

Rough Sets and Rule Induction in Imperfect Information Systems

Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara
Department of Management and Information Systems Science
Nagaoka University of Technology
Nagaoka, Japan

ABSTRACT

The original rough set theory deals with precise and complete data, while real applications frequently contain imperfect information. A typical imperfect data studied in rough set research is the missing values. Though there are many ideas proposed to solve the issue in the literature, the paper adopts a probabilistic approach, because it can incorporate other types of imperfect data including imprecise and uncertain values in a single approach. The paper first discusses probabilities of attribute values assuming different type of attributes in real applications, and proposes a generalized method of probability of matching. This probability is then used to define valued tolerance/similarity relations and to develop new rough set models based on the valued tolerance/similarity relations. An algorithm for deriving decision rules based on the rough set models is also studied and proposed.

Keywords:

Imperfect Information Systems, Probability of Matching, Approximation Space, Rough Sets, Decision Rules

1. INTRODUCTION

Classical Rough set theory [21, 22] provides a mathematical tool to analyse databases under objects description. The original rough sets approach presupposes that all objects in an information system have precise and complete attribute values. Problems arise when information systems contain imperfect data, which occasionally happens in the real world. Controversial rough set research mostly considers that imperfect data in information systems comes from missing values [4, 5, 6, 8, 11, 12, 23, 24, 25, 26]. An information system with missing values is called incomplete information system [11, 12]. In incomplete information systems, Table 1 for example [8], objects may contain several unknown attribute values. Unknown values are denoted by special symbol “*”.

In studies of rough sets in incomplete information systems, probabilistic solutions have been introduced based on the possibility of “missing value” [4, 17, 18, 19, 23, 24]. Among them, some approaches [4, 23] suppose a priori assumption that there exists a uniform probability distribution on every attribute domain and compute valued tolerance (or similarity) classes based on the joint probability distribution. The paper aims at defining a general method of determining the probability that two object may be

Table 1. An example of a dataset with missing values

Cases	Temperature	Headache	Nausea	Flu
x_1	high	*	no	yes
x_2	very-high	yes	yes	yes
x_3	*	no	no	no
x_4	high	yes	yes	yes
x_5	high	*	yes	no
x_6	normal	yes	no	no
x_7	normal	no	yes	no
x_8	*	yes	*	yes

tolerant of (similar to) each other on an attribute. The probability of matching will be defined based on the probability that two objects may take the same values on an attribute in the dataset.

Besides the missing values, there are many reasons why imperfect data are produced in datasets [16]. *Imprecision* is another type of possible imperfect data. Stored information is imprecise when it denotes a set of possible values and the real value is one of elements of this set. Specific kinds of imprecise information include disjunctive information (e.g., John’s age is either 31 or 32), negative information (e.g., John’s age is not 30), range information (e.g., John’s age is between 30 and 35 or John’s age is over 30), and information with error margins (e.g., John’s age is 30 ± 1 years).

One more possible type of imperfect data is *Uncertainty*[16]. Whereas the statement “John’s age is either 31 or 32” is in the form of imprecision, the statement “John is probably 32” or “John’s age is 32 with confidence 0.6” denotes uncertainty. Both imprecise and uncertain values can be represented by probabilistic data [2]. For example, Toyota might have demographic information indicating that customers living in a certain region are likely to purchase a Corolla with probability 0.7 or a Celica with probability 0.3. Table 2 illustrates an information system [1] with probabilistic data.

The last of imperfect data listed in [16] is *Error*. Stored information is erroneous when it is different from the true information. If errors in the given information are identified, they can be removed and the rest is treated as information with missing values. However, in the case where errors are unidentified, the reliability of the whole information will be lost. Approximations in the rough set theory are derived from the information available. Hence, the paper does not deal with errors. The term imperfect data, hereafter, will represent the case of missing, uncertain and imprecise information.

Table 2. An example of a dataset with uncertainty

Employees	Deterministic Department	Stochastic Quality Bonus	Stochastic Sales
Jon Smith	Toy	0.4[Great Yes] 0.5[Good Yes] 0.1[Fair Yes]	0.3[\$30-34K] 0.7[\$35-39K]
Fre Jones	Housewares	1.0[Good Yes]	0.5[\$20-24K] 0.5[\$25-29K]

Now, in information systems containing missing, imprecise and uncertain values, it is not appropriate to apply to the systems a method that can deal only with missing values. An possible solution could be a combination of transforming any type of imperfect data to probabilistic values [20] and applying an probabilistic method [17, 18, 20, 23, 24]. This paper will introduce a representation of imperfectness including missing values, uncertainty as well as imprecision, then discuss the valued tolerance/similarity relation based rough set (VRRS) model in an imperfect information system, which is first introduced in [20]. Another important target is to find a method to obtain decision rules in imperfect information systems.

The paper is organized as follows; Section 2 suggests a representation of imperfect data. Section 3 summarizes rough set theory as well as valued tolerance/similarity relation definitions. Section 4 adopts several methods to determine *Probability of object attribute values* as well as a general definition of *Probability of matching* between objects, which can be used to define approximation spaces using VRRS approaches. The last but not least part proposes a rule induction method for imperfect information systems.

2. IMPERFECT DATA REPRESENTATION

An information system in the rough set study is defined as a pair $I = (U, A)$, where U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes such that $f_a : U \rightarrow V_a$ for every $a \in A$ [21, 22]. The non-empty discrete value set V_a is called the domain of a . The original rough set theory deals with complete information systems in which $\forall x \in U, a \in A, f_a(x)$ is a precise value.

Any information system of the form $I = (U, A \cup \{d\})$ is called decision table where $d \notin A$ is called decision and the elements of A are called conditions. The paper assumes $V_d = \{d_1, \dots, d_k\}$. The decision d determines a partition $\{C_1, C_2, \dots, C_k\}$ of the universe U , where $C_i = \{x \in U | f_d(x) = d_i\}, 1 \leq i \leq k$. The set C_i is called the i -th decision class or concept on U .

Now, for an information system, in which some attribute values of objects are missing and/or associated with probabilistic data, the attribute values of an object might be represented as follows:

DEFINITION 1. In an imperfect information system $I = (U, A)$, let $t_{a,i}^x \subseteq V_a$ be the i -th set of overall "s" possible value sets of "x" on "a" and $p_{a,i}^x > 0$ be its probability. Then the pair (T_a^x, P_a^x) , where $T_a^x = \{t_{a,i}^x | 1 \leq i \leq s\}$, $P_a^x = \{p_{a,i}^x | \sum_i p_{a,i}^x = 1\}$, represents imperfect values of object x on a .

In the above, $t_{a,i}^x$ are not necessarily be mutually disjoint. Obviously, for this representation of imperfect values, it is able to present any type of imperfectness discussed in Section 1. A value is uncertain when any set of possible of values is singleton. In this case $|t_{a,i}^x| = 1$. Some types of missing values may have a pre-defined probability distribution and the imperfectness could be regarded as uncertainty. One example is a game of four people

Table 3. Dice game

score	
Players	Score
Terry	3
David	6
Tom	*
Anna	8

playing with dice. Their scores can be calculated based on the sum of two dice thrown for each of them. Table 3 shows their scores. In this table, the score of Tom is unknown due to some reasons. However, the probability of each value for Tom's score can be identified by a probability distribution for the sum of two dice. The probability that Tom's score is 7, for example, is 1/6. On the other hand, the probability that his score equals to 11 is 1/18. In this case $\forall v_i \in V_a, t_{a,i}^x = \{v_i\}, p_{a,i}^x = \lambda_a(v_i)$ when $f_a(x) = *$, where $\lambda_a(v)$ is the probability mass function on a .

A value is imprecise when there is only a set of multiple possible values and the probability of this set is also 1, formally $|T_a^x| = 1, p_{a,1}^x = 1$. A precise value and a missing value with no pre-defined probability distribution can be considered as two extreme kinds of imprecision [16]. A value is precise if the set of possible values is singleton. In this case $|t_{a,1}^x| = 1$. Missing values without pre-defined probability distribution could be regarded as imprecise information where the set of possible values encompasses the entire attribute domain, such that $t_{a,1}^x = V_a$.

More importantly, if an object attribute value contains both uncertainty and imprecision, it is possible to use the imperfect representation of Definition 1 to show this type of value. To illustrate, the "Stochastic sales" value of Jon Smith in Table 2 is represented as follows: $t_{Sales,1}^{Smith} = \{30, 31, 32, 33, 34\}, p_{Sales,1}^{Smith} = 0.3$ and $t_{Sales,2}^{Smith} = \{35, 36, 37, 38, 39\}, p_{Sales,2}^{Smith} = 0.7$.

3. VALUED TOLERANCE/SIMILARITY RELATION AND ROUGH SET

A solution of defining valued tolerance/similarity relations can be stated as follows: first, the pair (T_a^x, P_a^x) of object x is defined for each attribute a , and the probability that two objects are tolerant of (similar to) each other on the attribute is determined. Then, the degree that two objects are tolerant of (similar to) each other on a set of attributes is calculated, for example, using the joint probabilities assuming that all the attributes are independent of one another. This section will summarize concepts in valued tolerance/similarity relation definitions as well as a rough set approach with this kind of relations. Problems of valued tolerance/similarity relation based rough set models in the related work are also addressed in this section.

In complete information systems, the relation $EQU_P(x, y), P \subseteq A$ denotes a binary relation between objects that are equivalent in terms of values of attributes in P [21]. The equivalence relation is reflexive, symmetric, and transitive. Let $E_P(x) = \{y \in U | EQU_P(y, x)\}$ be the set of all objects that are equivalent to x by P , and be called equivalence class.

Now, for an information system, in which some attribute values of objects are missing and/or associated with uncertainty or imprecision, the paper will define probabilities of attribute values. For a discrete attribute, *Probability of object attribute value* denoted by $Pr_a(f_a(x) = v)$ represents the probability that $x \in U$ takes the value $v \in V_a$ on attribute $a \in A$. Two methods to estimate

the probabilities of object attribute values will be discussed in the next section.

Based on the probability estimated, **Probability of matching** between two objects $x, y \in U$ on attribute $a \in A$ denoted by $\theta_a(x, y)$ defines the probability that object x takes the same value as object y on attribute a . In [4, 23, 24], it is supposed that there is an uniform probability distribution on an attribute, and the probability of matching is defined as $\theta_a(x, y) = Pr_a(f_a(x) = v_i) \bullet Pr_a(f_a(y) = v_i) = 1/|V_a|^2$ where v_i is a value in the domain of attribute a . The definition is clearly inadequate when the attribute values of both “ x ” and “ y ” are supposed to be missing on “ a ”. The definition of probability of matching is discussed and calculated in general in the next section.

From the probability of matching between two objects, the degree that $x, y \in U$ are tolerant of (similar to) each other on a set of attributes $P \subseteq A$, which is denoted by $\phi_P(x, y)$ can be induced. The degree of tolerance/similarity can be defined as the probability that two objects have the same values on all attributes in set P and is calculated by joint probability $\phi_P(x, y) = \prod_{a \in P} \theta_a(x, y)$ assuming independence among attributes. Other methods of tolerance (similarity) degree definitions can be found in [19].

Now, it is able to define a relation $R_P(x, y)$ between objects x and y by controlling the degree of tolerance(similarity) using threshold α , such that $R_P(x, y) \Leftrightarrow \phi_P(x, y) \geq \alpha$. Based on that, a neighbourhood, which consists of successor and predecessor sets, of an object [10, 27] is determined. The successor set of x is the set of objects to which x is similar:

$$sucR_P(x) = \{y \in U | R_P(x, y)\} \quad (1)$$

The predecessor set of x is the set of objects which is similar to x :

$$preR_P(x) = \{y \in U | R_P(y, x)\} \quad (2)$$

For complete information systems, the lower and upper approximations are defined on the basis of an *indiscernibility* relation and should be the same among the different three definitions: singleton, subset and concept definitions [21, 22]. In the case of non-equivalence relations, which may not be reflexive, symmetric nor transitive, approximation spaces defined may lead to variant results [10]. Formally, approximations based on singleton, subset and concept approaches with a valued tolerance/similarity relation can be defined as follows:

Singleton lower approximation:

$$\underline{SingleAppr}_P(X) = \{x \in U | R_P(x) \subseteq X\} \quad (3)$$

Singleton upper approximation:

$$\overline{SingleAppr}_P(X) = \{x \in U | R_P(x) \cap X \neq \emptyset\} \quad (4)$$

Subset lower approximation:

$$\underline{SubsetAppr}_P(X) = \cup\{R_P(x) | x \in U \wedge R_P(x) \subseteq X\} \quad (5)$$

Subset upper approximation:

$$\overline{SubsetAppr}_P(X) = \cup\{R_P(x) | x \in U \wedge R_P(x) \cap X \neq \emptyset\} \quad (6)$$

Concept lower approximation:

$$\underline{ConceptAppr}_P(X) = \cup\{R_P(x) | x \in X \wedge R_P(x) \subseteq X\} \quad (7)$$

Concept predecessor upper approximation:

$$\overline{ConceptAppr}_P(X) = \cup\{R_P(x) | x \in X \wedge R_P(x) \cap X \neq \emptyset\} \quad (8)$$

where $R_P(x)$ denotes either successor and predecessor neighbourhood sets of x .

The difference between subset and concept definitions may be missed easily. In subset definition, extended tolerance classes of all elements in the universal set are examined, while only elements in X are examined in the case of concept definition.

Obviously, singleton lower and upper approximations of X are subsets of the subset lower and upper approximations of X , respectively. The subset lower approximation is the same set as the concept lower approximation. The concept upper approximation, however, is a subset of the subset upper approximation.

4. PROBABILITY OF MATCHING

This section will show how to define the probability of matching in an imperfect information system discussed in Section 2 in order to apply VRRS methods to this type of information system. According to Section 3, the probability of each object attribute value needs to be defined before calculating probability of matching.

4.1 Probability of object attribute values

In general, if there is no information about probability distribution of attribute values, it is possible to make the hypothesis that the probability is determined by an uniform distribution.

DEFINITION 2. *In an imperfect information system $I = (U, A)$, let the pair (T_a^x, P_a^x) present an imperfect attribute value of object x on a . Then the probability that an object x takes the value $v \in V_a$ on a can be calculated as follows:*

$$Pr_a(f_a(x) = v) = \sum_i p_{a,i}^x \bullet \frac{|\{v\} \cap t_{a,i}^x|}{|t_{a,i}^x|} \quad (9)$$

In this equation, the probability mass distributed equally in possible value set $t_{a,i}^x$ will be added to the probability of attribute value if $v \in t_{a,i}^x$, such that $|\{v\} \cap t_{a,i}^x| = 1$. Obviously, in case of uncertainty where a probability distribution is given, it is not necessary to calculate the probability of object attribute values. In this case $Pr_a(f_a(x) = v) = p_{a,i}^x$ because $t_{a,i}^x = \{v\}$. In case of the missing value without any pre-defined probability distribution, $Pr_a(f_a(x) = v) = 1/|V_a|$ for any $v \in V_a$.

However, even there is no pre-defined probability distribution, the probability of attribute values in some cases still can be estimated. The next step will summarize two possible solutions discussed in [20] with the adaptation of the representation of imperfect values in Definition 1.

4.2 Method of the frequency of attribute value

The approach is based on the notion of “The most common method”. This is a method of handling missing values summarized by Grzymala-Busse [9, 10], in which, missing values are replaced by the most common value of the attribute. The method of handling missing attribute values is implemented, e.g., in well-known machine learning algorithm CN2 [3].

Suppose the value domains are known. First, the probability that each value of the attribute appears based on the frequency of the available values in a dataset is defined. The probability that a value $v \in V_a$ appears as a value of a certain object is define by:

$$\rho_a(v) = \begin{cases} \frac{|V_a(v)|}{|U - V_a(?)|} & \text{if } V_a(?) \subset U, \\ \frac{1}{|V_a|} & \text{otherwise.} \end{cases} \quad (10)$$

where $V_a(v)$ and $V_a(?)$ are the sets of objects whose attribute value is “ v ” and the set of objects whose value on “ a ” is imperfect, respectively. The symbol “ \subset ” denotes a proper subset. As seen in the equation, the probability $\rho_a(v)$, $v \in V_a$ is defined by the ratio of the value v among objects whose values are not imperfect. If $V_a(?) = U$, that is, values of attribute a are imperfect in all objects, the equal probability distribution is given. The value of $\rho_a(v)$ is greater than zero if there is at least an object such that $f_a(x) = v$. Since it could be zero for many values if the size of U is small, the size of U should be large enough when using the approach.

Now, it is possible to define the probability of object attribute values by the frequency of values in a dataset. Formally, in imperfect information system $I = (U, A)$, an attribute $a \in A$ and its domain V_a , $\rho_a(v)$ denotes the frequency of each value $v \in V_a$ in the dataset. Given an object $x \in U$ with imperfect value on a the probability $Pr_a(f_a(x) = v)$ for any $v \in V_a$ as follows:

$$Pr_a(f_a(x) = v) = \begin{cases} \sum_i p_{a,i}^x \cdot |\{v\} \cap t_{a,i}^x| \cdot \frac{\rho_a(v)}{\sum_{v' \in t_{a,i}^x} \rho_a(v')} & \text{if } \sum_{v' \in t_{a,i}^x} \rho_a(v') \neq 0, \\ \sum_i p_{a,i}^x \cdot |\{v\} \cap t_{a,i}^x| \cdot \frac{1}{|t_{a,i}^x|} & \text{otherwise.} \end{cases} \quad (11)$$

Of course, the idea could be applied to the missing value with no predefined probability distribution. However, it should not be applied to attributes where uncertainty or a probability distribution is derived from a theoretical point of view, e.g. in the case of dice game mentioned before.

4.3 Method of the frequency of attribute value related to concepts

This is an extension of the method in the previous subsection. Observing some systems, sometimes it is possible to recognize that attribute values might depend on some concepts. Supposed the value domains are known, the probability that a value $v \in V_a$ appears as a value of objects contained in a concept $X \subseteq U$ is defined as follows:

$$\rho_a(v)_X = \begin{cases} \frac{|V_a(v)_X|}{|X - V_a(?)_X|} & \text{if } V_a(?)_X \subset X, \\ \frac{1}{|V_a|} & \text{otherwise.} \end{cases} \quad (12)$$

where $V_a(v)_X$ and $V_a(?)_X$ are the set of objects in concept X whose attribute value is “ v ” and the set of objects whose value on “ a ” is imperfect, respectively.

Table 4. Probability of attribute values given a concepts

Attributes	Values	Probability in concepts	
		Flu=Yes	Flu=No
Temperature	very-high	0.33	0.00
Temperature	high	0.67	0.33
Temperature	normal	0.00	0.67
Headache	yes	1.00	0.33
Headache	no	0.00	0.67
Nausea	yes	0.67	0.50
Nausea	no	0.33	0.50

Table 4 shows that flu relates to high and very-high temperature, headache and nausea. On the other hand non-flu supports the cases of low temperature and no headache.

In the same way as the previous method, it is possible to define the probability of object attribute values by the frequency of values in the dataset. Formally, in imperfect information system $I = (U, A)$, an attribute $a \in A$ and its domain V_a , $\rho_a(v)_X$ denotes the frequency of each value $v \in V_a$ in concept X . Given an object $x \in X$ with imperfect value on a the probability $Pr_a(f_a(x) = v)$ for any $v \in V_a$ as follows:

$$Pr_a(f_a(x) = v) = \begin{cases} \sum_i p_{a,i}^x \cdot |\{v\} \cap t_{a,i}^x| \cdot \frac{\rho_a(v)_X}{\sum_{v' \in t_{a,i}^x} \rho_a(v')_X} & \text{if } \sum_{v' \in t_{a,i}^x} \rho_a(v')_X \neq 0, \\ \sum_i p_{a,i}^x \cdot |\{v\} \cap t_{a,i}^x| \cdot \frac{1}{|t_{a,i}^x|} & \text{otherwise.} \end{cases} \quad (13)$$

4.4 Obtaining Probability of matching

This section will re-define the degree that two object have the same value on an attribute if at least one of the two objects has the missing, imprecise or uncertain value on the attribute.

DEFINITION 3. Given an information system $I = (U, A)$, on an attribute $a \in A$ with its domain V_a , the probability that the value of “ x ” is the same as the value of “ y ” on “ a ” is given by:

$$\theta_a(x, y) = \sum_{v \in V_a} Pr_a(f_a(x) = v | f_a(y) = v) Pr_a(f_a(y) = v) \quad (14)$$

when $x \neq y$. Otherwise $\theta_a(x, y) = \theta_a(x, x) = 1$. Note that $\theta_a(x, y) = 1$, if the two objects x and y have the same precise value on a , while it is zero if they have different precise values. $Pr_a(f_a(x) = v | f_a(y) = v)$ denotes the conditional probability of $f_a(x) = v$ given $f_a(y) = v$. Hereafter, the paper assumes that two events $f_a(x) = v$ and $f_a(y) = u$, $x, y \in U$, $a \in A$ are independent of each other for any $u, v \in V_a$.

The probability of matching for each type of missing, uncertain and imprecise values has been discussed in [20]. However, the concept

of imperfectness introduced in this paper generalizes the way to calculate the probability of matching as follows:

$$\theta_a(x, y) = \begin{cases} \sum_{v \in \Lambda(x, y)} Pr_a(f_a(x) = v)Pr_a(f_a(y) = v) & \text{if } \Lambda(x, y) \neq \emptyset, \\ 0 & \text{if } \Lambda(x, y) = \emptyset, \end{cases} \quad (15)$$

when $x \neq y$, $\Lambda(x, y) = [\cup_i t_{a,i}(x)] \cap [\cup_j t_{a,j}(y)]$. Otherwise $\theta_a(x, y) = \theta_a(x, x) = 1$.

Obviously, the probability of matching between two objects equals to zero unless there are common possible values for both two objects. Otherwise, the sum of product of probabilities should be taken that the two objects coincide to have a common value.

In short, the method of calculating probability that two objects have the same attribute values in case of imperfect information has been shown. This probability can be used to define a valued tolerance/similarity relation, then obtain approximation space in any published VRRS approach. Note that rough set definition is not in the scope of this paper. In the next parts, a rough set based rules induction for imperfect information systems will be introduced.

5. RULES INDUCTION

Rule induction is one of the most important knowledge discovery techniques in machine learning. A decision rule can be presented in the following expression:

$$r : \wedge_i (a_i = v) \rightarrow (d = w)$$

where $a_i \in A$, $v \in V_{a_i}$, and d, w is the decision attribute and a decision value respectively. The set $\mathcal{A}_r = \cup\{a_i\}$ and attribute d is called condition set and decision, respectively, of rule r . Hereafter, $f_a(r)$ and $f_d(r)$ represent the value of attribute $a \in \mathcal{A}_r$ and decision d , respectively, in r . $R_P(x, r)$, the same symbol for a relation between objects, is used to represent the relation between object x and the conditional part of r with respect to attribute set $P \subseteq A$.

In supervised learning, rules are obtained from information which consists of conditional and decisional attributes. However, due to imperfect data and/or some other reasons, rules may conflict with each other. In Table 1 for example, the rule from case x_4 : $(Temperature = high) \wedge (Headache = yes) \wedge (Nause = yes) \rightarrow (Flu = yes)$ conflicts with the rule of x_5 : $(Temperature = high) \wedge (Nause = yes) \rightarrow (Flu = no)$ if it is assumed that the missing value of *Headache* is "yes".

Rough sets, which describe a set of objects in the approximation space, play a vital role in rule induction. Rules induced from the certain region (lower approximation) and possible region (upper approximation) of a concept are called certain and possible rules respectively [7, 28].

Among published rule induction algorithms, LEM2 (Learning from Examples Module, version 2) of LERS (Learning from Examples using Rough Sets) is used commonly since it gives better results [7]. The algorithm is based on the idea of blocks of attribute-value pairs. For an attribute-value pair (a, v) , a block $[(a, v)]$ is a set of all cases from U such that for attribute a has value v . This algorithm can be also used for some rough set approaches in incomplete information systems [5, 24, 28] in which objects belong to the block $[(a, v)]$ if their values on a are tolerant of (similar to) v .

Table 5. An example incomplete table

Cases	a	b	d
x_1	a_1	*	d_1
x_2	*	b_1	d_1

Let B be a non-empty lower or upper approximation of a concept represented by a decision-value pair (d, w) . Let us say that the set B depends on a set T of attribute-value pairs if and only if

$$\emptyset \neq [T] = \cap_{(a,v) \in T} [(a, v)] \subseteq B \quad (16)$$

From the equation, it could be believed intuitively that the less cardinality of T is the more objects T covers. Thus, set T is called a minimal complex of B if and only if B depends on T and no proper subset T' exists such that B depends on T' . LEM2 algorithm defines each rule based on a minimal complex.

However, the belief mentioned above does not hold in some cases such as relations defined in [18, 19, 25, 26]. Let us illustrate the problem. Table 5 presents an incomplete information system. The following equation shows a relation called Extended Tolerance relation defined by Wang [25]:

$$R_P(x, y) \Leftrightarrow (\forall a \in P, f_a(x) = *, f_a(y) = *) \vee ((O_P(x) \cap O_P(y) \neq \emptyset) \wedge (\forall a \in O_P(x) \cap O_P(y) f_a(x) = f_a(y)))$$

where $O_P(x) = \{a | a \in P, f_a(x) \neq *\}$.

Using the Extended Tolerance relation, the pair $\{(a, a_1)\}$ covers only x_1 , $\{(b, b_1)\}$ covers only x_2 , yet $\{(a, a_1), (b, b_1)\}$ covers both x_1 and x_2 in Table 5, which violates the intuitive belief.

Taking imperfect data discussed in Section 2 into account, in the rest of this section, a method to obtain decision rules in imperfect information systems will be proposed.

DEFINITION 4. In an imperfect information system $I = (U, A)$, a candidate rule set suggested from object x is denoted by $\mathcal{S}(x)$ and defined by the following equation:

$$\mathcal{S}(x) = \{r | (a \in P \subseteq A, P \neq \emptyset, f_a(r) \in \cup_i t_{a,i}^x) \wedge R_P(x, r)\} \quad (17)$$

In the definition, the paper limits possible conditional values of the candidate rule r using the possible values of object x as well as the conditional part of the rule should be tolerant of (similar to) the object x with respect to attributes in \mathcal{A}_r . From this definition, the suggested rule set $\mathbb{S}(X)$ of an object set $X \subseteq U$ can be defined by:

$$\mathbb{S}(X) = \cup_{x \in X} \mathcal{S}(x) \quad (18)$$

On the other hand, for a rule r , a set of objects that rule r covers is defined by the following equation:

$$\mathcal{G}(r) = \{x | x \in U, R_{\mathcal{A}_r}(x, r)\} \quad (19)$$

A rule r is *optimal* if and only if no rule r' exist such that $\mathcal{A}_{r'} \subset \mathcal{A}_r$ and $\mathcal{G}(r') = \mathcal{G}(r)$. Let $\mathbb{G}(\mathcal{R}) = \cup_{r \in \mathcal{R}} \mathcal{G}(r)$ denotes the set of objects that a rule set \mathcal{R} covers, two regions are defined as follows:

DEFINITION 5. A rule set \mathcal{R} is called a lower covering of a set of objects B if only if the following conditions are satisfied:

Table 6. An example of a dataset with imperfect values

Cases	Temperature	Headache	Nausea	Flu
x_1	high	{yes:0.9; no:0.1}	no	yes
x_2	very-high	yes	yes	yes
x_3	{high:0.3; normal:0.7}	no	no	no
x_4	high	yes	yes	yes
x_5	high	{yes:0.1; no:0.9}	yes	no
x_6	normal	yes	no	no
x_7	normal	no	yes	no
x_8	{very-high:0.3; high:0.7}	yes	{yes:0.7; no:0.3}	yes

- (1) each member of \mathcal{R} is a optimal.
(2) $\forall r \in \mathcal{R}, \mathbb{G}(\mathcal{R} - \{r\}) \subset \mathbb{G}(\mathcal{R})$.
(3) $\mathbb{G}(\mathcal{R}) \subseteq B$
(4) and there is no rule set \mathcal{R}' such that $\mathbb{G}(\mathcal{R}) \subset \mathbb{G}(\mathcal{R}') \subseteq B$

DEFINITION 6. A rule set \mathcal{R} is called an upper covering of a set of objects B if only if the following conditions are satisfied:

- (1) each member of \mathcal{R} is a optimal.
(2) $\forall r \in \mathcal{R}, \mathbb{G}(\mathcal{R} - \{r\}) \subset \mathbb{G}(\mathcal{R})$.
(3) $\mathbb{G}(\mathcal{R}) \supseteq B$
(4) and there is no rule set \mathcal{R}' such that $\mathbb{G}(\mathcal{R}) \supset \mathbb{G}(\mathcal{R}') \supseteq B$

Based on the two definitions above, the paper suggests an algorithm to induce certain and possible rules by finding a lower and upper covering of a set of objects as follows:

Input: A set of object $X \subseteq U$

Output: lower covering $\underline{\mathcal{R}}$ and upper covering $\overline{\mathcal{R}}$ of X

Step 1:

Make a candidate rule set $L = \mathbb{S}(X)$;
Remove any rule in L which is not optimal;
 $L' := \{r | r \in L, \mathcal{G}(r) \subseteq X\}$;
 $L := L - L'$;
 $\underline{\mathcal{R}} := \emptyset$;
 $B := X$;

Step 2:

If $B = \emptyset$ or $L' = \emptyset$ then go to step 3;
Select the rule $r \in L'$ such that $\mathcal{G}(r) \cap B$ is the maximum;
If a tie occurs, select the rule with the smallest $|\mathcal{A}_r|$;
If another tie occurs, select the first rule;
 $\underline{\mathcal{R}} := \underline{\mathcal{R}} \cup \{r\}$;
 $B := B - (B \cap \mathcal{G}(r))$;
Remove from L' all rules r such that $B \cap \mathcal{G}(r) = \emptyset$;
Repeat the step 2;

Step 3:

$\overline{\mathcal{R}} := \underline{\mathcal{R}}$

Step 4:

Remove from L all rules r such that $B \cap \mathcal{G}(r) = \emptyset$;
If $B = \emptyset$ or $L = \emptyset$ then stop;
Select the rule $r \in L$ such that $\mathcal{G}(r) - B$ is the minimum;
If a tie occurs, select the rule such that $\mathcal{G}(r) \cap B$ is maximum;
If a tie occurs, select the rule with the smallest $|\mathcal{A}_r|$;
If another tie occurs, select the first rule;
 $\overline{\mathcal{R}} := \overline{\mathcal{R}} \cup \{r\}$;

Table 7. Candidate rules

Rules	Temperature	Headache	Nausea	Covered objects
r_1			no	x_1, x_3, x_6, x_7, x_8
r_2		yes		$x_1, x_2, x_4, x_5, x_6, x_8$
r_3		yes	no	x_1, x_6, x_8
r_4		no		x_1, x_3, x_5, x_7
r_5		no	no	x_1, x_3
r_6	high			x_1, x_3, x_4, x_5, x_8
r_7	high		no	x_1, x_3, x_8
r_8	high	yes		x_1, x_4, x_5, x_8
r_9	high	yes	no	x_1, x_8
r_{10}	high	no		x_1, x_3, x_5
r_{11}			yes	x_2, x_4, x_5, x_7, x_8
r_{12}		yes	yes	x_2, x_4, x_5, x_8
r_{13}	very-high			x_2, x_8
r_{14}	high		yes	x_4, x_5, x_8

The absence of attribute values in a rule mean those value are not exist in conditional part of that rule.

$$B := B - (B \cap \mathcal{G}(r));$$

Remove any rule $r \in \overline{\mathcal{R}}$ such that $\mathcal{G}(r) \subset \mathbb{G}(\overline{\mathcal{R}} - \{r\})$;

Repeat the step 4;

The information in Table 6 will be used to illustrate the algorithm. In this example, the simplest method - joint probability - is used to calculate the degree of tolerance, formally $\phi_P(x, y) = \prod_{a \in P} \theta_a(x, y)$. The threshold α is supposed to be 0.1. Thus, $R_P(x, y) \Leftrightarrow \phi_P(x, y) \geq 0.1$. Now it is possible to induce certain and possible rules for the concept $X = \{x | f_{Flu}(x) = yes\}$.

At first, the candidate rule set $\mathbb{S}(X)$ is calculated using equations (17) and (18), where any rules that can satisfy one or more cases in Table 6 are generated. Then rules, that are not optimal, will be removed. For example, rule $r' : (Temperature = very - high) \wedge (Headadge = yes) \rightarrow (Flu = yes)$ is not optimal because it covers the same set $\{x_2, x_8\}$ with the rule $r_{13} : (Temperature = very - high) \rightarrow (Flu = yes)$ while $\mathcal{A}_{r_{13}} \subset \mathcal{A}_{r'}$. The rules that are optimal are shown in Table 7. Then, to find certain rules, rules in a candidate set which cover only objects belonging to X should be chosen. Thus, the candidate set for certain rules is $L' = \{r_9, r_{13}\}$.

Going to the step 2, which is the step to induce certain rules, the maximal cardinality of $\mathcal{G}(r) \cap B$ is two. So the rule r_{13} , which has the smallest number in cardinality of conditional part, is selected such that $|\mathcal{A}_{r_{13}}| = 1$. Thus, the first certain rule presented by the first element of the lower covering is:

$$(Temperature = very - high) \rightarrow (Flu = yes)$$

Now, the rule is added to the certain rule set and remove all covered objects of this rule such that $B = B - \{x_2, x_8\} = \{x_1, x_4\}$. Then, rule r_9 is chosen and it covers x_1 . Hence the next certain rule is:

$$(Temperature = high) \wedge (Headache = yes) \\ \wedge (Nausea = no) \rightarrow (Flu = yes)$$

The step is stopped because $L' = \emptyset$.

Going through the step 3 and 4, the possible rule set of the concept X is obtained as follows:

$$(Temperature = high) \wedge (Headache = yes) \rightarrow (Flu = yes) \\ (Temperature = very - high) \rightarrow (Flu = yes)$$

From the above, it is possible to see that x_4 does not support any certain rule. This is because any candidate rule in $S(x_4)$ would cover some other objects with $Flu=no$. This type of rule may be present in possible rules instead.

6. CONCLUSION

In this paper, a new representation of multiple types of imperfect information including missing, uncertain and imprecise values is introduced. Using this representation, the paper defines the probability that two objects are tolerant of (similar to) each other. This probability can be used in various types of valued tolerance/similarity relations based rough set definitions that have been published so far.

In addition, a method of obtaining decision rules from an imperfect decision table was also discussed and proposed. At the same time, the algorithm also can produce both certain and possible rule without calculation of approximation space for a set of objects.

The imperfect representation could be utilized in some applications where data is described in imprecise as well as uncertain representation rather than precise values. One of typical fields with the need is Kansei Engineering or affective engineering [13, 14] where human feeling, subjective impression, affective images as well as imprecise specifications of target products are dealt with frequently in a decision table. In those systems, approximations using and the rule induction algorithm would be a great help for the analysis.

7. REFERENCES

- [1] D. Barbará, H. Garcia-Molina, and D. Porter. The management of probabilistic data. *IEEE Trans. on Knowl. and Data Eng.*, 4(5):487–502, October 1992.
- [2] Roger Cavallo and Michael Pittarelli. The theory of probabilistic databases. In *Proceedings of the 13th International Conference on Very Large Data Bases, VLDB '87*, pages 71–81, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.
- [3] Peter Clark and Tim Niblett. The cn2 induction algorithm. *Mach. Learn.*, 3(4):261–283, March 1989.
- [4] Yucui Feng, Wenhai Li, Zehua Lv, and Xiaoming Ma. Probabilistic approximation under incomplete information systems. In Zhongzhi Shi, K. Shimohara, and David Dagan Feng, editors, *Intelligent Information Processing*, volume 228 of *IFIP*, pages 73–80. Springer, 2006.
- [5] Jerzy W. Grzymala-Busse. Characteristic relations for incomplete data: A generalization of the indiscernibility relation. 3700:58–68, 2005.
- [6] Jerzy W. Grzymala-Busse. A rough set approach to data with missing attribute values. In Guoyin Wang, James F. Peters, Andrzej Skowron, and Yiyu Yao, editors, *RSKT*, volume 4062 of *Lecture Notes in Computer Science*, pages 58–67. Springer, 2006.
- [7] Jerzy W. Grzymala-Busse. Rule induction. In Maimon and Rokach [15], pages 249–265.
- [8] Jerzy W. Grzymala-Busse and Witold J. Grzymala-Busse. Handling missing attribute values. In Maimon and Rokach [15], pages 33–51.
- [9] Jerzy W. Grzymala-Busse and Ming Hu. A comparison of several approaches to missing attribute values in data mining. In *Revised Papers from the Second International Conference on Rough Sets and Current Trends in Computing, RSCTC '00*, pages 378–385, London, UK, UK, 2001. Springer-Verlag.
- [10] Jerzy W. Grzymala-Busse and Wojciech Rzasca. Definability and other properties of approximations for generalized indiscernibility relations. *T. Rough Sets*, 11:14–39, 2010.
- [11] Marzena Kryszkiewicz. Rough set approach to incomplete information systems. *Inf. Sci.*, 112(1-4):39–49, December 1998.
- [12] Marzena Kryszkiewicz. Rules in incomplete information systems. *Inf. Sci.*, 113(3-4):271–292, 1999.
- [13] Nagamachi M. Kansei engineering: a new ergonomics consumer-oriented technology for product development. *International Journal of Industrial Ergonomics*, 15:3–10, 1995.
- [14] Nagamachi M. Kansei engineering as a powerful consumer-oriented technology for product development. *Fundamental Reviews in Applied Ergonomics*, 33(3):289–294, 2002.
- [15] Oded Maimon and Lior Rokach, editors. *Data Mining and Knowledge Discovery Handbook, 2nd ed.* Springer, 2010.
- [16] A. Motro. *Uncertainty management information systems: From needs to solutions.* Kluwer Acad., 1997.
- [17] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Knowledge reduction in incomplete decision tables using probabilistic similarity-based rough set model. In *12th International Symposium on Advanced Intelligent Systems (ISIS 2011)*, pages 147–150, 2011.
- [18] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Rough set model based on parameterized probabilistic similarity relation in incomplete decision tables. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, pages 577–582, 2012.
- [19] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. Extended tolerance relation to define a new rough set model in incomplete information systems. *Advances in Fuzzy Systems*, 2013, 2013. Article ID 372091.
- [20] Do Van Nguyen, Koichi Yamada, and Muneyuki Unehara. On probability of matching in probabilistic based rough set definitions. In *IEEE-SMC2013*, pages 449–454, Manchester, The UK, 2013.
- [21] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.
- [22] Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data.* Kluwer Acad., 1991.

- [23] Jerzy Stefanowski and Alexis Tsoukis. On the extension of rough sets under incomplete information. In Ning Zhong, Andrzej Skowron, and Setsuo Ohsuga, editors, *RSFDGrC*, volume 1711 of *Lecture Notes in Computer Science*, pages 73–81. Springer, 1999.
- [24] Jerzy Stefanowski and Alexis Tsoukis. Incomplete information tables and rough classification. *Computational Intelligence*, 17(3):545–566, 2001.
- [25] Guoyin Wang. Extension of rough set under incomplete information systems. In *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, volume 2, pages 1098–1103, 2002.
- [26] Xibei Yang, Xiaoning Song, and Xiaohua Hu. Generalisation of rough set for rule induction in incomplete system. *IJGCRSIS*, 2(1):37–50, 2011.
- [27] Y. Y. Yao. On generalizing rough set theory. In Guoyin Wang, Qing Liu, Yiyu Yao, and Andrzej Skowron, editors, *RSFDGrC*, volume 2639 of *Lecture Notes in Computer Science*, pages 44–51. Springer, 2003.
- [28] Y. Y. Yao. Three-way decisions with probabilistic rough sets. *Information Sciences*, 2010(3):341–353, 2010.