ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 3, No. 3, 2008, *pp*. 189-198

Approach to approximate distribution reduct in incomplete ordered decision system

YunSong Qi [1 2 +], Huaijiang Sun [1], XiBei Yang [1],Yuqing Song [1], Quansen Sun [1]

[1] School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing, Jiangsu Province, China, 210094

[2] School of Electronics and Information, Jiangsu University of Science and Technology Zhenjiang, Jiangsu Province, China, 212003

(*Received March 24, 2008, accepted July 30 2008*)

Abstract. The original rough set model cannot be used to deal with the incomplete information systems. Nevertheless, by relaxing the indiscernibility relation to more general binary relations, many improved rough set models have been successfully applied into the incomplete information systems for knowledge acquisition. This article presents an explorative research focusing on the transition from incomplete decision system to a more complex system---the incomplete ordered decision system. In such incomplete decision system, all attributes have preference-ordered domains. With introduction of the concept of approximate distribution reduct into the incomplete ordered decision system, four new notions of approximate distribution reduct are proposed. They are the minimal subsets of condition attributes, which preserve lower and upper approximations of all upward unions and downward unions of the decision classes respectively. The judgment theorems and discernibility matrices associated with these approximate distribution reducts are also obtained. For further illustration, an example is analyzed. The research is meaningful both in the theory and in applications for the acquisition of rules in complex information systems.

Keywords: incomplete information system, incomplete ordered decision system, approximate distribution reduct, dominance-based rough set.

1. Introduction

It is well known that Pawlak's rough set[1-4] generalizes the classical set theory by allowing an alternative to formulate set with imprecise boundaries. In recent years, the rough set theory has been widely used in various fields, such as machine learning[5], data mining[6], pattern recognition[7] and knowledge discovery, etc. An important concept related to rough set is the information system (attribute-value system). It is noticeable that the traditional rough set model can only be used in the analysis of data presented in terms of the complete information, i.e. every object in the universe has a real and certain value for every attribute. However, the incomplete information systems[5,8-17] other than the complete information systems can be seen everywhere in real-world applications. By an incomplete information system we mean a system with unknown attribute value[13]. In this paper, we assume that the unknown values are regard as lost[8]. Incomplete information system in which all unknown values are lost, from the viewpoint of rough set theory, was studied for the first time in Ref. 9. Subsequently, Kryszkiewicz[11] formed her expanded rough set based on the tolerance relation (reflexive, symmetric). The tolerance relation is corresponding to the idea that unknown value is just "missed", but it does exist. It is our imperfect knowledge that obliges us to deal with a partial information table. Each object potentially has a complete description, but we just miss some information for the moment[16]. Therefore, the unknown value is considered as to be comparable to any one of the values in the domain of the corresponding attribute.

Nevertheless, the tolerance relation is failing at the point where attributes with preference-ordered domains (criteria)18-20. Attributes with preference-ordered domains are commonly seen in the Multi-Criteria Decision Making (MCDM)18,19 problems, like sorting, choice or ranking. In this paper, we call the

+ Corresponding author. Tel.: +86-511-84401153.
E-mail address: qys@ujs.edu.cn

incomplete information system in which all attributes are criteria the Incomplete Ordered Information System (IOIS). In Ref. 14, Shao has made an investigation of the Dominance-based Rough Set Approach (DRSA) to the IOIS. He did not only present the new definition of dominance relation in IOIS, but also study the approaches to knowledge reductions in IOIS and consistent Incomplete Ordered Decision System (IODS). However, Shao's research is not so all-around because most of the decision systems are inconsistent other than consistent for various factors21 such as noise in data, compact representation, prediction capability, etc. Therefore, new approach to knowledge reduction in the inconsistent IODS has become a necessity.

At the present time, without consideration of the preference-ordered domains of attributes, many authors have studied different forms of knowledge reduction in complete inconsistent system. For instance, Kryszkiewicz investigated and compared five notions of knowledge reduction in Ref. 22. Wang studied the relationship of the definitions of rough reduction in algebra view and information view23. Ślęzak proposed the concept of approximate entropy reduct in Ref. 24. Zhang proposed the (maximum) distribution reducts21. Mi defined the approximate distribution reduct based on variable precision rough set model25. In this paper, the approximate distribution reduct will be introduced into our IODS because it has significant advantage: an approximate distribution reduct of a decision system is a minimal subset of condition attributes which preserves the lower or upper approximation of all decision classes. Therefore, it can keep decision rules derived from the approximate distribution consistent set are compatible with the ones derived from original system25.

What should be noticed is that the knowledge approximated in the IODS are collections of upward and downward unions of decision classes and the granules of knowledge are sets of objects defined using a dominance relation. Therefore, it is required to provide new definitions of approximate distribution reduct in the IODS. In this paper, we will present four new notions of approximate distribution reduct. These approximate distribution reducts are the minimal subsets of attributes, which preserve the lower and upper approximation of all upward and downward unions of classes respectively. The corresponding judgment theorems and discernibility matrices associated with these reducts are also obtained. So we provide practical approaches to knowledge reductions in the IODS.

2. Incomplete Decision System

A complete decision system is a 4-tuple $\Omega=<U, AT\cup D, V, f>$, where U is a non-empty finite set of objects called universe and AT is a non-empty finite set of condition attributes, such that $\forall a\in AT : U\rightarrow V_a$ where V_a is the domain of single attribute a, D is a non-empty finite set of decision attributes where $AT\cap D = \varnothing$, V is regarded as the domain of all attributes and then $V= V_{AT}\cup V_D$. Moreover, for $\forall x\in U$, let us denote f (x, a) by the value of x holds on a ($a\in AT\cup D$).

A decision system is called an incomplete one iff $\exists\ x\in U$ and $\exists a\in AT$ such that f (x, a) = *, the special value"*" is used to indicate that the value of a condition attribute is unknown for the object. It is assumed here that the unknown value is just "missed", but it does exist. Thus, $V= V_{AT}\cup V_D\cup\{*\}$. The incomplete decision system is still denoted without confusion by $\Omega=<U, AT\cup D, V, f>$.

Definition 1 [11]. Let Ω be an incomplete decision system, then for $\forall A \subseteq AT$, the tolerance relation is defined as follows:

$T(A)= \{(x, y) \in U^2 : \forall a \in A, f(x, a) = f(y, a) \vee f(x, a) = * \vee f(y, a) = *\}$.

It is clear that the tolerance relation $T(A)$ is only reflexive and symmetric, but not necessarily transitive. Furthermore, for $\forall x\in U$, let us denote $T_A(x)$ by the set of objects for which $T(A)$ holds, i.e. $T_A(x)=\{y\in U : (x, y)\in T(A)\}$. In other words, $T_A(x)$ is the maximal set of objects which are possibly indiscernible by A with x, it is called the tolerant class of x.

Derivation of decision rules from decision systems is to examine whether the objects possessing the same condition attribute values will posses the same decision attribute values. So the criterion of classification is to classify the objects possessing the same attribute values to the same class. Kryszkiewicz made use of $T_A(x)$ to classify the universe of discourse in the incomplete information systems, but such a classification approach has the following two drawbacks[10]:

(1) objects in $T_A(x)$ may have no one common attribute value, because objects in $T_A(x)$ are all tolerant with x but may not tolerant with each other;

(2) for $x \neq y$, there may exist inclusion relation between $T_A(x)$ and $T_A(y)$.

To solve the above two problems, Leung made used of the maximal tolerance classification approach to classify objects in the incomplete information system. The maximal tolerance class, which is called maximal consistent block[13], is a well-known concept in discrete mathematics. For more details about knowledge reduction based on maximal consistent block, please quote Ref. 10, 13.

3. Incomplete Ordered Decision System

3.1. Ordered Decision System

Dominance-based Rough Sets Approach (DRSA) was firstly proposed by Greco to take into account the ordering properties of criteria. This innovation is mainly based on substitution of the indiscernibility relation by a dominance relation[20] and as a result, it can be used in multi-criteria decision analysis. A decision system is called an Ordered Decision System (ODS) iff condition and decision attributes are all criteria, it is denoted by Ω^O.

In an ordered decision system Ω^O, let $\succeq_a$ be a weak preference relation on U (often called outranking) representing a preference on the set of objects with respect to criterion a ; $x \succeq_a y$ means "x is at least as good as y with respect to criterion a". We say that x dominates y with respect to $A \subseteq AT$, (or, x A-dominates y), denoted by $x D_A y$, if $x \succeq_a y$ for all $a \in A$. Assuming, without loss of generality, that domains of all criteria are ordered such that preference increases with the value, $x D_A y$ is equivalent to: $f(x, a) \geq f(y, a)$. Therefore, the "granules of knowledge" used in DRSA are[20]:

- A set of objects dominating x, called A-dominating set, $D_A^+(x)=\{y \in U : y D_A x\}$.
- A set of objects dominated by x, called A-dominated set, $D_A^-(x)=\{y \in U : x D_A y\}$.

Moreover, we also assume here that the set of decision attributes D makes a partition of U into a finite number of classes. Let CL={CLt, t∈T }, T={1, 2, …, n}, be a set of these classes that are ordered, that is, for ∀ r, s∈T such that r > s, the objects from CLr are preferred to the objects from CLs. The set to be approximated is an upward union and a downward union of classes, which are defined respectively as CL≥t =∪s≥t CLs , CL≤t = ∪s≤t CLs, t =1, …, n. The statement x∈CL≥t means "x belongs to at least class CL≥t", where x∈CL≤t means "x belongs to at most class CL≥t" 20.

Definition 2 [20]. Let Ω^O be an ODS in which $A \subseteq AT$, for $\forall CL^{\geq}_t$ $(1 \leq t \leq n)$, the lower and upper approximate sets are defined as:

$$\underline{A}(CL^{\geq}_t)=\{x \in U : D_A^+(x) \subseteq CL^{\geq}_t\}, \quad \overline{A}(CL^{\geq}_t)=\{x \in U : D_A^-(x) \cap CL^{\geq}_t \neq \varnothing\},$$

for $\forall$ $CL^{\leq}_t$ $(1 \leq t \leq n)$, the lower and upper approximate sets are defined as:

$$\underline{A}(CL^{\leq}_t)=\{x \in U : D_A^-(x) \subseteq CL^{\leq}_t\}, \quad \overline{A}(CL^{\leq}_t)=\{x \in U : D_A^+(x) \cap CL^{\leq}_t \neq \varnothing\}.$$

3.2. DRSA in Incomplete Ordered Decision System

Similar to Section 2, an ordered decision system is called an incomplete one iff $\exists$ $x \in U$ and $\exists a \in AT$ such that $f(x, a) = *$. The Incomplete Ordered Decision System (IODS) is still denoted without confusion by Ω^O. By considering the unknown values "*", the definition of dominance relation should be improved.

Definition 3 [14]. Let Ω^O be an IODS in which $A \subseteq AT$, then the dominance relation in terms of A is defined as follows:

$$R^{\geq}(A)= \{(x, y) \in U^2 : \forall a \in A, f(x, a) \geq f(y, a) \vee f(x, a) = * \vee f(y, a) = *\}$$

It is clear that dominance relation $R^{\geq}(A)$ is only reflexive, but not necessarily symmetric and transitive.

Therefore, let us denote by

- $[x]^{\geq}_A=\{y \in U : (y, x) \in R^{\geq}(A)\}$ is the set of objects that may dominating x in terms of A,
- $[x]^{\leq}_A=\{y \in U : (x, y) \in R^{\geq}(A)\}$ is the set of objects that may be dominated by x in terms of A.

Theorem 1. Let ΩO be an IODS in which A ⊆ AT, then we have R≥(A)=∩a∈A R≥(a), R≥(AT) ⊆ R≥(A).

Definition 4.Let Ω^O be an IODS in which $A \subseteq AT$, *for* $\forall CL^{\geq}_t$ $(1 \leq t \leq n)$, the lower and upper approximate sets are defined as:

$$A_{\sim}(CL^{\geq}_t)=\{x \in U : [x]^{\geq}_A \subseteq CL^{\geq}_t\}, \quad A^{\sim}(CL^{\geq}_t)=\{x \in U : [x]^{\leq}_A \cap CL^{\geq}_t \neq \varnothing\},$$

for $\forall\ CL^{\leq}_t (1\leq t \leq n)$, the lower and upper approximate sets are defined as:

$$A_{\sim}(CL^{\leq}_t)=\{x\in U : [x]^{\leq}_A \subseteq CL^{\leq}_t\},\quad A^{\sim}(CL^{\leq}_t)=\{x\in U : [x]^{\geq}_A \cap CL^{\leq}_t \neq\varnothing\}.$$

Theorem 2. Let Ω^O be an IODS in which $A \subseteq AT$, then we have

$$A^{\sim}(CL^{\geq}_t)= \bigcup_{x\in CL^{\geq}_t}[x]^{\geq}_A,\quad A^{\sim}(CL^{\leq}_t)= \bigcup_{x\in CL^{\leq}_t}[x]^{\leq}_A.$$

Proof: By Def. 4, for ∀x∈A~(CL≥t), we have [x]≤A∩CL≥t ≠∅, it follows that there must be y∈U such that y∈[x]≤A and y∈CL≥t , that is, x∈[y]≥A , x∈ $\bigcup_{y\in CL^{\geq}_t}[y]^{\geq}_A$. Conversely, for ∀y∈ $\bigcup_{x\in CL^{\geq}_t}[x]^{\geq}_A$, we have y∈[x]≥A and x∈CL≥t , then x∈[y]≤A , that is, [y]≤A∩CL≥t ≠∅, i.e. y∈A~(CL≥t). From discussion above, A~(CL≥t)== $\bigcup_{x\in CL^{\geq}_t}[x]^{\geq}_A$. Similarity, it is not difficult to prove that A~(CL≤t)= $\bigcup_{x\in CL^{\leq}_t}[x]^{\leq}_A$.

Theorem 3. Let Ω^O be an IODS in which $A \subseteq AT$, then we have the following:

$$A_{\sim}(CL^{\geq}_t) =U-A^{\sim}(CL^{\leq}_{t-1}),\ t=2, \ldots n,\quad A_{\sim}(CL^{\leq}_t) = U-A^{\sim}(CL^{\geq}_{t+1}),\ t=1, \ldots n-1,$$

$$A^{\sim}(CL^{\geq}_t) =U-A_{\sim}(CL^{\leq}_{t-1}),\ t=2, \ldots n,\quad A^{\sim}(CL^{\leq}_t) = U-A_{\sim}(CL^{\geq}_{t+1}),\ t=1, \ldots n-1.$$

The dominance-based rough approximations of upward and downward unions of classes can serve to induce a generalized description of objects contained in the data table in terms of if--then decision rules. Intuitively, this is linked to the interpretation of the rough approximations in terms of implications[20]:

- $A_{\sim}(CL^{\geq}_t)=\{x\in U : (y, x)\in R^{\geq}(A)\Rightarrow y\in CL^{\geq}_t\}$, for t =2,…, n .
- $A^{\sim}(CL^{\geq}_t)=\{x\in U :$ if $(y, x)\in R^{\geq}(A)$, then y could belong $CL^{\geq}_t\}$, for t =2,…, n .
- $A_{\sim}(CL^{\leq}_t)=\{x\in U : (x, y)\in R^{\geq}(A)\Rightarrow y\in CL^{\leq}_t\}$, for t =1,…, n-1.
- $A^{\sim}(CL^{\leq}_t)=\{x\in U :$ if $(x, y)\in R^{\geq}(A)$, then y could belong $CL^{\leq}_t\}$, for t =1,…, n-1.

Definition 5. Let ΩO be an IODS, then ΩO is called a consistent one iff R≥(AT)⊆ R≥(D) where D is the set of decision attributes and R≥(D)= {(x, y) ∈U 2 : ∀d ∈D, f (x, d) ≥ f (y, d) }.

Definition 6. Let ΩO be a consistent IODS in which A ⊆ AT, then we say that A is a reduct of ΩO if the following two conditions hold:

- $R^{\geq}(A)\subseteq R^{\geq}(D)$,
- $R^{\geq}(B)\not\subseteq R^{\geq}(D)$ for $\forall B\subset A$.

In Def. 6, the reduct is the minimal subset of the condition attributes that preserves the consistent characteristic of the incomplete ordered decision system. Therefore, this kind of knowledge reduction can be only used in the analysis of consistent IODS.

As an example, Tab.1 show us an inconsistent incomplete ordered decision system. The director of a school must give a global evaluation to some students. This evaluation should be based on the level in Mathematics, Physics, History and Literature. However, some of evaluations are missing for some students. As we notice that (x7, x14)∈R≥(AT) while (x7, x14)∉R≥(e), this situation is the same to the pair of (x4, x9), in other words, R≥(AT)⊄R≥(D) holds in Tab.1. The approach to knowledge reduction in Def. 6 cannot be used in it for acquiring reduct. It is required to provide new definitions of knowledge reduction in the IODS. This is what will be discussed in the following.

Table.1. Student evaluations with missing values

U	Mathematics (a)	Physics (b)	History (c)	Literature (d)	Global evaluation (e)
x_1	Medium	Medium	Bad	Bad	Bad
x_2	Bad	Good	Bad	Bad	Bad
x_3	Bad	Bad	Medium	Bad	Bad
x_4	Medium	Medium	Good	*	Bad
x_5	Good	Medium	Bad	Medium	Medium
x_6	Medium	Good	Bad	Medium	Medium
x_7	Good	Medium	Medium	Good	Medium
x_8	Good	*	Medium	Bad	Medium
x_9	Medium	Medium	Medium	Good	Medium
x_{10}	Good	Bad	Good	Bad	Medium
x_{11}	Medium	*	Good	Bad	Medium
x_{12}	Good	Bad	Medium	Good	Good
x_{13}	Medium	Good	Medium	Good	Good
x_{14}	Good	Medium	*	Good	Good
x_{15}	Medium	Good	Good	Good	Good

4. Approximate distribution Reduct

In this section, we will make a further investigation of the approach to knowledge reduction in the IODS. In classical complete decision system, the concept of approximate distribution reduct[25] was firstly proposed by Mi. Approximate distribution reduct is based on the fundamental concept of Pawlak's rough set theory, i.e. lower and upper approximations. It is the minimal subset of condition attributes, which preserves the lower or upper approximate sets of all decision class. An important characteristic of approximate distribution reduct is that it can keep decision rules derived from the approximate distribution consistent set are compatible with the ones derived from original system.

In our IODS, attribute values are missing in some cases while the set that will be approximated is the upward or downward union of the decision classes. Therefore, how to expand the concept of approximate distribution reduct in the IODS is what will be discussed in the following.

Definition 7. Let Ω^O be an IODS in which $A \subseteq AT$, let us denote by

$$L^{\geq}_A = \{A_{\sim}(CL^{\geq}_1), A_{\sim}(CL^{\geq}_2),\ldots, A_{\sim}(CL^{\geq}_n)\},\ L^{\leq}_A=\{A_{\sim}(CL^{\leq}_1), A_{\sim}(CL^{\leq}_2), \ldots, A_{\sim}(CL^{\leq}_n)\},$$
$$H^{\geq}_A=\{A^{\sim}(CL^{\geq}_1), A^{\sim}(CL^{\geq}_2),\ldots, A^{\sim}(CL^{\geq}_n)\},\ H^{\leq}_A=\{A^{\sim}(CL^{\leq}_1), A^{\sim}(CL^{\leq}_2), \ldots, A^{\sim}(CL^{\leq}_n)\},$$

then

1. If $L^{\geq}_A = L^{\geq}_{AT}$, then A is referred to as the ≥-lower approximate distribution consistent set. If $L^{\geq}_A = L^{\geq}_{AT}$ and $L^{\geq}_B \neq L^{\geq}_A$ for $\forall B \subset A$, then A is referred to as a ≥-lower approximate distribution reduct of Ω^O.
2. If $L^{\leq}_A = L^{\leq}_{AT}$, then A is referred to as the ≤-lower approximate distribution consistent set. If $L^{\leq}_A = L^{\leq}_{AT}$ and $L^{\leq}_B \neq L^{\leq}_A$ for $\forall B \subset A$, then A is referred to as a ≤-lower approximate distribution reduct of Ω^O.
3. If $H^{\geq}_A = H^{\geq}_{AT}$, then A is referred to as the ≥-upper approximate distribution consistent set. If $H^{\geq}_A = H^{\geq}_{AT}$ and $H^{\geq}_B \neq H^{\geq}_A$ for $\forall B \subset A$, then A is referred to as a ≥-upper approximate distribution reduct of Ω^O.
4. If $H^{\leq}_A = H^{\leq}_{AT}$, then A is referred to as the ≤-upper approximate distribution consistent set. If $H^{\leq}_A = H^{\leq}_{AT}$ and $H^{\leq}_B \neq H^{\leq}_A$ for $\forall B \subset A$, then A is referred to as a ≤-upper approximate distribution reduct of Ω^O.

According to the above definition, we can see that a ≥-lower(upper) approximate distribution reduct is the minimal subset of condition attributes preserves lower(upper) approximations of all upward unions of decision classes.

On the other hand, a ≤-lower(upper) approximate distribution reduct is the minimal subset of condition attributes preserves lower(upper) approximations of all downward unions of decision classes.

Theorem 4. Let Ω^O be an IODS in which $A \subseteq AT$, then

1. A is ≥-lower approximate distribution reduct ⇔ A is ≤-upper approximate distribution reduct ;

2. A is ≤-lower approximate distribution reduct ⇔ A is ≥-upper approximate distribution reduct .

Proof: 1. "⇒" : If A is ≥-lower approximate distribution reduct, then we have L≥A = L≥AT and L≥B ≠ L≥A for ∀B⊂A, i.e., A∼(CL≥ t)= AT∼(CL≥t) for 1≤ t ≤ n. Moreover, by Theorem 3, we have A∼(CL≥t)

=U−A∼(CL≤t-1), t =2, …n , then A∼(CL≤t-1)= AT∼(CL≤t-1) holds for t =2, …n. Since CL≤n=U, then A∼(CL≤n)= AT∼(CL≤n)= U. Thus, H≤A = H≤AT and H≤B ≠ H≤A for ∀B⊂A hold, A is ≤-upper approximate distribution reduct.

"⇐" : If A is ≤-upper approximate distribution reduct, then we have H≤A = H≤AT and H≤B ≠ H≤A for ∀B⊂A, i.e., A~(CL≤t)= AT~(CL≤t) for 1≤ t ≤ n. Moreover, by Theorem 3, we have A∼(CL≤t) = U−A∼(CL≥t+1), t =1, …n−1, then A∼(CL≥t+1)= AT∼(CL≥t+1) holds for t =1, …n−1. Since CL≥1=U, then A∼(CL≥1)= AT∼(CL≥1) =U. Thus, L≥A = L≥AT and L≥B ≠ L≥A for ∀B⊂A hold, A is ≥-lower approximate distribution reduct.

Proof 2. The proof of 2 is similar to the proof of 1.

Definition 8. Let ΩO be an IODS in which A ⊆ AT, denote by

$D_1^{\geq} = \{(x, y) : \forall CL^{\geq}_t, \forall x \in AT_{\sim}(CL^{\geq}_t), \forall y \in U - CL^{\geq}_t\}$,

$D_1^{\leq} = \{(x, y) : \forall CL^{\leq}_t, \forall x \in AT_{\sim}(CL^{\leq}_t), \forall y \in U - CL^{\leq}_t\}$,

$D_2^{\geq} = \{(x, y) : \forall CL^{\geq}_t, \forall x \in U - AT^{\sim}(CL^{\geq}_t), \forall y \in CL^{\geq}_t\}$,

$D_2^{\leq} = \{(x, y) : \forall CL^{\geq}_t, \forall x \in U - AT^{\sim}(CL^{\leq}_t), \forall y \in CL^{\leq}_t\}$,

Where

$$D_1^{\geq}(x,y) = \begin{cases} \{a \in AT : f(x,a) > f(y,a)\}, & (x,y) \in D_1^{\geq} \\ AT, & (x,y) \notin D_1^{\geq} \end{cases}$$

$$D_1^{\leq}(x,y) = \begin{cases} \{a \in AT : f(x,a) < f(y,a)\}, & (x,y) \in D_1^{\leq} \\ AT, & (x,y) \notin D_1^{\leq} \end{cases}$$

$$D_2^{\geq}(x,y) = \begin{cases} \{a \in AT : f(x,a) < f(y,a)\}, & (x,y) \in D_2^{\geq} \\ AT, & (x,y) \notin D_2^{\geq} \end{cases}$$

$$D_2^{\leq}(x,y) = \begin{cases} \{a \in AT : f(x,a) > f(y,a)\}, & (x,y) \in D_2^{\leq} \\ AT, & (x,y) \notin D_2^{\leq} \end{cases}$$

then $D_l^{\geq}(x, y)$ are called ≥-lower(upper) approximate distribution discernibility attributes sets respectively, where l = 1, 2, $D_l^{\leq}(x, y)$ are called ≤-lower(upper) approximate distribution discernibility attributes sets, respectively, $D_l^{\geq}$ are called ≥-lower(upper) approximate distribution discernibility matrices respectively and $D_l^{\leq}$ are called ≤-lower(upper) approximate distribution discernibility matrices respectively.

Theorem 5. Let ΩO be an IODS in which A ⊆ AT, then

1. *A* is ≥-lower approximate distribution consistent set ⇔ for $\forall (x, y) \in D_1^{\geq}$, $A \cap D_1^{\geq}(x, y) \neq \emptyset$ holds;
2. *A* is ≤-lower approximate distribution consistent set ⇔ for $\forall (x, y) \in D_1^{\leq}$, $A \cap D_1^{\leq}(x, y) \neq \emptyset$ holds;
3. *A* is ≥-upper approximate distribution consistent set ⇔ for $\forall (x, y) \in D_2^{\geq}$, $A \cap D_2^{\geq}(x, y) \neq \emptyset$ holds;
4. *A* is ≤-upper approximate distribution consistent set ⇔ for $\forall (x, y) \in D_2^{\leq}$, $A \cap D_2^{\leq}(x, y) \neq \emptyset$ holds.

Proof 1: "⇒": Suppose that x∈AT∼(CL≥t), y∈U−CL≥t such that A∩D1≥ (x, y) = ∅, then there must be (y, x)∈R≥(A), y∈[x]≥A . Since A is ≥-lower approximate distribution consistent set, then for ∀CL≥t, A∼(CL≥ t)= AT∼(CL≥t) holds, i.e. [x]≥A ⊆CL≥t ⇔ [x]≥AT ⊆CL≥t , y∈CL≥t , this is contrary to the assumption that y∈U−CL≥t .

"⇐": Suppose that A is not the ≥-lower approximate distribution consistent set, then L≥A ≠ L≥AT holds. Since A ⊆ AT, then there must be CL≥t such that [x]≥AT ⊆CL≥t and [x]≥A ⊄CL≥t . Here, [x]≥AT ⊆CL≥t ⇒ x∈ AT∼(CL≥t). On the other hand, [x]≥A ⊄CL≥t , then there must be y∈U such that (y, x)∈ R≥(A) and y∈U−CL≥t , that is, A∩D1≥ (x, y)=∅. From discussion above, we have the following : for ∀CL≥t, ∀x∈AT∼(CL≥t), ∀y∈U−CL≥t , if A∩D1≥ (x, y) ≠ ∅, then A is ≥-lower approximate distribution consistent set.

Proof 2: The proof of 2 is similar to the proof of 1.

Proof 3: "⇒": Suppose that x∈U−AT∼(CL≥t), y∈CL≥t such that A∩D2≥ (x, y) = ∅, then there must be (x, y)∈R≥(A), y∈[x]≤A . Since A is ≥-upper approximate distribution consistent set, then for ∀CL≥t, A∼(CL≥ t) = AT∼(CL≥t) holds, i.e. [x]≤A∩CL≥t =∅ ⇔ [x] ≤AT∩CL≥t =∅, y∉CL≥t , this is contrary to the assumption that y∈CL≥t .

"⇐": Suppose that A is not the ≥-upper approximate distribution consistent set, then H≥A ≠ H≥AT holds. Since A ⊆ AT, then there must be CL≥t such that [x]≤AT∩CL≥t = ∅ and [x]≤A∩CL≥t ≠∅. Here, [x]≤AT ∩CL≥t = ∅ ⇒ x∈U－AT∼(CL≥t). On the other hand, [x]≤A∩CL≥t ≠∅ , then there must be y∈U such that (x, y)∈R≥(A)and y∈CL≥t , that is, A∩D2≥ (x, y)=∅. From discussion above, we have the following : for ∀CL≥t, ∀x∈U－AT∼(CL≥t), ∀y∈CL≥t , if A∩D2≥ (x, y) ≠ ∅, then A is ≥-upper approximate distribution consistent set.

Proof 4: The proof of 4 is similar to the proof of 3.

Definition 9. Let Ω^O be an IODS, denote by

$$F_1^{\geq} = \vee\{\wedge\{a: a\in D_1^{\geq}(x, y)\}\},\ F_1^{\leq} = \vee\{\wedge\{a: a\in D_1^{\leq}(x, y)\}\},$$

$$F_2^{\geq} = \vee\{\wedge\{a: a\in D_2^{\geq}(x, y)\}\},\ F_2^{\leq} = \vee\{\wedge\{a: a\in D_2^{\leq}(x, y)\}\},$$

then $F_l^{\geq}$ are called the ≥-lower(upper) distribution discernibility functions where l = 1, 2, $F_l^{\leq}$ are called the ≤-lower(upper) distribution discernibility functions.

Theorem 6. Let Ω^O be an IODS, the minimal disjunctive normal form of each discernibility function $F_l^{\geq}$, $F_l^{\leq}$ (l = 1, 2) is

$$F_l^{\geq} = \bigvee_{k=1}^{t}(\bigwedge_{s=1}^{q_k} a_{ls}^{\geq}),\ F_l^{\leq} = \bigvee_{k=1}^{t}(\bigwedge_{s=1}^{q_k} a_{ls}^{\leq}),$$

denoted by $B_{lk}^{\geq} = \{a_{ls}^{\geq}: s = 1, 2, \ldots, q_k\}$, $B_{lk}^{\leq} = \{a_{ls}^{\leq}: s = 1, 2, \ldots, q_k\}$, *then* $\{B_{lk}^{\geq}: k=1, 2, \ldots, t\}$ are, respectively, the set of all the ≥-lower(upper) distribution reducts, $\{B_{lk}^{\leq}: k=1, 2, \ldots, t\}$ are, respectively, the set of all the ≤-lower(upper) distribution reducts.

Proof: It follows directly from Theorem 5 and the definition of minimal disjunctive normal forms of the discernibility functions.

5. An Illustrative Example

Let us employ the inconsistent incomplete ordered decision system showed in Tab.1 to illustrate the approach to knowledge reduction.

In Tab.1, $U=\{x_1, x_2, \ldots, x_{15}\}$ is the universe; $AT=\{a, b, c, d\}$ is the set of condition attributes where a = Mathematics, b= Physics, c= History, d=Literature; $D=\{e\}$ is the decision attribute such that e = Global evaluation; $V_a= V_b = V_c = V_d = V_e$ ={Bad, Medium, Good} is the domain of all attributes.

Suppose that the decision attribute e partitions the universe into the set such that CL={CL_1, CL_2, CL_3}={{Bad}, {Medium}, {Good}}={{x_1, x_2, x_3, x_4}, {x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}}, {x_{12}, x_{13}, x_{14}, x_{15}}}. Therefore,

- $CL^{\leq}_1= CL_1$, i.e. the class of (at most) bad students,
- $CL^{\leq}_2= CL_1\cup CL_2$, i.e. the class of at most medium students,
- $CL^{\geq}_2= CL_2\cup CL_3$, i.e. the class of at least medium students,
- $CL^{\geq}_3= CL_3$, the class of (at least) good students.

By Def. 3, we have

[x1]≥AT ={x1, x4, x5, x6, x7, x8, x9, x11, x13, x14, x15}, [x2]≥AT ={x2, x6, x7, x8, x11, x13, x15}, [x3]≥AT ={x3, x4, x8, x10, x11, x12, x13, x14, x15}, [x4]≥AT ={x4, x11, x14, x15}, [x5]≥AT ={x5, x7, x14}, [x6]≥AT ={x6, x13, x15}, [x7]≥AT ={x7, x14}, [x8]≥AT ={x7, x8, x10, x12, x14}, [x9]≥AT ={x4, x7, x9, x13, x14, x15}, [x10]≥AT ={x10, x14}, [x11]≥AT ={x4, x10, x11, x14, x15}, [x12]≥AT ={x7, x12, x14}, [x13]≥AT ={x13, x15}, [x14]≥AT ={x7, x14}, [x15]≥AT ={x15}.

[x1]≤AT ={x1}, [x2]≤AT ={x2}, [x3]≤AT ={x3}, [x4]≤AT ={x1, x3, x4, x9, x11}, [x5]≤AT ={x1, x5}, [x6]≤AT ={x1, x2, x6}, [x7]≤AT ={x1, x2, x5, x7, x8, x9, x12, x14}, [x8]≤AT ={x1, x2, x13, x8}, [x9]≤AT ={x1, x3, x9}, [x10]≤AT ={x3, x8, x10, x11}, [x11]≤AT ={x1, x2, x3, x4, x11}, [x12]≤AT ={x3, x8, x12}, [x13]≤AT ={x1, x2, x3, x6, x9, x13}, [x14]≤AT ={x1, x3, x4, x5, x7, x8, x9, x10, x11, x12, x14}, [x15]≤AT ={x1, x2, x3, x4, x6, x9, x11, x13, x15}.

By computation, we have

A∼(CL≥1) =U, A∼(CL≥2) ={x5, x6, x7, x8, x10, x12, x13, x14, x15}, A∼(CL≥3) ={x13, x15},

A∼(CL≤1) ={x1, x2, x3}, A∼(CL≤2) ={x1, x2, x3, x4, x5, x6, x9, x10, x11}, A∼(CL≤3) =U,

A∼(CL≥1) =U, A∼(CL≥2) ={x4, x5, x6, x7, x8, x9, x10, x11, x12, x13, x14, x15}, A∼(CL≥3) ={x7, x8, x12, x13, x14, x15},

A∼(CL≤1) ={x1, x2, x3, x4, x9}, A∼(CL≤2) ={x1, x2, x3, x4, x5, x6, x7, x8, x9, x10, x11, x12, x14},

A∼(CL≤3) =U.

According to the above results, we have

D1≥ = D2≤ ={{x5, x1},{x5, x2},{x5, x3},{x5, x4},{x6, x1},{x6, x2},{x6, x3},{x6, x4},{x7, x1},{x7, x2},{x7, x3},{x7, x4},{x8, x1},{x8, x2},{x8, x3},{x8, x4},{x10, x1},{x10, x2},{x10, x3},{x10, x4},{x12, x1},{x12, x2},{x12, x3},{x12, x4},{x13, x1},{x13, x2},{x13, x3},{x13, x4},{x14, x1},{x14, x2},{x14, x3},{x14, x4},{x15, x1},{x15, x2},{x15, x3},{x15, x4},{x13, x5},{x13, x6},{x13, x7},{x13, x8},{x13, x9},{x13, x10},{x13, x11}, {x15, x5},{x15, x6},{x15, x7},{x15, x8},{x15, x9},{x15, x10},{x15, x11}}.

D1≤ = D2≥ ={{x1, x5},{x1, x6},{x1, x7},{x1, x8},{x1, x9},{x1, x10},{x1, x11},{x1, x12},{x1, x13},{x1, x14},{x1, x15},{x2, x5},{x2, x6},{x2, x7},{x2, x8},{x2, x9},{x2, x10},{x2, x11},{x2, x12},{x2, x13},{x2, x14},{x2, x15},{x3, x5},{x3, x6},{x3, x7},{x3, x8},{x3, x9},{x3, x10},{x3, x11},{x3, x12},{x3, x13},{x3, x14},{x3, x15},{x4, x12},{x4, x13},{x4, x14},{x4, x15},{x5, x12},{x5, x13},{x5, x14},{x5, x15},{x6, x12},{x6, x13},{x6, x14},{x6, x15},{x9, x12},{x9, x13},{x9, x14},{x9, x15},{x10, x12},{x10, x13},{x10, x14},{x10, x15},{x11, x12},{x11, x13},{x11, x14},{x11, x15}}.

Therefore, we can get the ≥-lower approximate distribution discernibility matrix of Tab.1 such as Tab.2 shows. What should be noticed is that only pairs in D1≥ are presented in Tab.2.

Based on Def. 9, we have F1≥ = F2≤ = $a\wedge b\wedge(a\vee d)\wedge(a\vee b)\wedge(a\vee c)\wedge(b\vee d)\wedge(b\vee c)\wedge(c\vee d)\wedge(a\vee b\vee d)\wedge(a\vee c\vee d)\wedge(b\vee c\vee d)= a\wedge b\wedge d$, that is, {a, b, d} is the ≥-lower approximate distribution reduct and ≤-upper approximate distribution reduct of Tab.1.

Similarly, it is not difficult to work out that {a, b, c, d} is the ≤-lower approximate distribution reduct and ≥-upper approximate distribution reduct of Tab.1. In other words, no attribute can be omitted in order to keep the invariability of the lower approximate sets of all downward unions of decision classes and the upper approximate sets of all upward unions of decision classes.

Based on the ≥-lower approximate distribution reduct, we can get the following certain rules:

- If Mathematics is good, both Physics and Literature are medium or better, then Global evaluation is medium or better. / / Supported by the objects x_5 and x_7 in $A_{\sim}(CL^{\geq}_{2})$.
- If both Mathematics and Literature are medium or better, Physics is good, then Global evaluation is medium or better. / / Supported by the objects x_6, x_{13}, x_{15} in $A_{\sim}(CL^{\geq}_{2})$.
- If Mathematics is good, both Physics and Literature are bad or better, then Global evaluation is medium or better. / / Supported by the objects x_8, x_{10}, x_{12}, x_{14} in $A_{\sim}(CL^{\geq}_{2})$.
- If Mathematics is medium or better, both Physics and Literature are good, then Global evaluation is good. / / Supported by the objects x_{13}, x_{15} in $A_{\sim}(CL^{\geq}_{3})$.
- Based on the ≤-lower approximate distribution reduct, we can get the following certain rules:
- If both Mathematics and Physics are medium or worse, both History and Literature are bad, then Global evaluation is bad. / / Supported by the object x_1 in $A_{\sim}(CL^{\leq}_{1})$.
- If Mathematics, History and Literature are all bad, Physics is bad or worse, then Global evaluation is bad. / / Supported by the object x_2 in $A_{\sim}(CL^{\leq}_{1})$.
- If Mathematics, Physics and Literature are all bad, History is medium or worse, then Global evaluation is bad. / / Supported by the object x_3 in $A_{\sim}(CL^{\leq}_{1})$.
- If both Mathematics and Physics are medium or worse, both History and Literature are good or worse, then Global evaluation is medium or worse. / / Supported by the objects x_1, x_3, x_9, x_4 in $A_{\sim}(CL^{\leq}_{2})$.
- If both Mathematics and Literature are medium or worse, Physics is good or worse and History is bad, then Global evaluation is medium or worse. / / Supported by the objects x_2, x_6 in $A_{\sim}(CL^{\leq}_{2})$.

- If Mathematics is good or worse, both Physics and Literature are medium or worse and History is bad, then Global evaluation is medium or worse. / / Supported by the object x_5 in $A_{\sim}(CL^{\le}_2)$.
- If both Mathematics and Physics are good or worse, History is medium or worse and Literature is bad, then Global evaluation is medium or worse. / / Supported by the object x_8 in $A_{\sim}(CL^{\le}_2)$.
- If both mathematics and History are good or worse, both Physics and Literature are bad, then Global evaluation is medium or worse. / / Supported by the object x_{10} in $A_{\sim}(CL^{\le}_2)$.
- If Mathematics is medium or worse, both Physics and History are good or worse, Literature is bad, then Global evaluation is medium or worse. / / Supported by the object x_{11} in $A_{\sim}(CL^{\le}_2)$.

Table 2. ≥-lower approximate distribution discernibility matrix of Table 1

x \ y	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1															
x_2															
x_3															
x_4															
x_5	*a,d*	*a,d*	*a,b,d*	*a*											
x_6	*b,d*	*a,d*	*a,b,d*	*b*											
x_7	*a,c,d*	*a,c,d*	*a,b,d*	*a*											
x_8	*a*	*a,b*	*a*	*a*											
x_9															
x_{10}	*a,c*	*a,c*	*a,c*	*a*											
x_{11}															
x_{12}	*a,c,d*	*a,c,d*	*a,d*	*a*											
x_{13}	*b,c,d*	*a,c,d*	*a,b,d*	*b*	*b,c,d*	*b,d*	*b*	*d*	*b*	*b,d*	*d*				
x_{14}	*a,d*	*a,d*	*a,b,d*	*a*											
x_{15}	*b,c,d*	*a,c,d*	**AT**	*b*	*b,c,d*	*c,d*	*b,c*	*c,d*	*b,c*	*b,d*	*d*				

6. Conclusions

To deal with the incomplete information system by rough set theory, many researchers have generalized the indiscernibility relation to more general relations. It is noticeable that the attributes in the incomplete information system have not been considered as the criteria in most recent research literatures. Based on the expanded dominance relation that was proposed in Ref. 14, this paper presents an explorative research focusing on the approach to knowledge reduction in the incomplete decision system in which all attributes are regarded as criteria. Four new notions of approximate distribution reduct are proposed in the IODS. These approximate distribution reducts are the minimal sets of attributes, which preserve lower and upper approximations of all the downward and upward unions of decision classes respectively. From discussion above, this paper provides a qualitative theoretical framework that may be important for analysis of rules' acquisition in incomplete information system with the ordering properties of criteria.

In our further research, we will develop the proposed approaches to knowledge reduction to the incomplete ordered decision system in which unknown values have some other semantic explanations, e.g. the unknown values are "do not care" conditions [8].

7. References

[1] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*. 1984, **11**: 341~356.

[1] Z. Pawlak. Rough set theory and its applications to data analysis. *Cybernetics and Systems*. 1998, **29**: 661~688.

[2] Z. Pawlak. Rough sets and intelligent data analysis. *Information Sciences*. 2002, **147**: 1~12.

[3] Z. Pawlak, A. Skowron. Rudiments of rough sets. *Information Sciences*. 2007, **177**: 3~27.

[4] T. P. Hong, L. H. Tseng, S. L. Wang. Learning rules from incomplete training examples by rough sets. *Expert Systems with Application*. 2002, **22**: 285-293.

[5] C. C. Chan. A rough set approach to attribute generalization in data mining. *Information Sciences*. 1998, **107**: 169~176.

[6] D. Kim. Data classification based on tolerant rough set. *Pattern Recognition*. 2001, **34**: 1613~1624.

[7] J. W. Grzymala-Busse. Data with missing attribute values: generalization of indiscernibility relation and rule rnduction. *Transactions on Rough Sets I, Lecture Notes in Computer Science*, Springer-Verlag, Berlin. 2004, vol. **3100**: pp. 78~95.

[8] J. W. Grzymala-Busse, A. Y. Wang. Modified algorithms LEM1 and LEM2 for rule induction from data with missing attribute values, in: *Proceeding of the Fifth International Workshop on Rough Sets and Soft Computing at the Third Joint Conference on Information Sciences*, Research TrianglePark. NC, 1997, pp. 69~72.

[9] Y. Y. Guan, H. K. Wang. Set-valued information systems. *Information Sciences*. 2006, **176**: 2507~2525.

[10] M. Kryszkiewicz. Rough set approach to incomplete information systems. *Information Sciences*. 1998, **112**: 39~49.

[11] M. Kryszkiewicz. Rules in incomplete information systems. *Information Sciences*. 1999, **113**: 271~292.

[12] Y. Leung, D. Y. Li. Maximal consistent block technique for rule qcquisition in incomplete information systems. *Information Sciences*. 2003, **15**: 85~106.

[13] M. W. Shao, W. X. Zhang. Dominance relation and rules in an incomplete ordered information system. *International Journal of Intelligent System*. 2005, **20**: 13~27.

[14] J. Stefanowski, A. Tsoukias. Incomplete information tables and rough classification. *Computational Intelligence*. 2001. **17**: 545-566.

[15] G. Y. Wang, Extension of rough set under incomplete information systems, in: *Proceedings of the 11th IEEE International Conference on Fuzzy Systems*, Hawaii, USA. May 12-17, 2002, pp. 1098~1103.

[16] W. Z. Wu, W. X. Zhang, H. Z. Li. Knowledge acquisition in incomplete fuzzy information systems via the rough set approach. *Expert Systems*. 2003, **20**(5): 280~286.

[17] J. Błaszczynski, S. Greco, R. Słowinski. Multi-criteria classification-a new scheme for application of dominance-based decision rules. *European Journal of Operational Research*. 2007, **181**: 1030~1044.

[18] S. Greco, B. Matarazzo, R. Słowinski. Rough sets theory for multicriteria decision analysis. *European journal of operational research*. 2002, **129**: 1~47.

[19] S. Greco, B. Matarazzo, R. Slowinski. Rough approximation by dominance relations. *International journal of intelligent systems*. 2002, **17**: 153~171.

[20] W. X. Zhang, J. S. Mi, W. Z. Wu. Approaches to knowledge reductions in inconsistent systems. *International Journal of Intelligent Systems*. 2003, **18**: 989~1000.

[21] M. Kryszkiewicz. Comparative study of alternative types of knowledge reduction in inconsistent systems. *International Journal of Intelligent Systems*. 2001, **16**: 105~120.

[22] G. Y. Wang. Rough reduction in algebra view and information view. *International Journal of Intelligent Systems*. 2003, **18**: 679~688.

[23] D. Slezak. Approximate entropy reducts. *Fundamenta Informaticae*. 2002, 53: 365~390.

[24] J. S. Mi, W. Z. Wu, W. X. Zhang. Approaches to knowledge reduction based on variable precision rough set model. *Information Sciences*. 2004, 159: 255~272.