

Rules Mining Based on Rough Set of Compatible Relation

Weiyan Xu^{*1}, Ming Zhang², Bo Sun¹, Mengyun Lin¹, Rui Cheng¹

¹School of Mathematics and Physics, Jiangsu University of Science and Technology, Zhenjiang 212003, China

²School of Computer science and Engineering, Jiangsu University of Science and Technology, Zhenjiang 212003, China,

*Corresponding author, e-mail: xwy_yan@hotmail.com

Abstract

Rough set model based on tolerance relation, has been widely used to deal with incomplete information systems. However, this model is not so perfect because not all of the elements in a tolerant class are mutually tolerant, but they are all tolerant with the generating element of this class. To mend this limitation, the compatible relation is redefined, and then the concept of maximal complete compatible class in incomplete information system is presented for the purpose that any two elements in the same compatible module are mutually compatible. Furthermore, two methods are put forward in the interest of selecting optimal compatible class for an object, which can be used in knowledge reduction. Besides, coverings on universe produced by tolerance and compatible relations are deeply investigated and compared. Finally, a medical decision table is analyzed, some compact rules are mined.

Keywords: rough set, incomplete information system, tolerance relation, compatible relation, optimal compatible class, Reduction

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

Rough set theory has developed since Pawlak's paper [1-2] as a new mathematical tool for analyzing vague and imprecise descriptions of objects. It uses indiscernibility (equivalence) relation to represent classification. In recent years, rough set theory has been successfully applied to so many fields such as Artificial Intelligence, Data Mining, Machine Learning, Pattern Recognition, Knowledge Acquisition and so on [2-11]. Rough set proposed by Pawlak is based on the assumption of complete information systems, i.e. there are no unknown values in the information table. However, in practical applications, incomplete information systems can be seen everywhere for a lot of unpredictable reasons. Therefore, mining rules from incomplete information systems is one of the important directions for the development of rough set.

In general incomplete information systems, unknown values may have two different explanations: in the first case, all unknown values are "do not care" condition; in the second case, all unknown values are lost. In Reference [10], Grzymala-Busse firstly studied the unknown value ("do not care") from the viewpoint of rough set theory, consequently, Kryszkiewicz [12] transformed the indiscernibility relation to tolerance relation (reflective, symmetric). On the other hand, incomplete information systems in which all unknown values are lost, from the viewpoint of rough set theory, were studied for the first time in Reference [12], where two algorithms for rule induction were presented. Based on Grzymala-Busse's work, Stefanowski [13] advanced the non-symmetric similarity relation (reflective, transitive).

In this paper, all unknown values are looked as "do not care", that is to say, each unknown value could be replaced by all values from the domain of the attribute, therefore, what we have done are all based on the further investigation of tolerance relation. In the classification produced by tolerance relation, not all of the elements in the same tolerant class are mutual tolerant, but they are all tolerant with the generating element of this tolerant class.

Owing to such limitation of tolerance relation, a binary relation called compatible relation is re-defined. According to compatible relation, a complete covering on universe can be got. That is to say, any two elements in the same compatible class are mutually compatible. It is clear that this kind of classification of universe in incomplete information systems meets with the

practical applications more than that is based on tolerance relation. Furthermore, since any one object in the universe is likely to be included into two or more different compatible classes, two different methods are presented for choosing optimal compatible class.

2. Basic Concepts

An incomplete information system is a quadruple $S = \langle U, AT, V, f \rangle$, where U is a nonempty finite set of objects called universe and AT is a nonempty finite set of attributes, such that $\forall a \in AT : U \rightarrow V_a = \bigcup_{a \in AT} V_a$ where V_a is called the value set of a ; any attribute domain V_a may contain special symbol "*" to indicate that the value of an attribute is unknown; V is regarded as the value set of all attributes and then $V = \bigcup_{a \in AT} V_a$; let us define f as an information function such that $f(x, a) \in V_a$ for any $a \in AT$ and $x \in U$.

Definition 1. Let $S = \langle U, AT, V, f \rangle$ be an incomplete information system, $\forall A \subseteq AT$, a binary relation $SIM(A)$ can be defined as [12]:

$$SIM(A) = \{(x, y) : \forall a \in A, f(x, a) = f(y, a) \vee f(x, a) = * \vee f(y, a) = *\} \quad (1)$$

The $SIM(A)$ is a tolerance relation since it is reflexive and symmetric. Furthermore, let us denote by $S_A(x)$ the set of objects for which $SIM(A)$ holds, In other words, $S_A(x)$ is the maximal set of objects which are possibly indiscernible by A with x and any one element in $S_A(x)$ has a tolerance relation with x , it is called the tolerant class of x . It is clear that $S_A(x)$ is actually a kind of neighborhood of x .

Let $U/SIM(A)$ denotes classification for $\forall A \subseteq AT$, which is the family set $\{S_A(x) : x \in U\}$. What should be noticed is that all tolerant classes in $U/SIM(A)$ do not constitute a partition in general, but a covering on universe U , namely, $\bigcup_{x \in U} S_A(x) = U$ and $S_A(x) \neq \emptyset (x \in U)$.

Table 1. Incomplete Information System

Car	Price	Mileage	Size	Max-speed
1	High	High	Full	Low
2	Low	*	Full	Low
3	*	*	Compact	High
4	High	*	Full	High
5	*	*	Full	High
6	Low	High	Full	*

In Table 1, $AT = \{Price, Mileage, Size, Max-speed\}$, then we have $U/SIM(A) = \{S_{AT}(1), S_{AT}(2), S_{AT}(3), S_{AT}(4), S_{AT}(5), S_{AT}(6)\} = \{\{1\}, \{2, 6\}, \{3\}, \{4, 5\}, \{4, 5, 6\}, \{2, 5, 6\}\}$. Tolerant classes are the basis of defining lower and upper approximations of a set $X \subseteq U$. The A -lower approximation and the A -upper approximation of X are:

$$\underline{A}(X) = \{x \in U : S_A(x) \subseteq X\} \text{ and } \overline{A}(X) = \{x \in U : S_A(x) \cap X \neq \emptyset\} \quad (2)$$

Even though tolerance relation has been widely used in dealing with incomplete information system, it has the following drawbacks. In the first place, we can see that different two tolerant classes may have inclusion relation. For instance, in Table 1, $S_{AT}(2) \subseteq S_{AT}(6)$ and $S_{AT}(4) \subseteq S_{AT}(5)$ hold, this kind of situation sometimes is unreasonable when defining approximate sets. Furthermore, for all objects in $S_{AT}(x)$, they may have no common attribute values. For example, in Table 1, $S_{AT}(5) = \{4, 5, 6\}$, $f(4, Price) = \{high\}$ while $f(6, Price) = \{Low\}$. From this point of view, it is clear that objects 4 and 6 are discernible. From what have been discussed above, we should make a more reasonable classification in incomplete information system.

3. Rough Set Based on Optimal Compatible Class

3.1. Compatible Relation

Definition 3. Let S be an incomplete information system, each subset of attributes $A \subseteq AT$ determines a compatible relation $COM(A)$.

$$COM(A) = \{(x, y) : \forall a \in A, f(x, a) = f(y, a) \vee f(x, a) = * \vee f(y, a) = *\} \quad (3)$$

Definition 4. Let S be an incomplete information system and $A \subseteq AT$, then $U/COM(A)$ (represents classification) is defined as follows:

$$U/COM(A) = \{B \subseteq U : B \times B \subseteq COM(A), (\forall x \in U \wedge x \notin B \rightarrow (B \cup \{x\})^2 \not\subseteq COM(A))\} \quad (4)$$

In Definition 3, compatible relation is really same to the tolerance relation. However, $U/COM(A)$ in Definition 2 is quite different with $U/SIM(A)$ because any two elements in B are mutually compatible. Furthermore, if any other element in U is added into the compatible class, compatible relation in this compatible class will be destroyed anyway. This kind of compatible class meets with the actual needs more than tolerant class and as a result, it is called the **maximal complete compatible class**.

For example, Let us consider Table 1, $U/COM(AT) = \{\{1\}, \{2,6\}, \{3\}, \{4,5\}, \{5,6\}\}$. For any $B \in U/COM(AT)$, B is the maximal complete compatible class.

Property 1. If $COM(A)$ is a compatible relation, then:

$$COM(A) = \bigcap_{a \in A} COM(\{a\}) \quad (5)$$

Theorem 1. Let S be an incomplete information system and $A \subseteq C \subseteq AT$, for any $M \in U/COM(C)$, there must be $N \in U/COM(A)$ such that $M \subseteq N$.

Proof. By $A \subseteq C \subseteq AT$ and property 1, there must be $COM(C) = (COM(A) \cap (\bigcap_{c \in C-A} COM(\{c\})))$, then $COM(C) \subseteq COM(A)$. If $M \in U/COM(C)$, then we have $M^2 \subseteq COM(C) \subseteq COM(A)$. It means that any two elements in M are mutually compatible on set of attributes A . If $M \in U/COM(A)$, then the theorem is obvious. If $M \notin U/COM(A)$, then, according to the knowledge of discrete mathematics, there must be one class such that $M \subseteq N$ and $M \in U/COM(A)$.

As far as $U/SIM(A)$ is concerned, we can say about tolerant class for any $x \in U$, and for $U/COM(A)$ we can only say about non-exclusive coverage of U by all maximal complete compatible classes. However, the following theorem tells us that maximal complete compatible classes are so tightly related with tolerant classes.

Theorem 2. Let S be an in incomplete information system and $A \subseteq AT$, then for any $x \in U$, $\bigcup \{B : B \in U/COM, x \in B\} = S_A(x)$ holds.

Proof. $\forall y \in \bigcup \{B : B \in U/COM(A), x \in B\}$, we have $(x, y) \in COM(A) = SIM(A)$. By section 2, $S_A(x) = \{y \in U : (x, y) \in SIM(A)\}$, then there must be $y \in S_A(x)$. Since y is arbitrary, then we must have $\bigcup \{B : B \in U/COM(A), x \in B\} \subseteq S_A(x)$.

For any $y \in S_A(x)$, $(x, y) \in SIM(A) = COM(A)$ holds. If $\{x, y\}$ is the maximal complete compatible class B , then the theorem is true. If $\{x, y\}$ is not the maximal one, there must be B such that $\{x, y\} \subseteq B$ and $B \in U/COM(A)$, then $\forall y \in \bigcup \{B : B \in U/COM(A), x \in B\}$ holds. Since y is arbitrary, then $S_A(x) \subseteq \bigcup \{B : B \in U/COM(A), x \in B\}$.

From the above discussed, the theorem 2 is proved.

3.2. Optimal Compatible Classes

Let S be an incomplete information system, $U/COM(A) = \{B_1, B_2, \dots, B_n\}$ where $A \subseteq AT$, for any $x \in U$, we call B_i the maximal complete compatible class of x if and only if $1 \leq i \leq n$ and $x \in B_i$. Nevertheless, it is not difficult to find that there may be two or more maximal complete compatible classes for $\forall x \in U$.

For example, the object 5 in Table 1, $\{4,5\}$ and $\{5,6\}$ are all it's maximal complete compatible classes.

It is natural to consider how to choose a class (called **optimal compatible class**) from several maximal complete compatible classes of object x for computing reduction. In the following, two different methods are presented.

Method 1. The first method of choosing optimal compatible class roots from the basic idea of valued tolerance relation [13]. Simply, for $\forall x \in U$, we should choose a maximal complete compatible class in which elements are most possibly having same values of attributes as x has. Assuming that the set of possible values on each attribute is discrete, we make hypothesis that there exists a uniform probability distribution among such values. Consider $\forall a \in AT$ in incomplete information system S and associate it to the set $\{a_1, a_2, \dots, a_m\}$ of all possible values, given an object $x \in U$, if $f(x, a) = *$, then we assume that the probability $f(x, a) = a_j$ ($j=1, 2, \dots, m$) is equal to $1/\text{Card}(\{a_1, a_2, \dots, a_m\})$.

Definition 5. The probability of two objects x, y having same value on a set of attributes A is $pr_A(x, y) = \prod_{a \in A} pr_a(x, y)$ where $pr_a(x, y)$ is the probability of two objects having same value on single attribute a .

For any $x, y \in U$ and $a \in AT$, $pr_a(x, y)$ could be computed as follows:

$$Pr_a(x, y) = \begin{cases} 1/\text{card}(V_a) & : \neg(f(x, a) = * \wedge f(y, a) = *) \\ 1/(\text{card}(V_a))^2 & : f(x, a) = * \wedge f(y, a) = * \\ 1 & : f(x, a) = f(y, a) \\ 0 & : f(x, a) \neq f(y, a) \end{cases} \quad (6)$$

Definition 6. Let S be an incomplete information system in which $A \subseteq AT$, for any $x \in U$, the optimal compatible class of x is $S_A^{OPT}(x) = B_i$ if and only if the value of $\sum_{y \in B_i - \{x\}} Pr_A(x, y) / \text{card}(B_i - \{x\})$ is maximal for any $B_i \in U/COM(A) \wedge x \in B_i$.

Note 1. For $\forall x \in U$, if there are two or more compatible classes have the same maximal value of $\sum_{y \in B_i - \{x\}} Pr_A(x, y) / \text{Card}(B_i - \{x\})$, then their intersection is acceptable as the optimal compatible class of x .

For example, the object 5 in Table 1, there are two maximal complete compatible classes, $\{4, 5\}$ and $\{5, 6\}$. For maximal complete compatible class $\{4, 5\}$, $pr_{AT}(5, 4) = \frac{1}{2} \times \frac{1}{4} = \frac{1}{8}$, for maximal complete compatible class $\{5, 6\}$, $pr_{AT}(5, 6) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$. Therefore, $S_{AT}^{OPT}(5) = \{4, 5\} \cap \{5, 6\} = \{5\}$.

Method 2. In method 2, we only consider those values are all known. In all maximal complete compatible classes of $\forall x \in U$, there is at least one in which elements have the maximal numbers of attributes whose certain values are equal to the attributes' certain values of x .

In incomplete information system S , let $x, y \in U$ and $a \in A \subseteq AT$, suppose that $t=1$ if and only if $f(x, a) = f(y, a)$ while $f(x, a)$ and $f(y, a)$ are all known, otherwise $t=0$. Therefore, for any $x \in U$, $A \subseteq AT$, we define $\lambda_{B_i}(x) = \sum_{y \in B_i - \{x\}} \sum_{a \in A} t$ where $x \in B_i$.

Definition 7. In incomplete information system S , in which $A \subseteq AT$, for any $x \in U$, the optimal compatible class of x is $S_A^{OPT}(x) = B$ if and only if the value of $\lambda_{B_i}(x) / \text{Card}(B_i - \{x\})$ is maximal for any $B_i \in U/COM(A) \wedge x \in B_i$.

Note 2. For any $x \in U$, if there are two or more compatible classes have the same maximal value of $\lambda_{B_i}(x) / \text{Card}(B_i - \{x\})$, then their intersection is acceptable as optimal compatible class of x .

For example, in Table 1, there are two maximal compatible classes for object 5, $N_1 = \{4, 5\}$, $N_2 = \{5, 6\}$, respectively. According to definition 7, it is easy to work out $\lambda_{N_1}(5) = 2$ and $\lambda_{N_2}(5) = 1$, so $\text{Max}(\lambda_{N_1}(5), \lambda_{N_2}(5)) = 2$, $S_{AT}^{OPT}(5) = \{4, 5\}$.

Definition 8. Given an incomplete information system S and a non-empty subset of attributes $A \subseteq AT$, with each subset of objects $X \subseteq U$ we associate two sets:

$$\underline{A}^{OPT}(X) = \{x \in U : S_A^{OPT}(x) \subseteq X\} \text{ and } \overline{A}^{OPT}(X) = \{x \in U : S_A^{OPT}(x) \cap X \neq \emptyset\} \quad (7)$$

Note 3. According to the compatible relation of rough set theory, A-lower and A-upper of X have two instances in respect that two different methods of choosing optimal compatible class.

3.3. Comparison between Tolerance and Compatible Relations

From the viewpoint of granular computing [14], classes are the basic building blocks and called elementary granules [15]. They are the smallest nonempty subsets that can be defined, observed or measured. Of course, different binary relations may produce different elementary granules. From what have been discussed above, classes produced by tolerance and compatible relations respectively, all form coverings on the universe.

Each covering represents one granulated view of the universe. Due to the difference of methods in classifying, tolerance and compatible relations produce different coverings called π_1 and π_2 , respectively. According to Theorem 2, it is clear that covering π_2 is a refinement of covering π_1 or equivalently π_1 is a coarsening of π_2 , denoted by $\pi_1 \leq \pi_2$ or $\pi_2 \geq \pi_1$, for the reason that every block of π_2 is contained in some block of π_1 . Given two coverings π_1 and π_2 , their meet $\pi_1 \wedge \pi_2$ is the largest covering which is a refinement of both π_1 and π_2 , and their join $\pi_1 \vee \pi_2$ is the smallest covering which is a coarsening of both π_1 and π_2 . From what have been discussed, covering π_2 has a smaller level of granulation for problem solving than covering π_1 .

Let U be a non-empty finite set of objects, and let $R \subseteq U \times U$ denote a binary relation on U . The pair $\text{apr} = \langle U, R \rangle$ is called an approximation space [1]. The covering of the universe is called the quotient set induced by R and is denoted by U/R . Even though rough set data analysis is a symbolic method of analysis, it uses counting information provided by the classes of the binary relations under consideration. The inherent statistic of an approximation space $\langle U, R \rangle$ is the accuracy measure of rough set approximation [1] $\alpha(A) = \text{Card}(\underline{A}(X)) / \text{Card}(\overline{A}(X))$. It may be interpreted as the probability that an element belongs to the lower approximation, given that the element belongs to the upper approximation and expresses the degree of completeness of our knowledge of X .

Tolerance and compatible relations produce different accuracy measures of rough set approximation named as $\alpha(A)_1$ and $\alpha(A)_2$, respectively. According to theorem 2, it is easy to validate that $\underline{A}(X) \subseteq \underline{A}^{\text{OPT}}(X) \subseteq X \subseteq \overline{A}^{\text{OPT}}(X) \subseteq \overline{A}(X)$, therefore, $\text{Card}(\underline{A}(X)) \leq \text{Card}(\underline{A}^{\text{OPT}}(X))$ and $\text{Card}(\overline{A}(X)) \geq \text{Card}(\overline{A}^{\text{OPT}}(X))$, so we have $0 \leq \alpha(A)_1 \leq \alpha(A)_2 \leq 1$. That is to say, with compatible relation, we can get more knowledge of X than tolerance relation.

4. Knowledge Reduction

An incomplete decision table [4] is an incomplete information system $DT = \langle U, AT \cup \{d\}, V, f \rangle$ where d is called decision attribute and V_d is the value domain of the decision attribute d , correspondingly, elements in AT are called condition attributes. In addition, $AT \cap \{d\} = \emptyset$ and $V_d \neq \emptyset$.

Any decision table may be regarded as a set of decision rules of the form: $\wedge(a, v) \rightarrow (d, w)$, where $a \in AT$, $v \in V_c$, $w \in V_d$. Owing to difference of classification between compatible and tolerance relations, the computation of generalized decision function [12] that is used in knowledge reduction should be modified.

Definition 9. Let $DT = \langle U, AT \cup \{d\}, V, f \rangle$ be an incomplete decision system and then the generalized decision function is defined as follows:

$$\eta_A : U \rightarrow P(V_d), \forall A \subseteq AT, \eta_A = \{i : i = d(y), y \in S_A^{\text{OPT}}(x)\} \quad (8)$$

Definition 10. Let $DT = \langle U, AT \cup \{d\}, V, f \rangle$ be an incomplete decision system, $A \subseteq AT$ is a reduction of DT (relative reduction) iff $\eta_A = \eta_{AT}$ and for any $C \subseteq A$, $\eta_C = \eta_A$.

In the process of investigation, it is convenient to use discernibility function [12] to compute reduction in incomplete information and decision systems.

In incomplete decision system, for any $A \subseteq AT$, let $\alpha_A(x, y)$ be a set of attributes $a \in A$ such that $(x, y) \notin \text{COM}(\{a\})$ and as a result if $(x, y) \in \text{COM}(\{a\})$ then $\alpha_A(x, y) = \emptyset$. Let $\sum \alpha_A(x, y)$ be a

Boolean expression that is equal to 1 if $\alpha_A(x, y) = \emptyset$, otherwise, $\sum \alpha_A(x, y)$ be a disjunction of variable corresponding to attributes contained in $\alpha_A(x, y)$.

Definition 11. Δ^* is a discernibility function for incomplete decision system if:

$$\Delta^* = \prod_{(x,y) \in U \times \{z \in U: d(z) \in \eta_A(x)\}} \sum \alpha_A(x, y) \tag{9}$$

Definition 12. $\Delta^*(x)$ is a discernibility function for object x in incomplete decision system if:

$$\Delta^*(x) = \prod_{(y) \in \{z \in U: d(z) \in \eta_A(x)\}} \sum \alpha_A(x, y) \tag{10}$$

5. Illustrative Example

In this section, a medical treatment decision table will be analyzed by rough set based on compatible relation. Of course, two methods of choosing optimal compatible classes are all worked out.

Table 2 depicts an incomplete decision table about medical treatment decision. *Age, Sarcous pain, Fevered, and Headache* are the conditional attributes, *Remedial scheme* is the decision attribute (in the sequel, $a_1, a_2, a_3, a_4,$ and d will stand for *Age, Sarcous pain, Fevered, Headach* and *Remedial schem*, respectively). $V_{a_1} = \{adult, enfant, infant\} = \{1,2,3\}$, $V_{a_2} = \{no\ pain, pain\} = \{1,2\}$, $V_{a_3} = \{normal, hyperpyrexia\} = \{1,2\}$, $V_{a_4} = \{no\ headache, headache\} = \{1,2\}$, and $V_d = \{medical\ therapy, physical\ therapy, combination\ of\ medication\ and\ physics\} = \{1,2,3\}$.

Table 2. Medical Decision Table

U	a_1	a_2	a_3	a_4	d
1	1	*	*	2	1
2	1	2	*	2	1
3	1	*	2	2	1
4	1	1	1	2	2
5	1	*	1	*	2
6	2	1	1	2	2
7	*	2	1	*	2
8	*	2	2	1	3
9	2	1	2	1	3
10	3	*	*	1	3

The following are ten decision rules for Table 2:

- $r_1: (a_1, 1) \wedge (a_2, *) \wedge (a_3, *) \wedge (a_4, 2) \rightarrow (d, 1)$
- $r_2: (a_1, 1) \wedge (a_2, 2) \wedge (a_3, *) \wedge (a_4, 2) \rightarrow (d, 1)$
- $r_3: (a_1, 1) \wedge (a_2, *) \wedge (a_3, 2) \wedge (a_4, 2) \rightarrow (d, 1)$
- $r_4: (a_1, 1) \wedge (a_2, 1) \wedge (a_3, 1) \wedge (a_4, 2) \rightarrow (d, 2)$
- $r_5: (a_1, 1) \wedge (a_2, 1) \wedge (a_3, 1) \wedge (a_4, *) \rightarrow (d, 2)$
- $r_6: (a_1, 2) \wedge (a_2, *) \wedge (a_3, 1) \wedge (a_4, 2) \rightarrow (d, 2)$
- $r_7: (a_1, *) \wedge (a_2, 1) \wedge (a_3, 1) \wedge (a_4, *) \rightarrow (d, 2)$
- $r_8: (a_1, *) \wedge (a_2, 2) \wedge (a_3, 2) \wedge (a_4, 1) \rightarrow (d, 3)$
- $r_9: (a_1, 2) \wedge (a_2, 1) \wedge (a_3, 2) \wedge (a_4, 1) \rightarrow (d, 3)$
- $r_{10}: (a_1, 3) \wedge (a_2, *) \wedge (a_3, *) \wedge (a_4, 1) \rightarrow (d, 3)$

From Table 2. we have $U/COM(AT) = \{\{1,2,3\}, \{1,2,5,7\}, \{1,4,5\}, \{6\}, \{7,10\}, \{8,10\}, \{9\}\}$.

According to Method 1, we have:

$$S_{AT}^{OPT}(1) = \{1,4,5\}; S_{AT}^{OPT}(2) = \{1,2,3\}; S_{AT}^{OPT}(3) = \{1,2,3\}; S_{AT}^{OPT}(4) = \{1,4,5\}; S_{AT}^{OPT}(5) = \{1,4,5\};$$

$$S_{AT}^{OPT}(6) = \{6\}; S_{AT}^{OPT}(7) = \{1,2,5,7\}; S_{AT}^{OPT}(8) = \{8,10\}; S_{AT}^{OPT}(9) = \{9\}; S_{AT}^{OPT}(10) = \{8,10\}$$

According to Method 2, we have:

$$S_{AT}^{OPT}(1) = \{1, 2, 3\}; S_{AT}^{OPT}(2) = \{1, 2, 3\}; S_{AT}^{OPT}(3) = \{1, 2, 3\}; S_{AT}^{OPT}(4) = \{1, 4, 5\}; S_{AT}^{OPT}(5) = \{1, 4, 5\};$$

$$S_{AT}^{OPT}(6) = \{6\}; S_{AT}^{OPT}(7) = \{1, 2, 5, 7\}; S_{AT}^{OPT}(8) = \{8, 10\}; S_{AT}^{OPT}(9) = \{9\}; S_{AT}^{OPT}(10) = \{8, 10\}$$

Owing to we have two methods of choosing optimal compatible classes, then two different generalized decision functions (η_{AT1} , η_{AT2} , respectively) are worked out in Table 3.

Table 3. Generalized Decision Functions

UAT	1	2	3	4	5	6	7	8	9	10
η_{AT1}	1,2	1	1	1,2	1,2	2	1,2	3	3	3
η_{AT2}	1	1	1	2	1,2	2	1,2	3	3	3

Owing to different generalized decision functions, the computation of discernibility function will be different, too. Formally, using η_{AT1} we can compute out all of the relative reductions of Table 2:

$$\Delta^*(1) = a_4; \Delta^*(2) = a_2a_4; \Delta^*(3) = a_3a_4; \Delta^*(4) = a_1a_2 \vee a_1a_3; \Delta^*(5) = a_1a_3; \Delta^*(6) = a_1a_3 \vee a_1a_4;$$

$$\Delta^*(7) = a_3; \Delta^*(8) = a_3a_4; \Delta^*(9) = a_1a_3 \vee a_1a_2a_4; \Delta^*(10) = a_1; \Delta^* = a_1a_2a_3a_4.$$

From relative reductions of objects, we can get following compact rules:

$$r_1: (a_4, 2) \rightarrow (d, 1) \vee (d, 2)$$

$$r_2: (a_1, 1) \wedge (a_2, 2) \rightarrow (d, 2)$$

$$r_3: (a_3, 2) \rightarrow (d, 2)$$

$$r_4: (a_1, 2) \wedge (a_4, 2) \rightarrow (d, 2)$$

$$r_5: (a_3, 2) \wedge (a_4, 1) \rightarrow (d, 3)$$

$$r_6: (a_1, 2) \wedge (a_3, 2) \rightarrow (d, 3)$$

$$r_7: (a_1, 2) \wedge (a_2, 1) \wedge (a_4, 1) \rightarrow (d, 32)$$

$$r_8: (a_3, 2) \rightarrow (d, 3)$$

Using η_{AT2} we can also compute out all of the reductions in Table 2:

$$\Delta^*(1) = a_1a_2; \Delta^*(2) = a_2a_4; \Delta^*(3) = a_3a_4; \Delta^*(4) = a_1a_2 \vee a_1a_3 \vee a_4; \Delta^*(5) = a_1a_3; \Delta^*(6) = a_1a_3 \vee a_1a_4;$$

$$\Delta^*(7) = a_3; \Delta^*(8) = a_3a_4; \Delta^*(9) = a_1a_3 \vee a_1a_2a_4; \Delta^*(10) = a_1; \Delta^* = a_1a_2a_3a_4.$$

Similarly to rules inducing by method 1 of choosing optimal compatible classes, it is not hard to bring out the rules by method 2 and the outcomes are same to r_1', \dots, r_8' .

6. Conclusion

Rough set theory assumes that knowledge comes from the ability of classification. However, an explicit hypothesis in rough set is that all available objects can be completely described by the set of attributes. In order to manage objects who have incomplete descriptions of attributes, so many scholars have done excellent jobs. In this paper, the compatible relation and maximal complete compatible classes are presented. The main advantage of compatible class is that it can make sure that all elements in the same class are mutually compatible while tolerant class cannot. From the knowledge reduction based on compatible relation, some compact rules are mined. In the further researches, we are going to define some precise measurements to weigh the reliability of those rules mined from incomplete information systems.

Acknowledgements

This work is supported by the Natural Science Foundation of China (No.61100116), Natural Science Foundation of Jiangsu Province of China (No.BK2011492) and (No.BK20130472).

References

- [1] Pawlak Z. Rough sets. *International Journal of Computer and Information Sciences*. 1982; 11(5): 341-356.
- [2] Pawlak Z, Skowron A. Rough sets: some extensions, *Information Sciences*. 2007; 177(1): 28-40.
- [3] Qian YH, Liang JY, Pedrycz W, Dang CY. Positive approximation: an accelerator for attribute reduction in rough set theory. *Artificial Intelligence*. 2010; 174(9-10): 597-618.
- [4] Yang XB, Xie J, Song, XN, et al. Credible rules in incomplete decision system based on descriptors. *Knowledge-Based-Systems*. 2009; 22(1): 8-17.
- [5] Xiong Yushu. A Clustering Algorithm Based on Rough Set and Genetic Algorithm. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(10): 5782-5788.
- [6] Zhang Y. Rough Set and DEA of Strategic Alliance Stable Decision-making Model. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(12): 7295-7301.
- [7] Qian YH, Liang JY, Pedrycz W, Dang CY. An efficient accelerator for attribute reduction from incomplete data in rough set framework. *Pattern Recognition*. 2011; 44(8): 1658-1670.
- [8] Xu WH, Li Y, Liao XW. Approaches to attribute reductions based on rough set and matrix computation in inconsistent ordered information systems. *Knowledge-Based Systems*. 2012; 27: 78-91.
- [9] Liang JY, Wang F, Dang CY, Qian YH, An efficient rough feature selection algorithm with a multi-granulation view, *International Journal of Approximate Reasoning*. 2012; 53(6): 912-926.
- [10] Grzymala-Busse JW. Data with missing attribute values: generalization of indiscernibility relation and rule reduction. *Lecture Notes in Computer Science*. 2004; 3100: 78-95.
- [11] Grzymala-Busse JW, Wang AY. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values. *Proceeding of the 5th International Workshop on Rough Sets and Soft Computing at the 3rd Joint Conference on Information Sciences*, Research Triangle Park, NC. 1997: 69-72.
- [12] Kryszkiewicz M. Rules in incomplete information systems. *International Journal of Information Sciences*. 1999; 113(3-4): 271-292.
- [13] Stefanowski J. Incomplete information tables and rough classification. *International Journal of Computational Intelligence*. 2001; 17(3): 545-566.
- [14] Yao YY. Information granulation and rough approximation. *International Journal of Intelligent Systems*, 2001; 16(1): 87-104.
- [15] Xu WH, Sun WX, Zhang XY, Zhang WX. Multiple granulation rough set approach to ordered information systems. *International Journal of General Systems*. 2012; 41(5): 475-501.