Some issues on rough-set-based approaches to data containing incomplete information

Michinori NAKATA
Josai International University
1, Gumyo, Togane, Chiba 283-8555, Japan
nakatam@ieee.org

Abstract— Methods based on rough sets to data containing incomplete information are examined for whether strong correctness criterion is satisfied or not. It is clarified that the methods proposed so far do not satisfy the strong correctness criterion. Therefore, we show a new method that satisfies the strong correctness criterion.

1 Introduction

Rough sets, proposed by Pawlak[10], give suitable methods to knowledge discovery from data. Usually, approaches based on rough sets are applied to complete data not containing uncertainty and imprecision. However, there ubiquitously exists uncertainty and imprecision in the real world[9].

Researches handling uncertainty and imprecision are actively done on the field of databases[9], but are not so much on the field of knowledge discovery. Recently, some methods directly handling incomplete information by rough sets have been proposed[6, 7, 8, 14, 15, 16]. On the other hand, a method based on possible worlds is proposed[11, 12]. This method is to apply the conventional methods based on rough sets to possible data obtained from dividing incomplete data into possible worlds, and then to aggregate the obtained results.

The former methods have to give the same results as the latter. This is called *strong correctness criterion*. Therefore, we examine whether methods proposed so far to directly handle incomplete information by rough sets satisfy the strong correctness criterion or not.

2 Approaches based on rough sets

In a table t consisting of a set $\mathcal{A}(=\{A_1,\ldots,A_n\})$ of attributes, an indiscernibility relation $IND(X)$ for a subset X of attributes is

$$IND(X) = \{(o,o')|\forall A_i \in X \;\; o[A_i] = o'[A_i]\},$$

where $o[A_i]$ and $o'[A_i]$ are attribute values of objects o and o', respectively. Suppose that the family of all equivalence classes obtained from the indiscernibility relation $IND(X)$ is denoted by $\mathcal{E}(X)$ $(=\{E(X)_1,\ldots,E(X)_m\})$, where $E(X)_i$ is an equivalence class. When every value of attributes consisting of X is exact, $E(X)_i \cap E(X)_j = \emptyset$ with $i \neq j$. Thus, the objects are uniquely partitioned. An indiscernible set $S(X)_o \in \mathcal{E}(X)$ of objects for an attribute value $o[X]$ of an object o is

$$S(X)_o = \{o'|\forall A_i \in X \;\; o[A_i] = o'[A_i]\}.$$

When two objects contain incomplete information for some attributes, they does not always take the same actual value, even if they have the same expression. To what degree the two objects take the same actual value is obtained. The degree is an indiscernibility degree of the two objects. The above expression is replaced as follows:

$$\begin{aligned} IND(X) = \{&(o,o')(EQ(o[X],o'[X]))| \\ &\forall A_i \in X \;\; EQ(o[A_i],o'[A_i]) \neq 0\} \cup_{o\in t} \{(o,o)(1)\}, \\ S(X)_o = \{&o'(EQ(o[X],o'[X]))| \\ &\forall A_i \in X \;\; EQ(o[A_i],o'[A_i]) \neq 0\} \cup \{o(1)\}, \end{aligned}$$

where $EQ(o[X],o'[X])$ is an indiscernibility degree of $o[X]$ with $o'[X]$, which is contained in the interval $[0,1]$, and

$$EQ(o[X],o'[X]) = \bigotimes_{A_i\in X} EQ(o[A_i],o'[A_i]).$$

where the operator $\bigotimes$ depends on the properties of imprecise attribute values. When the imprecise attribute values are expressed with probability distributions, the operator is product. On the other hand, when the imprecise attribute values are expressed with possibility distributions, the operator is min.

The lower approximation and the upper approximation of $IND(Y)$ by $IND(X)$ are

$$\begin{aligned} \underline{IND(Y,X)} &= \cup_{i,j}\{E(X)_i|E(X)_i \subseteq E(Y)_j\}, \\ \overline{IND(Y,X)} &= \cup_{i,j}\{E(X)_i|E(X)_i \cap E(Y)_j \neq \emptyset\}. \end{aligned}$$

The lower approximation $\underline{IND(Y,X,o)}$ and the upper approximation $\overline{IND(Y,X,o)}$ of $S(Y)_o$ by $IND(X)$ are expressed by means of using $S(X)_{o'}$ as follows:

$$\begin{aligned} \underline{IND(Y,X,o)} &= \cup_{o'}\{S(X)_{o'}|S(X)_{o'} \subseteq S(Y)_o\}, \\ \overline{IND(Y,X,o)} &= \cup_{o'}\{S(X)_{o'}|S(X)_{o'} \cap S(Y)_o \neq \emptyset\}. \end{aligned}$$

By using them,

$$\begin{aligned} \underline{IND(Y,X)} &= \cup_o \underline{IND(Y,X,o)}, \\ \overline{IND(Y,X)} &= \cup_o \overline{IND(Y,X,o)}. \end{aligned}$$

A measure called *quality of approximation* to estimate to what extent the approximation is correct is used. This measure means to what degree a dependency of attributes Y to attributes X holds[10]; namely, to what degree a table t satisfies a dependency $X \Rightarrow Y$. The degree is

$$\kappa(X \Rightarrow Y)_t = |\underline{IND(Y,X)}|/|t|,$$

where $|t|$ is the cardinality of a table t; namely, the total number of objects in the table t. This degree can be

also calculated by means of summing a degree to which each object o in the table t satisfies $X \Rightarrow Y$. The degree $\kappa(X \Rightarrow Y)_o$ to which an object o satisfies $X \Rightarrow Y$ is

$$\kappa(X \Rightarrow Y)_o = \kappa(S(X)_o \subseteq S(Y)_o).$$

where $\kappa(S(X)_o \subseteq S(Y)_o)$ is a inclusion degree of $S(X)_o$ to $S(Y)_o$.
When all the values of the attributes in X and Y are exact, this degree is 0 or 1; namely,

$$\kappa(X \Rightarrow Y)_o = \begin{cases} 1 & S(X)_o \subseteq S(Y)_o, \\ 0 & otherwise. \end{cases}$$

When an attribute takes an imprecise value, some authors propose to calculate the inclusion degree $\kappa(S(X)_o \subseteq S(Y)_o)$ by using implications[14, 15, 16]. Anyway, when this degree is obtained,

$$\kappa(X \Rightarrow Y)_t = \Sigma_{o \in t} \kappa(X \Rightarrow Y)_o / |t|.$$

3 Methods handling incomplete information

Some pioneering work is done by Slowiński and Stefanowski[13] and Grzymala[3] to handle incomplete information by using rough sets. When we handle a table containing incomplete information, obtained equivalence classes overlap each other; namely, $E(X)_i \cap E(X)_j \neq \emptyset$ with $i \neq j$. Recently, several investigations have been made on this topic.

Kryszkiewicz applies rough sets to data containing incomplete information by interpreting a missing value expressing unknown as indiscernible with an arbitrary value[6, 7, 8]. An indiscernibility relation under the viewpoint is called a tolerance relation. In this approach an object in which some attribute values are missing values is indiscernible with every object for the attributes. The tolerance relation is reflexive, symmetric, and transitive. Slowiński and Tsoukiàs apply rough sets to a table containing incomplete information by making an indescernibility relation from the viewpoint that an object with an exact attribute value is similar to another object with the attribute value being missing, but the converse is not so[14, 16]. They call the indiscernibility relation a similarity relation. The similarity relation is reflexive and transitive, but not symmetric. The above two approaches handle incomplete information by deriving an indiscernibility relation from giving a missing value an interpretation and then by applying the conventional method of rough sets to the indiscernibility relation.

Furthermore, Stefanowski and Tsoukiás make an indiscernibility relation by introducing the probabilistic degree that two objects cannot be discerned under the premise that an attribute can equally take an arbitrary value included in the corresponding domain when the attribute value is a missing value[14, 15, 16]. The indiscernibility relation is called a valued tolerance relation and each element is a value in the interval $[0, 1]$. In the approach, they use Reichenbach implication in calculating a inclusion degree of two indiscernible sets.

Active researches are made into imperfect information in the field of databases[9]. Some extensions have to be made to operators in order to directly deal with imperferct information. In order to check whether the extended operators create correct results in query processing or not, the strong correctness criterion are used[1, 4, 5, 17]. In rough-set-based approaches the strong correctness criterion is checked as follows:

- To derive a set of possible tables from a table containing incomplete information.
- To aggregate the results obtained from applying the conventional operators to each possible table.
- To compare the aggregated results with ones obtained from directly applying the extended operator to the table.

Here, a possible table derived from a table is that of each missing value in the table being replaced by an element containing in the corresponding domain. When two results coincide, the strong correctness criterion is satisfied. This is formulated as follows:
Suppose that $rep(t)$ is a set of possible tables derived from a table t containing incomplete information. Let q' be the conventional operator applied to $rep(t)$, which corresponds to an extended operator q directly applied to a table t. The two results is the same; namely,

$$q(t) = q'(rep(t)).$$

When this is valid, the extended operator q gives correct results. In the next section, we examine the correctness of methods proposed so far according to this criterion through calculating a degree of dependency.

4 Comparative studies on methods handling incomplete information

4.1 Tables and possible tables

We suppose that table t containing incomplete information is given as follows:

t

O	A	B
1	x	a
2	x	a
3	@	b
4	@	a

Here, attribute O denotes the object identity and @ denotes a missing value that means *unknown*. Possible tables obtained from table t are those that the missing value @ is replaced by an element consisting of the corresponding domain. Suppose that domains $dom(A)$ and $dom(B)$ of attributes A and B are $\{x, y\}$ and $\{a, b\}$, respectively. The following four possible tables are derived:

$Poss(t)_1$

O	A	B
1	x	a
2	x	a
3	x	b
4	x	a

$Poss(t)_2$

O	A	B
1	x	a
2	x	a
3	x	b
4	y	a

$Poss(t)_3$

O	A	B
1	x	a
2	x	a
3	y	b
4	x	a

$Poss(t)_4$

O	A	B
1	x	a
2	x	a
3	y	b
4	y	a

We calculate a degree $\kappa(A \Rightarrow B)$ of a dependency $A \Rightarrow B$ in these possible tables. There exists no object that contributes to $A \Rightarrow B$ in $Poss(t)_1$. Only the fourth object contributes to $A \Rightarrow B$ in $Poss(t)_2$. All the objects contribute to $A \Rightarrow B$ in $Poss(t)_3$. The first and second objects contribute to $A \Rightarrow B$ in $Poss(t)_4$. Thus, the contributions of the objects to $A \Rightarrow B$ are as

follows:

$$\begin{aligned}\kappa(A \Rightarrow B)_{Poss(t)_1} &= 0,\\ \kappa(A \Rightarrow B)_{Poss(t)_2} &= 1/4,\\ \kappa(A \Rightarrow B)_{Poss(t)_3} &= 1,\\ \kappa(A \Rightarrow B)_{Poss(t)_4} &= 1/2.\end{aligned}$$

One of the possible tables is the actual table, but it is unknown which table is the actual one. In this point, they can be regarded as probabilistically equal. Therefore, a degree $\kappa(A \Rightarrow B)_t$ of a dependency $A \Rightarrow B$ is the average of the degrees in each possible table; namely,

$$\begin{aligned}\kappa(A \Rightarrow B)_t &= \sum_{i=1,4} \kappa(A \Rightarrow B)_{Poss(t)_i}/4\\ &= (0 + 1/4 + 1 + 1/2)/4 = 7/16.\end{aligned}$$

Contributions of each object o_i to this value of $A \Rightarrow B$ are as follows:

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_1} &= 1/2,\\ \kappa(A \Rightarrow B)_{o_2} &= 1/2,\\ \kappa(A \Rightarrow B)_{o_3} &= 1/4,\\ \kappa(A \Rightarrow B)_{o_4} &= 1/2.\end{aligned}$$

We examine whether the same value $\kappa(A \Rightarrow B)_{o_i}$ for each object o_i is obtained or not by means of using the methods proposed so far in the following subsections.

4.2 Methods by tolerance relations

Kryszkiewicz[6, 7, 8] regards a missing value as indiscernible with an arbitrary value contained in the corresponding domain. This corresponds to that two objects are indiscernible when there is the probability that an object is equal to another object. An indiscernibility relation is symmetric under the semantics. Indiscernibility relations $IND(A)$ and $IND(B)$ for attributes A and B in table t are, respectively,

$$IND(A) = \begin{pmatrix} 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 \end{pmatrix},$$

$$IND(B) = \begin{pmatrix} 1 & 1 & 0 & 1\\ 1 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Indiscernible sets of the objects for attribute A are,

$$\begin{aligned}S(A)_{o_1} &= \{o_1, o_2, o_3, o_4\},\\ S(A)_{o_2} &= \{o_1, o_2, o_3, o_4\},\\ S(A)_{o_3} &= \{o_1, o_2, o_3, o_4\},\\ S(A)_{o_4} &= \{o_1, o_2, o_3, o_4\}.\end{aligned}$$

Indiscernible sets of the objects for attribute B are,

$$\begin{aligned}S(B)_{o_1} &= \{o_1, o_2, o_4\},\\ S(B)_{t_2} &= \{o_1, o_2, o_4\},\\ S(B)_{t_3} &= \{o_3\},\\ S(B)_{t_4} &= \{o_1, o_2, o_4\}.\end{aligned}$$

The contributions of the objects are,

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_1} &= \kappa(S(A)_{o_1} \subseteq S(B)_{o_1}) = 0,\\ \kappa(A \Rightarrow B)_{o_2} &= \kappa(S(A)_{o_2} \subseteq S(B)_{o_2}) = 0,\\ \kappa(A \Rightarrow B)_{o_3} &= \kappa(S(A)_{o_3} \subseteq S(B)_{o_3}) = 0,\\ \kappa(A \Rightarrow B)_{o_4} &= \kappa(S(A)_{o_4} \subseteq S(B)_{o_4}) = 0.\end{aligned}$$

Thus, the degree of a dependency $A \Rightarrow B$ is,

$$\kappa(A \Rightarrow B)_t = (0 + 0 + 0 + 0)/4 = 0.$$

4.3 Methods by similarity relations

Stefanowski and Tsoukiàs[14, 16] make an indiscernibility relation under the interpretation that an exact value is similar to a missing value, but the missing value is not similar to every exact value, and the missing values are similar to each other. The interpretation corresponds to that the probability from exact values is absolutely accepted, but the probability from missing values is not so at all. Under this interpretation obtained indiscernibility relations are not symmetric. An indiscernibility relation $IND(A)$ for an attribute A in table t is

$$IND(A) = \begin{pmatrix} 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1\\ 0 & 0 & 1 & 1\\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

$IND(B)$ is unchanged. Indiscernible sets of the objects for attribute A are,

$$\begin{aligned}S(A)_{o_1} &= \{o_1, o_2, o_3, o_4\},\\ S(A)_{o_2} &= \{o_1, o_2, o_3, o_4\},\\ S(A)_{o_3} &= \{o_3, o_4\},\\ S(A)_{o_4} &= \{o_3, o_4\}.\end{aligned}$$

The indiscernible sets of the objects for attribute B are unchanged. The contributions of the objects are,

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_1} &= \kappa(S(A)_{o_1} \subseteq S(B)_{o_1}) = 0,\\ \kappa(A \Rightarrow B)_{o_2} &= \kappa(S(A)_{o_2} \subseteq S(B)_{o_2}) = 0,\\ \kappa(A \Rightarrow B)_{o_3} &= \kappa(S(A)_{o_3} \subseteq S(B)_{o_3}) = 0,\\ \kappa(A \Rightarrow B)_{o_4} &= \kappa(S(A)_{o_4} \subseteq S(B)_{o_4}) = 0.\end{aligned}$$

Thus, the degree of a dependency $A \Rightarrow B$ is,

$$\kappa(A \Rightarrow B)_t = (0 + 0 + 0 + 0)/4 = 0.$$

4.4 Methods by valued tolerance relations

Stefanowski and Tsoukiàs[14, 15, 16] take the interpretation that when an attribute value is a missing value, the actual value is one of elements in the domain of the attribute and which element is the actual value does not depend on an specified element; namely, each element has the same probability for that the element is the actual value. Under this interpretation an obtained indiscernibility relation is symmetric, but consists of values in the interval $[0, 1]$. An indiscernibility relation $IND(A)$ for attribute A in table t is

$$IND(A) = \begin{pmatrix} 1 & 1 & 1/2 & 1/2\\ 1 & 1 & 1/2 & 1/2\\ 1/2 & 1/2 & 1 & 1/2\\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

$IND(B)$ is unchanged. The indiscernible sets of the objects for attribute A are,

$$\begin{aligned}S(A)_{o_1} &= \{o_1(1), o_2(1), o_3(1/2), o_4(1/2)\},\\ S(A)_{o_2} &= \{o_1(1), o_2(1), o_3(1/2), o_4(1/2)\},\\ S(A)_{o_3} &= \{o_1(1/2), o_2(1/2), o_3(1), o_4(1/2)\},\\ S(A)_{o_4} &= \{o_1(1/2), o_2(1/2), o_3(1/2), o_4(1)\}.\end{aligned}$$

An indiscernible set of the objects for attribute B is unchanged.

Suppose that an object o belongs to sets S and S' with probabilistic degrees $P_{t,S}$ and $P_{t,S'}$, respectively. The degree $\kappa(S \subseteq S')$ that the set S is included in another set S' is,

$$\begin{aligned}\kappa(S \subseteq S') &= \prod_{o \in S} \kappa(o \in S \rightarrow o \in S') \\ &= \prod_{o \in S} (1 - P_{o,S} + P_{o,S} \times P_{o,S'}).\end{aligned}$$

In this formula, the inclusion degree of two sets is calculated by means of using Reichenbach implication ($u \rightarrow v = 1 - u + u \times v$). Now, S and S' are $S(A)_{o_i}$ and $S(B)_{o_i}$, respectively, and $P_{o_i,S(A)_{o_i}}$ and $P_{o_i,S(B)_{o_i}}$ are $EQ(o_i[A], o_j[A])$ and $EQ(o_i[B], o_j[B])$, respectively. Thus, the contributions of the objects are as follows:

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_1} &= \kappa(S(A)_{o_1} \subseteq S(B)_{o_1}) \\ &= 1 \times 1 \times (1 - 1/2 + 1/2 \times 0) \\ &\quad \times (1 - 1/2 + 1/2 \times 1) = 1/2.\end{aligned}$$

Similarly,

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_2} &= \kappa(S(A)_{o_2} \subseteq S(B)_{o_2}) = 1/2, \\ \kappa(A \Rightarrow B)_{o_3} &= \kappa(S(A)_{o_3} \subseteq S(B)_{o_3}) \\ &= (1 - 1/2 + 1/2 \times 0) \\ &\quad \times (1 - 1/2 + 1/2 \times 0) \\ &\quad \times 1 \times (1 - 1/2 + 1/2 \times 0) = 1/8, \\ \kappa(A \Rightarrow B)_{o_4} &= \kappa(S(A)_{o_4} \subseteq S(B)_{o_4}) \\ &= (1 - 1/2 + 1/2 \times 1) \\ &\quad \times (1 - 1/2 + 1/2 \times 1) \\ &\quad \times (1 - 1/2 + 1/2 \times 0) \times 1 = 1/2.\end{aligned}$$

Thus, the degree of dependency $A \Rightarrow B$ is

$$\kappa(A \Rightarrow B)_t = (1/2 + 1/2 + 1/8 + 1/2)/4 = 13/32.$$

The obtained values $\kappa(A \Rightarrow B)_{o_3}$ and $\kappa(A \Rightarrow B)_t$ are not equal to ones obtained from possible tables.

5 Methods satisfying strong correctness criterion

Why cannot the methods proposed so far satisfy the strong correctness criterion? The methods by tolerance relations and similarity relations deal with incomplete information not strictly, but approximately under some interpretations of missing values. Thus, these methods cannot satisfy the strong correctness criterion. On the other hand, the method by valued tolerance relations, which is proposed by Stefanowski and Tsoukiàs[14, 15, 16], strictly handles incomplete information. Why cannot this method by Stefanowski and Tsoukiàs satisfy the strong correctness criterion? Stefanowski and Tsoukiàs calculates the inclusion degree of two sets to which each element belongs with a probabilistic degree as follows:

- To calculate to what probabilistic degree every element belonging to a set also belongs to another set by using Reichenbach implication.
- To multiply the obtained degrees together.

The process shows that the total inclusion degree is obtained through aggregating the inclusion degrees for every element. This is valid under the condition that an inclusion degree for an element is determined independently of another element. Is this valid in the present situation?

In the previous section, the degree $\kappa(A \Rightarrow B)_{o_3}$ of a dependency $A \Rightarrow B$ for the third object o_3 does not coincide with the degree obtained from using possible tables. This is due to not taking into account the fact that when the third object is indiscernible with the first for attribute A, simultaneously it is indiscernible with the second; namely, the first and the second objects have to be dealt with together. This strongly suggests that the condition described above is not valid in the present situation.

Furthermore in order to examine this, we go into issues for using implication operators. In Reichenbach implication, a probability $Prob(a \rightarrow b)$ of a sentence $a \rightarrow b$ is equal to $1 - Prob(a) + Prob(a) \times Prob(b)$, when probabilities that a sentence a is valid and a sentence b is valid are given with $Prob(a)$ and $Prob(b)$, respectively. This comes from the followings: when the sentence a is valid with a probability $Prob(a)$, $a \rightarrow b$ is valid with $Prob(a) \times Prob(b)$; when a is invalid, $a \rightarrow b$ is valid regardlessly of b; namely, $a \rightarrow b$ is valid with $1 - Prob(a)$ when a is invalid. Thus, $Prob(a \rightarrow b)$ is $1 - Prob(a) + Prob(a) \times Prob(b)$ generally. Is it correct that $a \rightarrow b$ is valid regardlessly of b, when a is invalid in the present situation?

The fact that an object o_i belongs to $S(X)_{o_j}$ with a probabilistic degree $EQ(o_j[X], o_i[X])$ means that o_j is equal to o_i for a set X of attributes with the degree $EQ(o_j[X], o_i[X])$. In the method by Stefanowski and Tsoukiàs using an implication, Reichenbach implication, The degree that $o_i \in S(X)_{o_j} \rightarrow o_i \in S(Y)_{o_j}$ is valid is $1 - EQ(o_j[X], o_i[X]) + EQ(o_j[X], o_i[X]) \times EQ(o_j[Y], o_i[Y])$, when o_j is equal to o_i for a set X of attributes with a probabilistic degree $EQ(o_j[X], o_i[X])$ and o_j is equal to o_i for a set Y of attributes with a probabilistic degree $EQ(o_j[Y], o_i[Y])$. This calculation means that the dependency is valid regardlessly of a set Y of attributes when o_j is not equal to o_i for a set X of attributes with a probabilistic degree $1 - EQ(o_j[X], o_i[X])$. However, this is not correct if there exists another object o_k that is equal to o_j with a probabilistic degree for a set X of attributes, but that is not to o_i at all for X, as is shown in the following example.

We suppose that table t' containing incomplete information is given as follows:

t'

O	A	B
1	x	a
2	y	a
3	@	b
4	@	a

In table t' only the attribute value $o_2[A]$ is different from table t in the previous example. Notice there exists another object o_2 that is equal to o_3 for an attribute A, but that is not equal to o_1 for A. Results obtained from using possible tables are:

$$\begin{aligned}\kappa(A \Rightarrow B)_{o_1} &= 1/2, \\ \kappa(A \Rightarrow B)_{o_2} &= 1/2, \\ \kappa(A \Rightarrow B)_{o_3} &= 0,\end{aligned}$$

$$\kappa(A \Rightarrow B)_{o_4} = 1/2.$$

An indiscernibility relations $IND(A)$ for an attribute A in table t' is as follows:

$$IND(A) = \begin{pmatrix} 1 & 0 & 1/2 & 1/2 \\ 0 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1 & 1/2 \\ 1/2 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

$IND(B)$ is the same as in table t. The indiscernible sets of the objects for attribute A are,

$$\begin{aligned} S(A)_{o_1} &= \{o_1(1), o_3(1/2), o_4(1/2)\}, \\ S(A)_{o_2} &= \{o_2(1), o_3(1/2), o_4(1/2)\}, \\ S(A)_{o_3} &= \{o_1(1/2), o_2(1/2), o_3(1), o_4(1/2)\}, \\ S(A)_{o_4} &= \{o_1(1/2), o_2(1/2), o_3(1/2), o_4(1)\}. \end{aligned}$$

The indiscernible sets of the objects for attribute B are the same as in table t. We focus on the contribution of the third object o_3.

$$\begin{aligned} \kappa(A \Rightarrow B)_{o_3} &= \kappa(S(A)_{o_3} \subseteq S(B)_{o_3}) \\ &= (1 - 1/2 + 1/2 \times 0) \\ &\quad \times (1 - 1/2 + 1/2 \times 0) \times 1 \\ &\quad \times (1 - 1/2 + 1/2 \times 0) = 1/8. \end{aligned}$$

In the example, the contribution of the fact that o_3 is equal to o_1 for an attribute A with a probabilistic degree $EQ(o_3[A], o_1[A])$ is calculated by means of $1 - EQ(o_3[A], o_1[A]) + EQ(o_3[A], o_1[A]) \times EQ(o_3[B], o_1[B]$. However, the fact that o_3 is not equal to o_1 for an attribute A means that o_3 is equal to another object o_2 for an attribute A. Thus, when o_3 is not equal to o_1 for an attribute A with a probabilistic degree $1 - EQ(o_3[A], o_1[A])$, o_3 has to be unconditionally equal to o_2 for an attribute B. However, this is not valid in table t'. In other words, we cannot separate the two facts that o_3 is equal to o_1 for an attribute A with a probabilistic degree $EQ(o_3[A], o_1[A])$ and o_3 is equal to o_2 for an attribute A with a probabilistic degree $EQ(o_3[A], o_2[A])$. These two facts link with each other disjunctively. We simultaneously have to deal with the two facts.

From considering the above viewpoint, we propose a new method for calculating $\kappa(X \Rightarrow Y)_{o_i}$.

Let $ps(X)_{o_i,l}$ be an element of the power set $PS(X)_{o_i}$ of $S(X)_{o_i} \backslash o_i$.

$$\begin{aligned} \kappa(X \Rightarrow Y)_{o_i} = \sum_l \kappa(&\wedge_{o' \in ps(X)_{o_i,l}} (o_i[X] = o'[X]) \\ &\wedge_{o' \notin ps(X)_{o_i,l}} (o_i[X] \neq o'[X])) \\ &\times \kappa(\wedge_{o' \in ps(X)_{o_i,l}} (o_i[Y] = o'[Y]))), \end{aligned}$$

where $\kappa(f)$ is the probabilistic degree that a formula f is valid and $\kappa(f) = 1$ when there is no f.

In this formula, all the elements in an equivalence class are simultaneously handled. We recalculate the degree of dependency $A \Rightarrow B$ in table t. For the object o_1,

$$S(A)_{o_1} \backslash o_1 = \{o_2(1), o_3(1/2), o_4(1/2)\}.$$

For the power set $PS(X)_{o_1}$ of $S(A)_{o_1} \backslash o_1$,

$$\begin{aligned} PS(X)_{o_1} = \{&\emptyset, o_2(1), o_3(1/2), o_4(1/2), \{o_2(1), o_3(1/2)\}, \\ &\{o_2(1), o_4(1/2)\}, \{o_3(1/2), o_4(1/2)\}, \\ &\{o_2(1), o_3(1/2), o_4(1/2)\}\}. \end{aligned}$$

We omit the case of elements containing o_3 for the power set $SP(X)_{o_1}$, because $\kappa(o_1[B] = o_3[B]) = 0$. For the element $\emptyset$,

$$\kappa(o_1[A] \neq o_2[A] \wedge o_1[A] \neq o_3[A] \wedge o_1[A] \neq o_4[A]) = 0.$$

For the element $o_2(1)$,

$$\begin{aligned} &\kappa(o_1[A] = o_2[A] \wedge o_1[A] \neq o_3[A] \wedge o_1[A] \neq o_4[A]) = 1/4, \\ &\kappa(o_1[B] = o_2[B]) = 1. \end{aligned}$$

For the element $o_4(1/2)$,

$$\kappa(o_1[A] \neq o_2[A] \wedge o_1[A] \neq o_3[A] \wedge o_1[A] = o_4[A]) = 0.$$

For the element $\{o_2(1), o_4(1/2)\}$,

$$\begin{aligned} &\kappa(o_1[A] = o_2[A] \wedge o_1[A] \neq o_3[A] \wedge o_1[A] = o_4[A]) = 1/4, \\ &\kappa(o_1[B] = o_2[B] \wedge o_1[B] = o_4[B]) = 1. \end{aligned}$$

Thus,

$$\begin{aligned} &\kappa(X \Rightarrow Y)_{o_1} = \\ &\quad 0 + 1/4 \times 1 + 0 + 0 + 0 + 1/4 \times 1 + 0 + 0 = 1/2. \end{aligned}$$

Similarly,

$$\kappa(X \Rightarrow Y)_{o_2} = 1/2.$$

For the object o_3,

$$S(A)_{o_3} \backslash o_3 = \{o_1(1/2), o_2(1/2), o_4(1/2)\}.$$

For the power set $PS(X)_{o_3}$ of $S(A)_{o_3} \backslash o_3$,

$$\begin{aligned} &PS(X)_{o_3} = \\ &\quad \{\emptyset, o_1(1/2), o_2(1/2), o_4(1/2), \{o_1(1/2), o_2(1/2)\}, \\ &\qquad \{o_1(1/2), o_4(1/2)\}, \{o_2(1/2), o_4(1/2)\}, \\ &\qquad\quad \{o_1(1/2), o_2(1/2), o_4(1/2)\}\}. \end{aligned}$$

We calculate only for the element $\emptyset$, because $\kappa(o_3[B] = o_i[B]) = 0$ for $i = 1, 2$, and 4. For the element $\emptyset$,

$$\kappa(o_3[A] \neq o_1[A] \wedge o_3[A] \neq o_2[A] \wedge o_3[A] \neq o_4[A]) = 1/4.$$

Thus,

$$\begin{aligned} \kappa(X \Rightarrow Y)_{o_3} &= 1/4 \times 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\ &= 1/4. \end{aligned}$$

For the object o_4,

$$S(A)_{o_4} \backslash o_4 = \{o_1(1/2), o_2(1/2), o_3(1/2)\}.$$

For the power set $PS(X)_{o_4}$ of $S(A)_{o_4} \backslash o_4$,

$$\begin{aligned} &PS(X)_{o_4} = \\ &\quad \{\emptyset, o_1(1/2), o_2(1/2), o_3(1/2), \{o_1(1/2), o_2(1/2)\}, \\ &\qquad \{o_1(1/2), o_3(1/2)\}, \{o_2(1/2), o_3(1/2)\}, \\ &\qquad\quad \{o_1(1/2), o_2(1/2), o_3(1/2)\}\}. \end{aligned}$$

We omit the case of elements containing o_3 for the power set $SP(X)_{o_1}$, because $\kappa(o_4[B] = o_3[B]) = 0$. For the element $\emptyset$,

$$\kappa(o_4[A] \neq o_1[A] \wedge o_4[A] \neq o_2[A] \wedge o_4[A] \neq o_3[A]) = 1/4.$$

For the element $o_1(1/2)$,

$$\kappa(o_4[A] = o_1[A] \wedge o_4[A] \neq o_2[A] \wedge o_4[A] \neq o_3[A]) = 0.$$

For the element $o_2(1/2)$,

$$\kappa(o_4[A] \neq o_1[A] \wedge o_4[A] = o_2[A] \wedge o_4[A] \neq o_3[A]) = 0.$$

For the element $\{o_1(1/2), o_2(1/2)\}$,

$$\kappa(o_4[A] = o_1[A] \wedge o_4[A] = o_2[A] \wedge o_4[A] \neq o_3[A]) = 1/4,$$
$$\kappa(o_4[B] = o_1[B] \wedge o_4[B] = o_2[B]) = 1.$$

Thus,

$$\kappa(X \Rightarrow Y)_{o_4} = 1/4 \times 1 + 0 + 0 + 0 + 1/4 \times 1 + 0 + 0 + 0 = 1/2.$$

The obtained results coincide with ones from possible tables.

Proposition
This method satisfies the strong correctness criterion.

6 Conclusions

We examine methods by tolerance relations, by similarity relations, and by valued tolerance relations for calculating a degree of dependency, a measure of *quality of approximation*, in tables containing incomplete information for whether they satisfy the strong correctness criterion. The methods by tolerance relations and by similarity relations do not strictly, but approximately handle incomplete information. The method by valued tolerance relations strictly handles incomplete information, but does not simultaneously handle all the elements in an equivalence class. By the example, it is shown that these methods do not satisfy the strong correctness criterion. Therefore, we have proposed a new method in which all the elements in an equivalence class are simultaneously dealt with. This method satisfies the strong correctness criterion.

Acknowledgment

This research has partially been supported by the Grant-in-Aid for Scientific Research (B), Japan Society for the Promotion of Science, No. 14380171.

References

[1] Abiteboul, S., Hull, R., and Vianu, V. [1995] Fundations of Databases, Addison-Wesley Publishing Company, 1995.

[2] Gediga, G. and Düntsch, I. [2001]Rough Approximation Quality Revusited, Artificial Intelligence, **132**, 219-234.

[3] Grzymala-Busse, J. W. [1991]On the Unknown Attribute Values in Learning from Examples, in Ras, M. Zemankova, (eds.), Methodology for Intelligent Systems, ISMIS '91, Lecture Notes in Artificial Intelligence 542, Springer-Verlag, 368-377.

[4] Imielinski, T. [1989]Incomplete Information in Logical Databases, Data Engineering, **12**, 93-104.

[5] Imielinski, T. and Lipski, W. [1984]Incomplete Information in Relational Databases, Journal of the ACM, **31**:4, 761-791.

[6] Kryszkiewicz, M. [1998]Properties of Incomplete Information Systems in the framework of Rough Sets, in L. Polkowski and A. Skowron, (ed.), Rough Set in Knowledge Discovery 1: Methodology and Applications, Studies in Fuzziness and Soft Computing 18, Physica Verlag, 422-450.

[7] Kryszkiewicz, M. [1999]Rules in Incomplete Information Systems, Information Sciences, **113**, 271-292.

[8] Kryszkiewicz, M. and Rybiński, H. [2000]Data Mining in Incomplete Information Systems from Rough Set Perspective, in L. Polkowski, S. Tsumoto, and T. Y. Lin, (eds.), Rough Set Methods and Applications, Studies in Fuzziness and Soft Computing 56, Physica Verlag, 568-580.

[9] Parsons, S. [1996] Current Approaches to Handling Imperfect Information in Data and Knowledge Bases, IEEE Transactions on Knowledge and Data Engineering, **8**3, 353-372.

[10] Pawlak, Z. [1991] Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers 1991.

[11] Sakai, H. [1998]Some Issues on Nondeterministic Knowledge Bases with Incomplete Information, in: Proceedings of RSCTC'98, Polkowski, L. and Skowron, A., eds., Lecture Notes in Artificial Intelligence Vol. 1424, Springer-Verlag 1998, pp. 424-431.

[12] Sakai, H. [1999]An Algorithm for Finding Equivalence Relations from Table Non-deterministic Information, in N. Zhong, A. Skowron, S. Ohsuga, (eds.), New Directions in Rough Sets, Data Mining and Granular-Soft Computing, Lecture Notes in Artificial Intelligence 1711, pp. 64-72.

[13] Słowiński, R. and Stefanowski, J. [1989]Rough Classification in Incomplete Information Systems, Mathematical and Computer Modelling, **12**:10/11, 1347-1357.

[14] Stefanowski, J. and Tsoukiàs, A. [1999]On the Extension of Rough Sets under Incomplete Information, in N. Zhong, A. Skowron, S. Ohsuga, (eds.), New Directions in Rough Sets, Data Mining and Granular-Soft Computing, Lecture Notes in Artificial Intelligence 1711, pp. 73-81.

[15] Stefanowski, J. and Tsoukiàs, A. [2000]Valued Tolerance and Decision Rules, in W. Ziarko and Y. Yao, (eds.), Rough Sets and Current Trends in Computing, Lecture Notes in Artificial Intelligence 2005, Springer-Verlag, pp. 212-219.

[16] Stefanowski, J. and Tsoukiàs, A. [2001]Incomplete Information Tables and Rough Classification, Computational Intelligence, **17**:3, 545-566.

[17] Zimányi, E. and Pirotte, A. [1997] Imperfect Information in Relational Databases, in Uncertainty Management in Information Systems: From Needs to Solutions, A. Motro and P. Smets, eds., Kluwer Academic Publishers, 1997, pp. 35-87.