Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (9): 1621-1628, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1621.1628

Set Approximation in Incomplete Data

Renpu Li
Department of Computer Science and Technology, School of Information and Electrical Engineering, Ludong University, Yantai 264025, China

Abstract: The problem of set approximation in incomplete data is addressed. Different with complete data where the upper/lower approximation of an object set is certain and can be given by one set, for incomplete data upper/lower approximation of a set is uncertain and needs to be bracketed by a set pair. From the completion view of incomplete data, the semantic interpretations of four boundaries used to approximate a set in incomplete data are given. It is illustrated that existing definitions based on tolerance class or covering are not enough to describe precisely the set approximation in incomplete data. Based on a concept of interval granule, new methods are presented for incomplete data to compute the four approximation boundaries of a set. This study provides a new view of granular computing on set approximation in incomplete data and is helpful for computing the uncertainty of a set more accurately.

Key words: Incomplete data, set approximation, completion, interval granule

INTRODUCTION

In the classical rough set modal (Pawlak, 1991), the concept of rough set approximation is based on the concept of equivalence relation. In general, an equivalence relation is firstly defined for an information table and the equivalence classes induced by the equivalence relation can be thought as elementary or basic sets defined on a set of attributes. A set is said to be definable if it is the union of some equivalence classes (Grzymala-Busse, 2005; Jarvinen and Kortelainen, 2004; Pawlak *et al*., 1995), otherwise it is undefinable. In order to characterize an undefinable set, a pair of lower and upper approximations is defined. The lower approximation of a set is the union of equivalence classes which is a subset of the set and its upper approximation is the union of those equivalence classes which has a non-empty overlap with the set.

Three different definitions of set approximations, called, respectively object-based, granule-based and subsystem-based, are discussed (Yao, 2003). For the complete data, three definitions are equivalent, but have different semantical interpretations. For the incomplete data, a direct extension of the object-based definition (Kryszkiewicz, 1998) is proposed by replacing equivalence classes of objects with similarity classes of objects. Another definition directly generalizing the granule-based definition is provided (Yao, 1998). In this definition, a covering on the object universe is used to substitute for the partition. However, upper and lower approximations in above two definitions are not dual in incomplete data. Two dual approximation pairs are studied in (Yao, 2003, 2001). They extend granule-based definition in two ways. Either the lower or the upper approximation operator is extended and the other one is defined by duality. It has been verified that granule-based and object-based approximations may not the same (Yao, 2001) in incomplete data. Couso and Dubois (2011), under the view that a covering corresponds to a family of possible partitions induce by the selection function, two dual approximation pairs proposed by Yao can be computed by using the union and intersection operators on upper and lower approximations of possible partitions. Also the upper/lower approximation of a set in incomplete data is bracketed by the pair combined by two upper/lower approximations of two dual pairs.

In this study, we studied the problem of set approximation in incomplete data from the view of completion. An incomplete information table is thought as a family of possible completions, one of them is the actual one. Follow the idea of (Couso and Dubois, 2011), the upper/lower approximation of a set has its own upper and lower approximation which are called upper-upper approximation, upper-lower approximation, lower-upper approximation and lower-upper approximation, respectively. These four measures have clear semantical interpretations. For example, upper-upper/upper-lower approximation of a set can be interpreted as a set of objects which certainly/possibly belongs to its upper approximation.

It is verified that existing definitions are not enough to compute precisely all four measures for approximating a set in incomplete data. Based on a concept of interval granule, new computation formulations are proposed for set approximation in incomplete data, they not only provides a way to compute precisely the uncertainty of a set in incomplete data, but gives a new view of applying the idea of granular computing to reasoning in incomplete data.

INFORMATION TABLES

In the rough set theory, data is usually given in the form of information table in which a finite set of objects is described by using a finite set of attributes. In this study, a basic assumption is that any object on any attribute possesses only one value. If the value is unknown its real value is one known value of the value domain. For an information table, if all values in an information table are known, it is complete, or it is incomplete.

Complete and incomplete information tables:

Definition 1: A complete information table T' is expressed as the tuple:

$$T' = (U, At, V, f')$$

where, U is a finite nonempty set of objects, At is a finite nonempty set of attributes, $V = \{V_a|a\in At\}$ and V_a is a nonempty set of values for an attribute $a\in At$, $f' = \{f'_a|a\in At\}$ and f'_a: $U\rightarrow V_a$ is an information function.

Definition 2: An incomplete information table T is expressed as the tuple:

$$T = (U, At, V, f')$$

where, U is a finite nonempty set of objects, At is a finite nonempty set of attributes, $V = \{V_a|a\in At\}$ and V_a is a nonempty set of values for an attribute $a\in At$, $f = \{f_a|a\in At\}$ and f_a: $U\rightarrow v_a\cup P(V_a)$ is an information function where $P(V_a)$ is the power set of V_a.

A single-valued information table is characterized by an information function f_a: $U\rightarrow V_a$ and a set-valued table by f_a: $U\rightarrow P(V_a)$. Our definition of an information table combines the standard definitions of single-valued and set-valued information tables (Orlowska, 1998; Pawlak, 1981). This enables us to study both complete and incomplete information regarding values of an object.

If $f_a(x)\in V_a\cup\{\phi\}$, we have complete information about the value of x on a. By the empty set ϕ, we denote that the attribute a is not applicable to x. For simplicity, we do not consider this case in the current study by assuming that information function is of the form f_a: $U\rightarrow V_a\cup(P(V_a)-\{\phi\})$. Furthermore, we assume that for any $x\in U$ and $a\in At$, x can only take exactly one value in V_a. However, due to a lack of knowledge, we may only have partial information about the actual value of x on a. Such partial information is expressed as a subset of V_a, $f_a(x)\in P(V_a)$. That is, we only know that the values of x on a is in the set $f_a(x)$, but do not know which one is the actual value.

Defnition 3: Let T = (U, At, V, f) be an information table. If for all $a\in At$ and $x\in U$, $f_a(x)\in V_a$, it is called a complete information table (CIT), otherwise it is an incomplete information table (IIT).

Example 1: Table 1 is an example of an incomplete information table T = (U, At, V, f), where $U = \{O_1, O_2, O_3, O_4\}$, At = {a, b}, $V_a = \{0, 1\}$, $V_b = \{0, 1\}$.

An incomplete table as a family of complete tables: In an incomplete information table, a set of values is used to indicate all possibilities of the actual value of an object. Within this framework, an incomplete information table can be interpreted as a family of complete information tables that are consistent with it (Couso and Dubois, 2011; Kryszkiewicz, 1999; Lipski, 1979; Nakamura, 1996).

Definition 4: Let T = (U, At, V, f) be an incomplete information table. A complete information table T' = (U, At, V, f') is called a completion of T if and only if for all $a\in At$ and $x\in U$, $f_a(x)\in V_a$ implies $f'_a(x) = f_a(x)$ and $f_a(x)\in P(V_a)$ implies $f'_a(x)\in f_a(x)$. The set of all completions of T is denoted as COMP(T).

Assume that the number of set-valued elements of T is m and the cardinalities of the set-valued elements are given by $N_i(i = 1, 2,..., m)$. The number of completions of T is:

$$\prod_{i=1} N_i$$

Example 2: Consider Table 1 given in example 1. The number of completions of IT can be computed by $|f_a(O_1)|\times|f_a(O_3)|\times|f_b(O_4)| = 2\times2\times2 = 8$, where |.| denotes the cardinality of a set. The eight completions from T1-T8 are given in Fig. 1.

Although, only one of COMPT(T) is the actual table, we do not know which one due to incomplete information. The set of all completions COMP(T) provides an interpretation of an incomplete information table which is

Table 1: An incomplete information table IT

U	a	b
o_1	{0, 1}	1
o_2	0	0
o_3	{0, 1}	1
o_4	0	{0, 1}

T1

U	a	b
o_1	0	1
o_2	0	0
o_3	0	1
o_4	0	0

T2

U	a	b
o_1	1	1
o_2	0	0
o_3	0	1
o_4	0	0

T3

U	a	b
o_1	0	1
o_2	0	0
o_3	1	1
o_4	0	0

T4

U	a	b
o_1	1	1
o_2	0	0
o_3	1	1
o_4	0	0

T5

U	a	b
o_1	0	1
o_2	0	0
o_3	0	1
o_4	0	1

T6

U	a	b
o_1	1	1
o_2	0	0
o_3	0	1
o_4	0	1

T7

U	a	b
o_1	0	1
o_2	0	0
o_3	1	1
o_4	0	1

T8

U	a	b
o_1	1	1
o_2	0	0
o_3	1	1
o_4	0	1

Fig. 1: Completions of incomplete Table 1

consistent with the interpretation of a set-valued element of T. Recall that any value $f'_a(x) \in V_a$ may be the actual value of x. By the definition of a completion, any complete table T'∈COMPT(T) may be the actual information table.

In other words, the available information only allow us to infer the possible values of a complete table, but not the actual values. If one considers a particular table from COMP(T), some extra information need to be introduced. Thus, it is more reasonable to interpret an incomplete information table T as a set of its completions COMP(T).

ROUGH SET APPROXIMATION

A logic language: Granular computing is an effective way of thinking and can be applied in many complex problems involves incomplete, uncertain, or vague information (Yao, 2001). A logic language has been proposed in (Yao and Zhou, 2007) and is very efficient on the set definition and approximation (Yao, 2007).

Let T = (U, At, V, f) be a complete information table, an atomic formula on an attribute a∈At is given as a = v, where v∈Va, an object x∈U satisfies it if $f_a(x) = v$, written x| = (a = v). All atomic formulas on a is denoted as a.

For a set of attribute A = $(a_1, a_2, \ldots, a_n) \subseteq At$, a compound formula ψ on A can be expressed $\varphi = (a_1 = v_1) \wedge (a_2 = v_2)$ where $v_i \in V_{a_i}$. An object x∈U satisfies ψ if $f_{a_i}(x) = v_i$ for ∀i∈{1, 2,..., n}, written x| = φ. All formulas on A is denoted as A. At is simplified as in this study if there is no confusion.

If ψ is a formula, the set of objects m(ψ) is defined by:

$$m(\psi) = \{x \in U | x| = \psi\}$$

For ∀x∈U, $m^{-1}(x)$ is defined by:

$$m^{-1}(x) = \{\psi \in \mathcal{L} \mid x \models \psi\}$$

m(ψ) includes all objects satisfying ψ and $m^{-1}(x)$ is the formula satisfied by x.

Approximation in complete information tables: Let T = (U, At, V, f) be a complete information table and X⊆U. Upper and lower approximation of X can be defined, respectively as:

$$\underline{apr}(X) = \{x \in U \mid [x] \subseteq X\} = \cup\{m(\psi) \mid \psi \in \mathcal{L}, m(\psi) \subseteq X\}$$

$$\overline{apr}(X) = \{x \in U \mid [x] \cap X \neq \phi\} = \cup\{m(\psi) \mid \psi \in \mathcal{L}, m(\psi) \cap X \neq \phi\}$$

where $[x] = m(m^{-1}(x))$ is the set of objects which satisfies the same formula with x.

For complete information tables $\underline{apr}(X)$ and $\overline{apr}(X)$ are dual to each other in the sense:

$$\underline{apr}(X) = (\overline{apr}(X^c))^c$$

$$\overline{apr}(X) = (\underline{apr}(X^c))^c$$

where X^c = U-X is the complement of X.

Approximation in incomplete information tables: As far as the approximation of a set X⊆U in an incomplete information table T, assume that R∈COMP(T) be the actual table of T and the upper and lower approximations of X be $\overline{apr}_R(X)$ and $\underline{apr}_R(X)$, respectively. Because we do not know in COMP(T) which one is R, so we can not obtain the certain values of $\overline{apr}_R(X)$ and $\underline{apr}_R(X)$. However, under the condition that the actual table is one of COMP(T), we can compute the upper and lower approximations for $\overline{apr}_R(X)$ and $\underline{apr}_R(X)$, respectively.

Let $\overline{apr}^u(X)$ and $\overline{apr}^l(X)$ be the upper and lower approximation of $\overline{apr}_R(X)$, $\overline{apr}^u(X)$ and $\overline{apr}^l(X)$ be the upper and lower approximation of $\underline{apr}_R(X)$. We have following definitions:

$$\overline{apr}_T{}^u(X) = \bigcup_{T' \in COMP(T)} \overline{apr}_{T'}(X)$$
$$= \{x \in U \mid \exists T' \in COMP(T)[x \in \overline{apr}_{T'}(X)]\}$$

$$\overline{apr}_T^{\,l}(X) = \bigcap_{T' \in COMP(T)} \overline{apr}_{T'}(X)$$
$$= \{x \in U \mid \forall T' \in COMP(T)[x \in \overline{apr}_{T'}(X)]\}$$

Above four measures have clear semantic interpretations for the set approximation of a set of an incomplete information table. For example, $\overline{apr}_T^{\,u}(X)$ contains a set of objects which possibly belong to $\overline{apr}_R(X)$. In other words, for any object $x \in \overline{apr}_T^{\,u}(X)$, there exists at least one completion T'COMP(T) such that $x \in \overline{apr}_{T'}(X)$. While v∊Va contains a set of objects which certainly belong to $\overline{apr}_R(X)$, that is, for any completion T'∊COMP(T), $\forall x \in \overline{apr}_T^{\,l}(X)$, we have $x \in \overline{apr}_{T'}(X)$. The similar interpretations can be induced for $\underline{apr}^u{}_T(X)$ and $\forall x \in \overline{apr}_T^{\,l}(X)$. For an incomplete information table T, it is clear that $\underline{apr}^l{}_T(X)$.

$$\underline{apr}_T^{\,l}(X) \subseteq \underline{apr}_R(X) \subseteq \underline{apr}_T^{\,u}(X)$$

$$\overline{apr}_T^{\,l}(X) \subseteq \overline{apr}_R(X) \subseteq \overline{apr}_T^{\,u}(X)$$

Example 3: Consider Table 1 given in example 1. Let X = (O_2, O_3, O_4), based on its eight completions in Fig. 1, four measures for approximation of X can be computed as follows:

$$\underline{apr}_{T_1}(X) = (o_2, o_4), \overline{apr}_{T_1}(X) = (o_1, o_2, o_3, o_4)$$

$$\underline{apr}_{T_2}(X) = (o_2, o_3, o_4), \overline{apr}_{T_2}(X) = (o_2, o_3, o_4)$$

$$\underline{apr}_{T_3}(X) = (o_2, o_3, o_4), \overline{apr}_{T_3}(X) = (o_2, o_3, o_4)$$

$$\underline{apr}_{T_4}(X) = (o_2, o_4), \overline{apr}_{T_4}(X) = (o_1, o_2, o_3, o_4)$$

$$\underline{apr}_{T_5}(X) = (o_2), \overline{apr}_{T_5}(X) = (o_1, o_2, o_3, o_4)$$

$$\underline{apr}_{T_6}(X) = (o_2, o_3, o_4), \overline{apr}_{T_6}(X) = (o_2, o_3, o_4)$$

$$\underline{apr}_{T_7}(X) = (o_2, o_3), \overline{apr}_{T_7}(X) = (o_1, o_2, o_3, o_4)$$

$$\underline{apr}_{T_8}(X) = (o_2, o_4), \overline{apr}_{T_8}(X) = (o_1, o_2, o_3, o_4)$$

And then, we can get:

$$\underline{apr}_T^{\,l}(X) = \{o_2\}$$

$$\underline{apr}_T^{\,u}(X) = \{o_2, o_3, o_4\}$$

$$\overline{apr}_T^{\,l}(X) = \{o_2, o_3, o_4\}$$

$$\overline{apr}_T^{\,u}(X) = \{o_1, o_2, o_3, o_4\}$$

COMPUTATION OF APPROXIMATIONS

For an incomplete information table, it is inefficient to obtain the values of four measures of set approximation through computing the upper and lower approximation of all completions. In this section, by extending the logic language to incomplete information table, a new concept called interval granule is constructed for describing the granulation structure of an incomplete information table T. It can approximate the granules induced by the formulas in COMP(T) from upper and lower directions. Using this structure, four measures for set approximation of incomplete information tables can be computed efficiently.

Interval granules

Definition 5: Let T = (U, At, V, f) be an incomplete information table. For an atomic formula (a = v) where a∊At and v∊V_a, an object x∊U certainly satisfies it if $f_a(x) = v$, written $x \models^{\Box} (a = v)$. x possibly satisfies it if v∊$f_a(x)$, written $x \models^{\Diamond} (a = v)$.

Definition 6: Let T = (U, At, V, f) and A = {a_1, a_2,...,a_n}⊆At, a compound formula ψ on A can be expressed ($a_1 = v_1 \wedge a_2 = v_2 \wedge ... \wedge a_n = v_n$) where v_i∊V_{a_i} i = 1, 2,..., n. An object x∊U is called certainly satisfied to ψ if $f_{a_i}(x) = v_i$ for any i written $x \models^{\Box}{}_A \Psi$, x possibly satisfies it if v_i∊$f_{a_i}(x)$ or $f_{a_i}(x) = v_i$ for any i, written $x \models^{\Diamond}{}_A \Psi$.

Definition 7: Let T = (U, At, V, f) be an information table and ψ a formula of $\mathcal{L}_{At}$, two sets $\underline{m}(\psi)$ and $\overline{m}(\psi)$ are defined, respectively by:

$$\underline{m}(\psi) = \{x \in U \mid x \models^{\Box} \psi\} = \{x \in U \mid \forall T' \in COMP(T)[x \models \psi]\}$$

$$\overline{m}(\psi) = \{x \in U \mid x \models^{\Diamond} \psi\} = \{x \in U \mid \exists T' \in COMP(T)[x \models \psi]\}$$

The set pair $m(\psi) = (\underline{m}(\psi), \overline{m}(\psi))$ is called an interval granule of ψ. Let $\mathcal{L} = \{\psi_1, \psi_2, ..., \psi_k\}$ and k is the number of formules on At, the granular structure on U is:

$$\tau = \{m(\psi_1), m(\psi_2), ... m(\psi_k)\} = \{(\underline{m}(\psi_1), \overline{m}(\psi_1)), (\underline{m}(\psi_2), \overline{m}(\psi_2)), ... (\underline{m}(\psi_k), \overline{m}(\psi_k))\}$$

Definition 8: Let T = (U, At, V, f) be an information table and x∈U, two sets $\underline{m}^{-1}(x)$ and $\overline{m}^{-1}(x)$ are defined, respectively by:

$$\underline{m}^{-1}(x) = \{\psi \in \mathcal{L} | x \vDash^{\square} \psi\} = \{\psi \in \mathcal{L} | \forall T' \in COMP(T)[x \vDash \psi]\}$$

$$\overline{m}^{-1}(x) = \{\psi \in \mathcal{L} | x \vDash^{\diamond} \psi\} = \{\psi \in \mathcal{L} | \exists T' \in COMP(T)[x \vDash \psi]\}$$

Example 4: For the incomplete table IT in Table 1, the formulas and granular structure induced by attribute a are as follows:

$$\mathcal{L}_a = \{a = 0, a = 1\}$$

$$\tau_a = \{(\{o_2, o_4\}, \{o_1, o_2, o_3, o_4\}), (\varphi, \{o_1, o_3\})\}$$

For the object o_4:

$$\underline{m}^{-1}(o_4) = \{a = 0\}$$

$$\overline{m}^{-1}(o_4) = \{a = 0\}$$

Example 5: For the Table 1 the formulas and granular structure induced by attribute At are as follows:

$$\mathcal{L} = \{(a = 0 \wedge b = 0), (a = 0 \wedge b = 1), (a = 1 \wedge b = 1)\}$$

$$\tau = \{(\{o_2\}, \{o_2, o_4\}), (\phi, \{o_1, o_3, o_4\}), (\phi, \{o_1, o_3\})\}$$

For the object o_4:

$$\underline{m}^{-1}(o_4) = \{\phi\}$$

$$\overline{m}^{-1}(o_4) = \{\psi_2, \psi_3\} = \{(a = 0 \wedge b = 0), (a = 0 \wedge b = 1)\}$$

For a complete information table, such as T'∈COMP(T), for $\forall \psi \in \mathcal{L}$, $\underline{m}(\psi) = \overline{m}(\psi)$. Let $I_T(\psi) = \{x | x \in U, x \vDash \psi\}$, called the equivalence class induced by ψ on T'. While it is clear that for ∀x∈U, $\underline{m}^{-1}(x) = \overline{m}^{-1}(x) = m^{-1}(x)$, for simplication, they are denoted as $m^{-1}_{T'}(x)$.

Theorem 1: Let T = (U, At, V, f) be a complete information table, x∈U and $\psi \in \mathcal{L}$. If $\psi = m^{-1}_{T'}(x)$, then $I_T(\epsilon) = [x]_T$.

Theorem 2: Let T = (U, At, V, f) be an incomplete information table and $\psi \in \mathcal{L}$, for any T'∈COMP(T), $\underline{m}_T(\psi) \subseteq I_{T'}(\psi) \subseteq \overline{m}_T(\psi)$.

Theorem 3: Let T = (U, At, V, f) be an incomplete information table and $\psi \in \mathcal{L}$. For ∀X satisfying $\underline{m}(\psi) \subseteq X \subseteq \overline{m}(\psi)$, ∃T'∈COMP(T) such that $I_{T'}(\psi) = X$.

Proposition 1: Let T = (U, At, V, f), x∈U, $\psi \in m^{-1}(x)$ and U. For ∀X satisfying $\underline{m}(\psi) \subseteq X \subseteq \overline{m}(\psi)$, ∃T'∈COMP(T) such that $[x]_T = x$.

Approximation computation based on interval granules: Let T = (U, At, V, f), X⊆U, $\mathcal{L}$ be the formula set of U on At. Based on the concept of interval granule, new formulations for computing four measures to approximate a set of incomplete information table are given by the following theorems.

Theorem 4:

- $\underline{apr}^l(X) = (\underline{apr}^u(X^c)) = \left(\bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) | \overline{m}(\psi) \not\subseteq X\} \right)^c$
- $\underline{apr}^u(X) = \bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) \cap X | \underline{m}(\psi) \subseteq X\}$
- $\overline{apr}^l(X) = (\bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) - X | \underline{m}(\psi) \cap X = \phi\})^c$
- $\overline{apr}^u(X) = \bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) | \overline{m}(\psi) \cap X \neq \phi\}$

Proof: Let:

$$P_1 = \left(\bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) | \overline{m}(\psi) \not\subseteq X\} \right)$$

For $\forall x \in P_1 \Rightarrow \exists \psi \in \mathcal{L}$ such that $x \in \overline{m}(\psi)$ and $\overline{m}(\psi) \not\subset X$ where $\psi \in m^{-1}(x)$, according to Proposition 1⇒∃T'∈COMP(T) such that $x \in [x]_T = \overline{m}(\psi) \not\subset X \Rightarrow$ x does not belong to $\underline{apr}(X)_{T'}$, according to the definition of $\underline{apr}^l(X) \Rightarrow x \in (\underline{apr}^l(X))^c \Rightarrow P_1 \subseteq (\underline{apr}^l(X))^c$.

On the other hand, for $\forall x \in (\underline{apr}^l(X))^c \Rightarrow \exists T' \in COMP(T)$ such that x does not belong to $\underline{apr}(X)_{T'} \Rightarrow [x]_{T'} \not\subseteq X$, let $\psi = m^{-1}_{T'}(x) \Rightarrow I_{T'}(\psi) = [x] \not\subseteq X$, according to theorem 2 $\Rightarrow I_{T'}(\psi) \subseteq \overline{m}(\psi) \not\subset X \Rightarrow (\underline{apr}^l(X))^c \subseteq P_1$.

Finally we have $(\underline{apr}^l(X))^c = P_1$, $\underline{apr}^l(X) = (P_1)^c$ is proved Let:

$$P_2 = \bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) \cap X | \underline{m}(\psi) \subseteq X\}$$

For $\forall x \in P_2 \Rightarrow \exists \psi \in \mathcal{L}$ such that $\psi \in m^{-1}(x) \Rightarrow x \in X$.

Because $\underline{m}(\psi) \subseteq X$ and $x \in X \Rightarrow (x \cup \underline{m}(\psi)) \subseteq X$. Also, $\psi \in m^{-1}(x) \Rightarrow \underline{m}(\psi) \subseteq (x \cup \underline{m}(\psi)) \subseteq \overline{m}(\psi)$, according to theorem 1⇒∃T'∈COMP(T) such that $[x]_T = (x \cup \underline{m}(\psi)) \subseteq X \Rightarrow x \in \underline{apr}(X)_{T'} \Rightarrow x \in \underline{apr}^u(X)_T \Rightarrow P_2 \subseteq \underline{apr}^u(X)_T$.

On the other hand, for $\forall x \in \underline{apr}^u(X)_T \Rightarrow \exists T' \in COMP(T)$ such that $x \in \underline{apr}(X)_{T'} \Rightarrow [x]_{T'} \subseteq X$. Assume $\psi = m_T^{-1}(x), \Rightarrow I_{T'}(\psi) = [x]_{T'} \subseteq X$, according to theorem 2 $\Rightarrow \underline{m}(\psi) \subseteq I_{T'}(\psi) \subseteq X$ and $x \in I_{T'}(\psi) \subseteq \overline{m}(\psi)$. Because $[x]_{T'} \subseteq X \Rightarrow x \in X \Rightarrow x \in (\overline{m}(\psi) \cap X) \Rightarrow x \in P_2 \Rightarrow \underline{apr}^u(X)_T \subseteq P_2$.

So, $\underline{apr}^u(X)_T = P_2$ is proved.

Let:

$$P_3 = (\bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) - X \mid \underline{m}(\psi) \cap X = \phi\})$$

According to the definition $\overline{apr}^1(X) \Rightarrow \forall x \in (\overline{apr}^1(X))^c \Rightarrow \exists T'$ COMP(T) of such that x does not belong to $\overline{apr}_{T'}(X) \Rightarrow [x]_{T'} \cap X = \phi$. Let $\psi = m_{T'}^{-1}(x) \Rightarrow \underline{m}(\psi) \subseteq I_{T'}(\psi) = [x]_{T'} \Rightarrow \underline{m}(\psi) \cap X = \phi$. Because $I_{T'}(\psi) = [x]_{T'} \subseteq \overline{m}(\psi)$ and $[x]_{T'} \cap X = \phi \Rightarrow x \in \overline{m}(\psi)$ and $x \in X^c \Rightarrow x \in (\overline{m}(\psi) - X) \Rightarrow x \in P_3 \Rightarrow (\overline{apr}^1(X))^c \subseteq P_3$.

On the other hand, for $\forall x \in P_3 \Rightarrow \exists \psi$ such that $x \in (\overline{m}(\psi) - X)$ and $\underline{m}(\psi) \cap X = \phi \Rightarrow x \in X^c$ and $x \in \overline{m}(\psi) \Rightarrow (x \cup \underline{m}(\psi)) \cap X = \phi$. Because $\underline{m}(\psi) \subseteq (x \cup \underline{m}(\psi)) \subseteq \overline{m}(\psi)$, according to theorem 2$\Rightarrow \exists T' \in$COMP(T) such that $[x]_{T'} = I_{T'}(\psi) = (x \cup \underline{m}(\psi)) \Rightarrow [x]_{T'} \cap X = \phi \Rightarrow x \in (\overline{apr}^1(X))^c \Rightarrow P_3 \subseteq (\overline{apr}^1(X))^c$.

Finally, $(\overline{apr}^1(X))^c = P_3$, $\overline{apr}^1(X) = (P_3)^c$ is proved.

Let:

$$P_4 = \bigcup_{\psi \in \mathcal{L}} \{\overline{m}(\psi) \mid \overline{m}(\psi) \cap X \neq \phi\}$$

For $\forall x \in P_4 \Rightarrow \exists \psi \in \mathcal{L}$ such that $x \in \overline{m}(\psi)$ and $\overline{m}(\psi) \cap X \neq \phi$. According to theorem 1$\Rightarrow \exists T' \in$COMP(T) such that $[x]_{T'} = \overline{m}(\psi) \Rightarrow [x]_{T'} \cap X \neq \phi$ and $x \in \overline{apr}_{T'}(X) \Rightarrow P_4 \subseteq \overline{apr}_T^u(X)$.

On the other hand, for $\forall x \in \overline{apr}_T^u(X) \Rightarrow \exists T' \in COMP(T)$ such that $x \in \overline{apr}_{T'}(X) \Rightarrow [x]_{T'} \cap X \neq \phi$. Let $\psi = m_{T'}^{-1}(x) \Rightarrow \underline{m}(\psi) \subseteq [x]_{T'} \subseteq \overline{m}(\psi) \Rightarrow \overline{m}(\psi) \cap X \neq \phi \Rightarrow x \in P_4 \Rightarrow \overline{apr}_T^u(X) \subseteq P_4$.

Finally $\overline{apr}_T^u(X) = P$ is proved.

COMPARISON WITH EXISTING APPROXIMATION COMPUTATION

In order to reasoning in an incomplete information table, classical rough set theory needs to be generalized. Correspondingly, several generalized definitions for set approximation have been presented by using tolerance classes or coverings.

Let T = (U, At, V, f) be an incomplete information table, $X \subseteq U$ and $A \subseteq At$. Two straightforward generalized definitions for set approximations are given as follows:

Object-based definition:

$$\underline{X}^1 = \{x \mid x \in U, S_A(x) \subseteq X\}$$

$$\overline{X}^1 = \{x \mid x \in U, S_A(x) \cap X \neq \phi\}$$

Granule-based definition:

$$\underline{X}^2 = \bigcup \{S_A(x) \mid x \in U, S_A(x) \subseteq X\}$$

$$\overline{X}^2 = \bigcup \{S_A(x) \mid x \in U, S_A(x) \cap X \neq \phi\}$$

Where:

$$S_A(x) = \{y \mid y \in U, \forall a \in A \; [(f_a(x) \in f_a(y)) \vee (f_a(y) \in f_a(x)) \vee (f_a(x) \cap f_a(y) \neq \phi)]\}$$

However, above definitions are no longer dual. Let C is a covering of the universe U, two pairs of upper and lower approximations (Yao, 1998), called as tight pair and loose pair, respectively (Couso and Dubois 2011), are defined as follows:

Loose pair:

$$\overline{X}^3 = \bigcup \{B \mid B \in \mathcal{C}, B \cap X \neq \phi\}$$

$$\underline{X}^3 = ((\overline{X^c})^3)^c = \{x \mid x \in U, \forall B \in \mathcal{C}[x \in B \Rightarrow B \subseteq X]\}$$

Tight pair:

$$\underline{X}^4 = \bigcup \{B \mid B \in \mathcal{C}, B \subseteq X\}$$

$$\overline{X}^4 = ((\underline{X^c})^4)^c = \{x \mid x \in U, \forall B \in \mathcal{C}[x \in B \Rightarrow B \cap X \neq \phi]\}$$

According to the existing generalized definitions for set approximation of incomplete information tables, we can get these results as follows:

$$\underline{X}^1 = \{o_2\}, \overline{X}^1 = \{o_1, o_2, o_3, o_4\},$$
$$\underline{X}^2 = \{o_2, o_4\}, \overline{X}^2 = \{o_1, o_2, o_3, o_4\}$$

For an incomplete information table T = (U, At, V, f), many methods can be used to generate a covering of U, two of them are most commonly applied. One is based on tolerance classes of objects (Kryszkiewicz, 1998) and another is based on the values of attributes (Couso and Dubois 2011). For example, we can obtain two different coverings on the of U Table 1 as follows.

Based on tolerance classes:

$$\mathcal{C}_1 = \{\{o_1, o_3, o_4\}, \{o_2, o_4\}, \{o_1, o_3, o_4\}, \{o_1, o_2, o_3, o_4\}\}$$

Based on attribute values:

$$\mathcal{C}_2=\{\{o_2,o_4\},\{o_1,o_3,o_4\},\{o_1,o_3\}\}$$

So the following results can be easily obtained on, $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively:

$$\underline{X}^3_{\mathcal{C}_1}=\phi,\ \overline{X}^3_{\mathcal{C}_1}=\{o_1,o_2,o_3,o_4\}$$

$$\underline{X}^4_{\mathcal{C}_1}=\{o_2,o_4\},\ \overline{X}^4_{\mathcal{C}_1}=\{o_1,o_2,o_3,o_4\}$$

$$\underline{X}^3_{\mathcal{C}_2}=\{o_2\},\ \overline{X}^3_{\mathcal{C}_2}=\{o_1,o_2,o_3,o_4\}$$

$$\underline{X}^4_{\mathcal{C}_2}=\{o_2,o_4\},\ \overline{X}^4_{\mathcal{C}_2}=\{o_1,o_2,o_3,o_4\}$$

For the two measures $\underline{apr}^l(X)$ and $\overline{apr}^u(X)$, the following results have been easily obtained.

Theorem 5: $\underline{X}^1=\underline{apr}^1(X),\ \overline{X}^1=\overline{apr}^u(X)$

If $\mathcal{C}$ be a covering induced by attribute values.

Theorem 6: Couso and Dubois (2011):

$$\underline{X}^3_{\mathcal{C}}=\underline{apr}^1(X),\ \overline{X}^3_{\mathcal{C}}=\overline{apr}^1(X)$$

However, for another two measures $\underline{apr}^u(X)$ and $\overline{apr}^l(X)$, there is still no effective method to compute them based on the existing set definitions for incomplete information tables.

In fact, through examining the definitions of $\underline{apr}^u(X)$ and $\overline{apr}^l(X)$, it can been found that in order to discern whether an object x belongs X to or not, one need to compared all equivalence classes of x in all completions with X. However, the basic granules in existing set definitions, either the tolerance class of x or coverings induced by attribute value of x, are unions of some equivalence classes of x in some completions.

Example 6: Consider Table 1. For the object o_3, its tolerance class is $\{o_1, o_3, o_4\}$ (denoted as B_1) and the elements in the covering based on attribute values are $\{o_1, o_3, o_4\}$ and $\{o_1, o_3\}$ (denoted as, B_2 and B_3, respectively). While the eight equivalence classes of o_3 in T_1 to T_8 in Fig. 1 are $\{o_1, o_3\}$, $\{o_3\}$, $\{o_1, o_3\}$, $\{o_3\}$ and $\{o_1, o_3\}$, respectively. However, $o_3 \in \underline{apr}^u(X)$ is not included in $\underline{X}^4_{\mathcal{C}_1}$ or $\underline{X}^4_{\mathcal{C}_1}$ because the granules in $\mathcal{C}_1$ and $\mathcal{C}_2$ related to o_3, B_1, B_2 and B_3 are not a subset of X. Similarly, o_1 exists in $\overline{X}^4_{\mathcal{C}_1}$ and $\overline{X}^4_{\mathcal{C}_1}$ should not included in $\overline{apr}^u(X)$. Although, the granules in $\mathcal{C}_1$ and $\mathcal{C}_2$ related to o_1, $\{o_1, o_3\}$, $\{o_1, o_3, o_4\}$ and $\{o_1, o_2, o_3, o_4\}$ all have a non-empty intersection with X, the equivalence classes of o_1 in T_1, T_2, T_3 and T_4 are $\{o_1\}$ which has an empty intersection with X.

So granules in existing set definitions only provide an upper approximation for the equivalence classes in all completions which is not enough to compute the values of $\underline{apr}^u(X)$ and $\overline{apr}^l(X)$. It is necessary to construct a new method by describing the structure of equivalence classes in all completions from both upper and lower directions.

CONCLUDING REMARKS

In this study, four measures for set approximation in incomplete information tables are studied. From the view of completion, these measures have clear semantic interpretations. $\underline{apr}^l(X)$ and $\overline{apr}^l(X)$ contain the objects which certainly belong to the lower and upper approximation of X, respectively while $\underline{apr}^u(X)$ and $\overline{apr}^u(X)$ contain the objects which possibly belong to the lower and upper approximation of X, respectively. It has been illustrated and analyzed that existing definitions based on classes and covering can only be used to compute $\underline{apr}^l(X)$ and $\overline{apr}^u(X)$.

A new concept called interval granules is presented for approximating the information granules of completions from both upper and lower directions and then the new formulation to computing four measures are provided by theorems. Different with existing methods computing for every object, the new formulation is constructed only using the granulation structure of the incomplete information table, it provides a new way for set approximation in incomplete data from the view of granular computing.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China under Grant No. 61202111 and No. 61170161, the Natural Science Foundation of Shandong Province China under Grant No. ZR2011GQ001 and ZR2012FQ029, the Project Sponsored by the Scientific Research Foundation for the Returned Overseas Chinese Scholars under Grant No. 43 and the Project of Shandong Province Higher Educational Science and Technology Program under Grant No. J12LN05.

REFERENCES

Couso, I. and D. Dubois, 2011. Rough sets, coverings and incomplete information. Fundamenta Inform., 108: 223-247.

Grzymala-Busse, J.W., 2005. Incomplete Data and Generalization of Indiscernibility Relation, Definability and Approximations. In: Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, Slezak, D., G. Wang, M. Szczuka, I. Duntsch and Y. Yao (Eds.). Springer, Berlin, Germany, ISBN-13: 9783540286530, pp: 244-253.

Jarvinen, J. and J.A. Kortelainen, 2004. Note on Definability in Rough Set Theory. In: Current Issues in Data and Knowledge Engineering, De Baets, B., R. De Caluwe, G. De Tre, J. Fodor, J. Kacprzyk and S. Zadrozny (Eds.). Akademicka Oficyna Wydawnicza Exit, Warsaw, Poland, pp: 272-277.

Kryszkiewicz, M., 1998. Rough set approach to incomplete information systems. Inform. Sci., 112: 39-49.

Kryszkiewicz, M., 1999. Rules in incomplete information systems. Inform. Sci., 113: 271-292.

Lipski Jr., W., 1979. On semantic issues connected with incomplete information databases. ACM Trans. Database Syst., 4: 262-296.

Nakamura, A., 1996. A rough logic based on incomplete information and its application. Int. J. Approximate Reason., 15: 367-378.

Orlowska, E., 1998. Introduction: What You Always Wanted to Know about Rough Sets. In: Incomplete Information: Rough Set Analysis, Orlowska, E. (Ed.). Physica-Verlag, Heidelberg, Germany, pp: 1-20.

Pawlak, Z., 1981. Information systems theoretical foundations. Inform. Syst., 6: 205-218.

Pawlak, Z., 1991. Rough Sets: Theoretical Aspects of Reasoning About Data. 1st Edn., Kluwer Academic Publishers, UK., ISBN: 978-0792314721.

Pawlak, Z., J. Grzymala-Busse, R. Slowinski and W. Ziarko, 1995. Rough sets. Commun. ACM, 38: 88-95.

Yao, Y.Y. and B. Zhou, 2007. A logic language of granular computing. Proceedings of the 6th IEEE International Conference on Cognitive Informatics, August 6-8, 2007, Lake Tahoo, CA., USA., pp: 178-185.

Yao, Y.Y., 1998. Relational interpretations of neighborhood operators and rough set approximation operators. Inform. Sci., 111: 239-259.

Yao, Y.Y., 2001. Information granulation and rough set approximation. Int. J. Intell. Syst., 16: 87-104.

Yao, Y.Y., 2003. On generalizing rough set theory. Proceedings of the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, May 26-29, 2003, Chongqing, China, pp: 44-51.

Yao, Y.Y., 2007. A Note on Definability and Approximations. In: Transactions on Rough Sets VII, Peters, J.F., A. Skowron, V.W. Marek, E. Orlowska, R. Slowinski and W. Ziarko (Eds.). Springer, Berlin, Germany, pp: 274-282.