of rough set theory. The attribute core of a decision table is always the start point of information reduction. There are two main views and methods about rough set theory, that is, the algebra view and information view [3, 4, 5, 6]. In this paper, we will study the problem of calculating the attribute core of a decision table. Based on former research results on this problem in the algebra view and information view of rough set theory [7-11], we will further study their relationship. Our simulation experiments give a quantitative answer of the difference between these two views of rough set theory. Especially, we find that there will be great difference of the attribute cores of an inconsistent decision table in these two views. This result is much useful for uncertain information system reduction.

2. Basic Concepts
For the convenience of discussion, we introduce some basic notions about attribute reduction and attribute core of rough set theory at first.

Definition.1 An information system is defined as $S=<U, R, V, f>$, where $U$ is a finite set of objects and $R=C\cup D$ is a finite set of attributes. $C$ is the condition attribute set and $D$ is the decision attribute set, $V=\bigcup_{a \in V}$ is a union of the domain of each attribute of $R$. Each attribute has an information function $f: U \times R \rightarrow V$.

Definition.2 Given an information system $S=<U, R, V, f>$, let $X$ denote a subset of elements of the universe $U(X \subseteq U)$. The lower approximation of $X$ in $B(B \subseteq R)$ is defined as $B_-(X)$, the upper approximation of $X$ in $B$ is defined as $B_+(X) = \{Y | Y \subseteq U \wedge \forall_{b \in B} (b(x) = b(y)) \}$. where, $U|IND(B)=\{X | X \subseteq U \wedge \forall_{y \in Y} \exists_{x \in X} (b(x)=b(y)) \}$.

Definition.3 Given an information system $S=<U, R, V, f>$ and an attribute $r \in P(P \subseteq R)$, if $IND(P \setminus r) = IND(P)$, $r$ is said to be dispensable, otherwise indispensable.

Definition.4 Given an information system $S=<U, R, V, f>$ and an attribute $r \in P(P \subseteq C)$, if $IND(P \setminus r) = IND(P)$, $r$ is said to be dispensable, otherwise indispensable.

Definition.5 Given an information system $S=<U, R, V, f>$, the $C$-positive region $POS_C(D)$ is defined as $POS_C(D) = \{x \in U : \exists X | X \subseteq U \wedge IND(D) \wedge x \in X \subseteq C \}$.

Definition.6 Given an information system $S=<U, R, V, f>$, an attribute set $T \subseteq P(P \subseteq R)$ is a relative reduction of $P$ with reference to an attribute set $Q$ if $T$ is relatively orthogonal with reference to the attribute set $Q$ and $POS_T(Q)=POS_P(Q)$. We use the term $RED_Q(P)$ to denote the family of relative reductions of $P$ with reference to $Q$. $CORE_Q(P) = \cap RED_Q(P)$ is called the $Q$-core of $P$.

WSS03 Applications, Products and Services of Web-based Support Systems 159
3. Attribute Core in the Algebra View

Hu developed a method to calculate the attribute core of a decision table based on Skowron’s discernibility matrix [7].

**Definition.7** For a set of attributes \( B \subseteq C \) in a decision table \( S=(U, C \cup D, V, f) \), the discernibility matrix can be defined by \( C_0(B) = \{ C_0(i, j) \}_{i, j \in n} \), where

\[
C_0(i, j) = \begin{cases} a_k \mid a_k \in P \land a_k(X_i) \neq a_k(X_j), X_i, X_j \in U \land d(X_i) \neq d(X_j) & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}
\]

for \( i, j = 1, 2, \ldots, n \).

Hu drew the following conclusion in [7]: \( |C_0(i, j)| = 1 \) if and only if the attribute in \( i \) belongs to \( \text{CORE}_D(C) \). This conclusion is used in many later documents.

Ye proved Hu’s conclusion is not true in inconsistent decision tables by giving a counterexample. He developed an improved method to calculate the attribute core through improving the definition of discernibility matrix in the following way.

**Definition.8** For a set of attributes \( B \subseteq C \) in a decision table \( S=(U, C \cup D, V, f) \), the discernibility matrix can be defined by \( C_0'(B) = \{ C_0'(i, j) \}_{i, j \in n} \), where

\[
C_0'(i, j) = \begin{cases} C(i, j) & \text{if } |D(x_i)| \cdot |D(x_j)| = 1, \\ \emptyset & \text{else}, \end{cases}
\]

for \( i, j = 1, 2, \ldots, n \), \( |D(x_i)| = |\{ y \mid y \in [X_i] \} | \).

In [8], Ye drew another conclusion: \( |C_0'(i, j)| = 1 \) if and only if the attribute \( i \) belongs to \( \text{CORE}_D(C) \).

Yi proved that Ye’s method could be used to calculate the attribute core of a decision table in the algebra view of rough set theory [9].

4. Attribute Core in the Information View

For the convenience of later discussion, we introduce some basic concepts and theorems about the information view of rough set theory here [3].

**Definition.9** Given an information system \( S=(U, C \cup D, V, f) \), and a partition of \( U \) with classes \( X_i \), \( 1 \leq i \leq n \). The entropy of attributes \( B \subseteq C \cup D \) is defined as

\[
H(B) = -\sum_{i=1}^{n} p(X_i) \log(p(X_i)),
\]

where \( p(X_i) = |X_i|/|U| \).

**Definition.10** Given an information system \( S=(U, C \cup D, V, f) \), the conditional entropy of \( D \) given \( B \subseteq C \cup D \) is defined as

\[
H(D|B) = -\sum_{i=1}^{n} p(X_i) \sum_{y \in [X_i]} p(Y_i|X_i) \log(p(Y_i|X_i)),
\]

where \( p(Y_i|X_i) = |Y_i \cap X_i|/|X_i| \), \( 1 \leq i \leq n, 1 \leq j \leq m \).

The following 3 theorems are proved in [4].

**Theorem 1** Given a relatively consistent decision table \( S=(U, C \cup D, V, f) \), an attribute \( r \in C \) is relatively reducible if and only if \( H(D|B) = H(D|B \cup \{r\}) \).

**Theorem 2** Given a relatively consistent decision table \( S=(U, C \cup D, V, f) \), the attribute set \( B \subseteq C \) is relatively reduced if and only if \( H(D|B) = H(D|B \cup \{r\}) \) for all \( r \in C \).

**Theorem 3** Given a relatively consistent decision table \( S=(U, C \cup D, V, f) \), attribute set \( B \subseteq C \) is a relatively reduct of condition attribute set \( C \) if and only if

1. \( H(D|B) = H(D|C) \), and
2. \( H(D|B) = H(D|B \cup \{r\}) \) for any attribute \( r \in B \).

**Definition.11** Given a decision table \( S=(U, C \cup D, V, f) \), attribute set \( B \subseteq C \) is a relatively reduct of condition attribute set \( C \) if and only if

1. \( H(D|B) = H(D|C) \), and
2. \( H(D|B) = H(D|B \cup \{r\}) \) for all \( r \in C \).

**Definition.12** \( \text{CORE}_D(P) = \text{RED}_D(P) \) is called the \( Q \)-core of attribute set \( P \).

The attribute core of an inconsistent decision table in the information view can be calculated with the following theorem and algorithm.

**Theorem 4** Given a decision table \( S=(U, C \cup D, V, f) \), attribute set \( B \subseteq C \) is a core attribute if and only if

1. \( H(D|B) = H(D|C) \), and
2. \( H(D|B) = H(D|B \cup \{r\}) \) for all \( r \in C \).

In [9], Wang developed the following algorithm for calculating the attribute core of a decision table in the information view of rough set theory.

**Algorithm 1**: Input: A decision table \( S=(U, R, V, f) \). Output: The attribute core of \( S \) in the information view, \( \text{CORE}_D(C) \).

Step 1. \( \text{CORE}_D(P) \) = \( \phi \).

Step 2. For each condition attribute \( r \in C \), do

If \( H(D|C) < H(D|C \cup \{r\}) \), then

\( \text{CORE}_D(C) \) = \( \text{CORE}_D(C) \cup \{r\} \).

Step 3. Stop.

5. Difference of the Attribute Core of a Decision Table in the Algebra View and Information View

In the algebra view of rough set theory, a condition attribute is reducible if and only if the lower approximation of at least one decision class of the decision table will be changed after deleting it. That is, a condition attribute is reducible if and only if the
consistent part of the decision table will be changed after deleting it.

In the information view of rough set theory, a condition attribute is reducible if and only if the conditional entropy of the decision table will be changed after deleting it. However, the conditional entropy of the consistent part of a decision table is always 0. All conditional entropy of a decision table results from its inconsistent part. Thus, a condition attribute should be reducible in the information view if and only if the probability distribution of the whole decision table will not be changed after deleting it.

Let the attribute core of a decision table in the algebra view is \( \text{CORE}_1 \), and \( \text{CORE}_2 \) in the information view, we can draw the following conclusions [9]:

1. If a decision table is consistent, its attribute core in the algebra view is equivalent to that in the information view, that is, \( \text{CORE}_1 = \text{CORE}_2 \).

2. If a decision table is inconsistent, its attribute core in the algebra view is included by that in the information view, that is \( \text{CORE}_1 \subseteq \text{CORE}_2 \).

In addition, Ye drew a further conclusion about the difference between the algebra view and information view of rough set theory in [10]: if a decision table is partially inconsistent, that is, there exists an object, say \( X_k \in U \), having \( D(X_k) > 1 \), but \( \min_i \{D(X_i), D(X)\} = 1 \) for any pair of objects of the decision table, its attribute core in the algebra view should be equivalent to that in the information view, that is, \( \text{CORE}_1 = \text{CORE}_2 \). This result is consistent with Ye’s method in definition 8.

### 6. The Quantitative Difference of the Attribute Cores between the Algebra View and Information View

For comparing the attribute cores in the algebra view and information view of rough set theory, we have done the following four simulation experiments.

#### Table 1 Experiment Parameters

<table>
<thead>
<tr>
<th>Experiment No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Condition Attribute</td>
<td>4</td>
<td>9</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Domain of Condition Attribute</td>
<td>0,1</td>
<td>0,1,2</td>
<td>0,1,2,3</td>
<td>0,1,2,3,4</td>
</tr>
<tr>
<td>Number of Decision Attribute</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

In each simulation experiment, a lot of decision tables with different scale (number of samples) and complexity (uncertainty degree) are tested in both views. Table 1 shows the four sets of experiment parameters. For example, in experiment 1, there are 4 condition attributes and 1 decision attribute, the domain of each condition attribute is \{0, 1\}, and the domain of each decision attribute is \{0, 1\}. When the scale of decision table is small, a lot of decision tables are randomly generated and tested in order to get more accurate statistic result. The maximal number of decision tables with the same scale is 10000, whereas when the scale of decision table is large, due to the limitation of the memory capability of computer, lesser decision tables are randomly generated and tested. The minimal number of decision tables with the same scale is 5. In our experiments, all attribute cores in the algebra view are calculated using the method of Ye [8], while the attribute cores in the information view are calculated using algorithm 1 in section 5. The average number of the core attributes of these decision tables is taken as the statistic value for the number of core attributes under each scale of decision tables. In our experiments, the number of samples of a decision table is set to be 1 to 1000000. The processes of the other 3 experiments are the same. The results of these 4 experiments are shown in figure 1 to 4. In these figures, the horizontal axis indicates the logarithm of the number of samples of a decision table (decision table scale), the vertical axis indicates the average number of the core attributes of the decision tables in the same scale.

In these figures, blue dot lines (- - -) indicate the result of the algebra view of rough set theory (\( \text{CORE}_1 \)), red dash lines (---) indicate the result of the information view (\( \text{CORE}_2 \)), black solid lines indicate that red line overlays blue line, that is, \( \text{CORE}_1 = \text{CORE}_2 \).
According to the above experiment results, we can draw the following conclusions:

1. When the scale of a decision table is very small, the decision table is consistent and does not contain any inconsistent information. The redundancy degree of its condition attributes is high. Its attribute cores in the algebra view and information view are all empty. Obviously, there is no difference between them.

2. When the scale of a decision table increases gradually, the inconsistent information of the decision table increases, and the redundancy degree of its condition attribute decreases gradually. The number of core attributes of the decision table increases gradually, at the same time, the attribute core of a decision table in the information view includes that in the algebra view.

3. When the scale of a decision table increases to some degree, all condition attributes of the decision table are no more redundant, both the attribute cores in the algebra view and information view are all the full set of its condition attributes. There is no difference between them.

4. When the scale of a decision table is large, the inconsistent information of the decision table is larger also. The lower approximation and the number of core attributes of the decision table decrease gradually in the algebra view, while in the information view, the attribute core is still all the full set of its condition attributes, since every condition attribute influences the distribution of samples of the decision table.

5. When the scale of a decision table increases very large, the decision table is totally inconsistent. The lower approximation of the decision table is empty. Its attribute core in the algebra view is empty too. However, its attribute core in the information view is still the full set of its condition attributes. Every condition attribute influences the distribution of samples of the decision table.

### 7. Conclusion

Calculation of the attribute core of a decision table is the key of rough set theory, and the base of information reduction. In this paper, based on the difference between the definitions of attribute core of a decision table in the algebra view and information view, we quantitatively analyze their difference under different conditions. Our simulation experiment results prove that the attribute core in the information view includes that in the algebra view. Especially, we find that there will be great difference of the attribute cores of a decision table in these two views if it contains
much inconsistent information. This result is much useful for uncertain information system reduction

Acknowledgements

This paper is partially supported by National Science Foundation of China (No.69803014), National Climb Program of China, Foundation for University Key Teacher by the State Education Ministry of China (No.GG-520-10617-1001), Scientific Research Foundation for the Returned Overseas Chinese Scholars by the State Education Ministry of China, and Application Science Foundation of Chongqing.

References