Semi-supervised Based Rough Set to Handle Missing Decision Data

Thinh Cao*, Koichi Yamada†, Muneyuki Unehara*, Izumi Suzuki*, and Do Van Nguyen†
*Department of Management of Information Systems, Nagaoka University of Technology, Nagaoka, Japan
Email: caothinh@stn.nagaokaut.ac.jp, {yamada,unehara,suzuki}@kjs.nagaokaut.ac.jp
†Institute of Information Technology, Ha Noi, Vietnam, Email: ngdovan@gmail.com

Abstract—We have developed a rough set model for analyzing an information system in which some conditions as well as decision values are missing. Current studies have focused mainly on the missing of condition data but seem to ignore the missing of decision data. The common approach is to remove objects with no decision values because such objects are apparently considered fruitless from the decision-making standpoint. However, this deletion may lead to the risk of information loss. We observe that such a situation is somewhat similar to the semi-supervised situation in the sense that some objects are characterized by complete decision data while some are not. Considering both kinds of objects from a probabilistic view, we predict potential candidates for missing values by comparing measurements of two factors, local decision belief and universal decision belief, with a parameter threshold α. These possible decision candidates help to form a relative dissimilarity relation, which measures the unlikeness of pairs of objects rather than their likeness. Contrasting with the other approaches, rough set definitions based on this relation do not approximate the target set but its complement instead. The knowledge acquisition induced by the common approach and the proposed approach is compared, and the result shows that the latter can overcome some limitations of the former. This approach is new and flexible to deal with missing decision information.

TABLE I

AN EXAMPLE OF UNAVAILABLE DECISION VALUES

<table>
<thead>
<tr>
<th>Patient</th>
<th>Condition</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fever</td>
<td>Chills</td>
</tr>
<tr>
<td>1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

I. INTRODUCTION

The original rough set theory [1], [2] deals with precise and complete data while real applications often encounter incomplete data. In fact, problems on missing condition values have been studied in recent times and got positive results [3]–[27]. These studies can be divided into two branches: an indirect and direct approach. The indirect approach attempts to transform an incomplete information system to a complete one by some ways: removing objects with missing condition values; replacing them with average values, or all possible values, etc [8], [10]. The direct approach extends the concept of classical equivalence relation by relaxing requirements of reflexivity, symmetry, and transitivity. Most of the methods in this approach are based on either the tolerance relation [16] or the similarity relation [14]. While the former rests on the idea that missing values may equal to any value in the value domain, the latter considers they are really "lost" and can not be compared to other values. For example, valued tolerance relation [13], maximal consistent block [18] are constructed on the basis of the tolerance relation, whereas the difference relation [26] is built as an extension of the similarity relation. All the above methods, as well as many others, presume that the data in the decision attribute is perfect, i.e. without noise, inconsistency or missing data.

However, decision information may be unavailable due to various reasons. If an information system contains data on symptoms and diseases of patients in a hospital, some decisions may be missing when the patients stop coming for treatment, e.g., for financial reasons. Also, the data may be blank owing to the inadvertent discard by someone. Although this could occur in reality, there have been few studies on the theme of missing decision values. Medhat [28] suggested a method to restore missing decision values by measuring a similarity distance between objects with and without missing decision values. However, this similarity distance can apply only to numerical data. In addition, similar objects may not have the same decision values necessarily.

A common approach is to remove such objects with missing decision data. However, the induced knowledge after the removal would be different from the knowledge in the original table where those objects are not removed. For example, given the data as in Table I, we can detect easily that the second patient may take another decision value different from the one of the first. If we remove the second patient (and the fourth patient), the data as in Table I, we can detect easily that the second patient may take another decision value different from the first. However, decision information may be unavailable due to various reasons. If an information system contains data on symptoms and diseases of patients in a hospital, some decisions may be missing when the patients stop coming for treatment, e.g., for financial reasons. Also, the data may be blank owing to the inadvertent discard by someone. Although this could occur in reality, there have been few studies on the theme of missing decision values. Medhat [28] suggested a method to restore missing decision values by measuring a similarity distance between objects with and without missing decision values. However, this similarity distance can apply only to numerical data. In addition, similar objects may not have the same decision values necessarily.

Another weakness of this deletion is to alter the size of a
data set, thus changing the data distribution. For example, in the Fever attribute, since the value Yes appears more frequent than No, the appearance probability of the former is likely to be higher than the one of the latter. However, this probability would be changed if we remove those patients with missing decision values.

In addition, this deletion could break the relation between condition attributes. In the above table, without considering the patients with missing decision values, someone can draw a conclusion that Fever always occurs along with Chills on every patient when he or she tries to examine the relation between Fever and Chills attributes. However, the conclusion is inaccurate if considering those deleted objects. In a real application, e.g., health diagnosis system, attributes of a certain disease might not be independent but mutually dependent to a certain level. Maintaining relation between these attributes is important for analyzing the disease.

Since ignored objects with missing decision values could affect the knowledge induction, i.e., causing information loss, it is evident that they should not be removed or should be handled in a more proper way.

This study has been partly inspired by the recent advances in the field of semi-supervised learning. The semi-supervised paradigm rests on the idea that labeled data is rather arduous but unlabeled data is easy to collect. The usage of both kinds of data, when combined, can result in an improvement of the learning capacity. The problem of missing decision values is close to the semi-supervised context in the sense that some data is labeled by decisions and the others are not. The existing rough set approaches to an incomplete information system so far have been limited to the case where all data is clearly labeled by decisions but neglecting those are unlabeled. Our study, instead, utilizes these two types of data to exploit the missing data in a decision attribute. This new study tackles a more general problem of machine learning from an incomplete information system, where both labeled and unlabeled data coexist.

II. ROUGH SET PRELIMINARIES [1], [2], [16], [17]

A. Information system

An information system in the rough set study is formally described as $I = (U, \text{AT} \cup \{d\}, V, f)$, where $U$ is a non-empty set of objects, $\text{AT}$ is a non-empty set of condition attributes, $d \notin \text{AT}$ denotes a decision attribute, $f$ is an information function $f : U \times \text{AT} \cup \{d\} \rightarrow V$, $V = \cup V_t$ for any $t \in \text{AT} \cup \{d\}$. $f(x, a)$ and $f(x, d)$, $x \in U$, $a \in \text{AT}$ are represented by $f_a(x)$ and $f_d(x)$ respectively. $V_a$ and $V_d$ denotes the domain of $f_a$ and $f_d$ respectively. Any domain may contain special symbols "*" to indicate a missing value, i.e., the value of an object is unknown. Any value different from "*" will be called regular [18]. Denoted by CIS, an IS in which all data are regular, by IIS, an IS where unknown values appear only in the condition part, and by IDS, an IS where unknown values can appear in both condition and decision attributes. Table II is an example of IDS

<table>
<thead>
<tr>
<th>Car</th>
<th>Price(P)</th>
<th>Mileage(M)</th>
<th>Size(S)</th>
<th>Max-Speed(X)</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>low</td>
<td>high</td>
<td>full</td>
<td>low</td>
<td>excel</td>
</tr>
<tr>
<td>2</td>
<td>medium</td>
<td>medium</td>
<td>full</td>
<td>low</td>
<td>excel</td>
</tr>
<tr>
<td>3</td>
<td>low</td>
<td>medium</td>
<td>medium</td>
<td>*</td>
<td>poor</td>
</tr>
<tr>
<td>4</td>
<td>low</td>
<td>*</td>
<td>*</td>
<td>high</td>
<td>poor</td>
</tr>
<tr>
<td>5</td>
<td>high</td>
<td>low</td>
<td>full</td>
<td>high</td>
<td>good</td>
</tr>
<tr>
<td>6</td>
<td>high</td>
<td>*</td>
<td>full</td>
<td>high</td>
<td>good</td>
</tr>
<tr>
<td>7</td>
<td>high</td>
<td>low</td>
<td>full</td>
<td>high</td>
<td>poor</td>
</tr>
<tr>
<td>8</td>
<td>high</td>
<td>low</td>
<td>full</td>
<td>high</td>
<td>*</td>
</tr>
</tbody>
</table>

where $Price(P), Mileage(M), Size(S), Max-Speed(X)$ are condition attributes, $d$ is a decision attribute.

B. Tolerance relation

A tolerance relation, $TOL(A)$, $A \subseteq AT$, represents a binary relation between objects that are possibly indiscernible w.r.t the attribute set $A$. Formally,

$$TOL(A) = \{(x, y) \in U \times U | \forall a \in A, f_a(x) = f_a(y) \lor f_a(x) = * \lor f_a(y) = *\} \tag{1}$$

The tolerance relation is reflexive, symmetric but not necessarily transitive. Let $S_A(x) = \{y \in U | (x, y) \in TOL(A)\}$ be a set of objects tolerant of $x$ in terms of $A$, and be called a tolerance class. Suppose $A = AT$, tolerance classes from Table II are obtained as follows:

$S_A(1) = \{1\}, S_A(2) = \{2\}, S_A(3) = S_A(4) = \{3, 4\}, S_A(5) = S_A(6) = S_A(7) = S_A(8) = \{5, 6, 7, 8\}$.

C. Rough set

Let $X \subseteq U$ and $A \subseteq AT$. The lower approximation of $X$ w.r.t $A$ is defined by the next equation:

$$R(X) = \{x \in U | S_A(x) \subseteq X\} = \bigcup_{x \in X} \{S_A(x) | S_A(x) \subseteq X\} \tag{2}$$

The upper approximation of $X$ w.r.t $A$ is defined by the following equation:

$$\overline{R}(X) = \{x \in U | S_A(x) \cap X \neq \emptyset\} = \bigcup_{x \in U} \{S_A(x) | S_A(x) \cap X \neq \emptyset\} \tag{3}$$

The boundary region and outside region of $X$ are defined by

$$BND(X) = \overline{R}(X) - R(X), \quad OUT(X) = U - R(X) = \{x \in U | S_A(x) \cap X = \emptyset\}. \tag{4}$$

If the boundary region of $X$ is not empty, the pair $(R(X), \overline{R}(X))$ is called a rough set.
D. Generalized decision

Let $A \subseteq AT$. A generalized decision is a function $\partial_A : U \rightarrow 2^\omega$ represented by

$$\partial_A(x) = \{f_d(y) | y \in S_A(x)\}. \quad (5)$$

$\partial_A(x)$ defines a set of decision values of objects tolerant of $x$, or it is a set of possible decision values of objects which may have the same attribute values as $x$.

E. Reduct and Relative Reduct

In an IS, there may exist a subset of attributes that fully characterizes the same knowledge of the database as does the full attribute set. The minimal subsets among those are called reducts. The concept of reduct originally does not consider the decision attribute, the paper thus focuses on another concept relative reduct, which depends on the decision attribute. There are various definitions of relative reduct in rough set theory [21]. Such definitions rest on one of the three following preservation: positive region preservation, generalized decision preservation, and relative indiscernibility (or discernibility) relation preservation. It is crucial to note that the relative reduct aiming at positive region preservation introduced by Pawlak in [1] is just one of the definitions. Instead, we introduce another study conducted by M. Kryszkiewicz [16].

A set $A \subseteq AT$ is a relative reduct iff it meets the following conditions:

$$(1) \partial_A(x) = \partial_{AT}(x), \forall x \in U$$
$$(2) \forall B \subset A, \partial_B(x) \neq \partial_{AT}(x), \quad (6)$$

where $\subset$ denotes a proper subset.

According to this definition, a relative reduct is a minimal set of condition attributes satisfying that the generalized decision induced w.r.t $A$ is the same as that induced w.r.t $AT$. An information system may have no, one or several relative reducts.

III. ROUGH SET IN IDS

In this section, we present a progress to build up a rough set model for an IDS. Our approach is inspired by the idea that an object with a missing decision value could take one of the decision values of its neighbors, i.e. the objects "near" to it. The probability of these neighbor decision values might have correlation with their frequencies in the neighbors as well as in the whole set of objects. From this idea, we construct a relative dissimilarity relation and develop a new rough set model.

A. Relative dissimilarity relation

In studies on IIS, a binary relation is established based on pairwise comparison of condition attributes [3]–[7], [9], [15], [16], [18], [22]. In the case of IDS, we believe that not only condition attributes but also the decision attribute would play a certain role in defining the similarity (or dissimilarity) of objects. This leads us to the definition of relative dissimilarity relation.

First, let us start by defining the minimum probability that a decision value appears based on the frequency in the data set. If $t = f_d(x) \neq *$, i.e., the decision value of the object $x$ is not missing, the probability that $t$ appears is between $\frac{|V_d(x)\cap V_d(t)|}{|V_d(t)|}$ and $\frac{|V_d(t)|+|V_d(*)|}{|V_d|}$. $V_d(t)$ and $V_d(*)$ denotes sets of objects whose decision values are $t$ and unknown respectively, and $|A|$ denotes the cardinality of the set $A$. Then, the minimum probability that a decision value $t \neq *$ appears can be defined by:

$$Q(t) = \frac{|\{x; f_d(x) = t\}|}{|V_d|}. \quad (7)$$

We regard $Q(t)$ as the belief about the appearance of the decision value $t$ over the whole decision domain $V_d$, it is thus labeled as universal decision belief. In fact, the value $Q(t)$ is a probability in the case of CIS and IIS, but not in the case of IDS in general.

Using the notion of universal decision belief, we define $\alpha$-generalized decision, which is a substitute of the generalized decision devised for IDS. Mathematically, the $\alpha$-generalized decision w.r.t $A \subseteq AT$ is determined by the following equation:

$$\omega_A^\alpha(x) = \{t | y \in S_A(x), t = f_d(y) \neq *\} \quad \text{if} \quad f_d(x) \neq *,$$
$$\omega_A^\alpha(x) = \{t | P(x,t), Q(t) > \alpha, y \in S_A(x), t = f_d(y) \neq *\} \quad \text{if} \quad f_d(x) = *.$$ 

(8)

where

$$P(x,t) = \frac{|\{z | f_d(z) = t, z \in S_A(x)\}|}{|\{z | f_d(z) \neq *, z \in S_A(x)\}|}. \quad (9)$$

$P(x,t) = 0$ if the denominator of (9) is zero. The parameter $\alpha$ is a real value in the range $[0, 1]$. $\omega_A^\alpha(x)$ represents a set of possible decision values of the objects in the tolerance class $S_A(x)$. When $f_d(x) \neq *$, $\omega_A^\alpha(x)$ takes the decision values of the objects tolerant of $x$ if these decision values are not missing. When $f_d(x) = *$, $\omega_A^\alpha(x)$ takes only the decision values $t = f_d(y) \neq *, y \in S_A(x)$ such that the product $P(x,t). Q(t) > \alpha$, and the others are ignored. In this case, the parameter $\alpha$ plays as a threshold to filter decision values. The smaller $\alpha$ is, the more decision values $\omega_A^\alpha(x)$ could take and vice versa. Consequently, $\omega_A^\alpha(x)$ becomes an empty set if $P(x,t). Q(t) \leq \alpha \forall t$. Table III illustrates the $\alpha$-generalized decision with $\alpha = 0.1$.

$P(x,t)$ represents the probability that $f_d(x)$ takes the decision values $t$ of $x$'s neighbors if these decision values are not missing. Note that $x$'s neighbors are the objects similar to $x$ in terms of condition attributes, these objects thus belong to $S_A(x)$. In a sense, $P(x,t)$ implies a local decision belief comparing with the universal decision belief defined above.

The reason to employ both these beliefs is that the reliability of the local decision belief may be low when the number of objects in a tolerance class is small, while the cardinality of the whole object set is at least larger than that of the tolerance class. The reliability of the universal decision belief may also be low because it does not consider the similarity in the condition attributes of objects. Using both the local and
the universal decision belief thus compensates the weakness of each other.

The following theorems hold for the local, universal decision belief and the \( \alpha \)-generalized decision:

**Theorem 1.** If \( Q(t) = 0 \) then \( P(x, t) = 0 \) \( \forall x \in U \), but the opposite does not hold.

**Theorem 2.** If \( \exists x \in U \) s.t. \( P(x, t) > 0 \) then \( Q(t) > 0 \), but the opposite does not hold.

**Theorem 3.** If \( \alpha \geq \beta \), then \( \omega^\alpha_A(x) \subseteq \omega^\beta_A(x) \) \( \forall x \in U \).

---

**B. Rough set model based on relative dissimilarity relation**

A target set in the rough set study is normally a collection of objects with the same decision information, i.e., a member of a partition of the universe of discourse in terms of decision values. In the case where there is no missing value in a decision attribute, the target set could be defined clearly by decision attribute values. When some decision values are missing, the target set may not be defined by them necessarily. For simplicity, we restrict the target set is given by (a) regular decision value(s). This means 1) the target set contains only the objects with a regular decision value, and 2) objects with the same decision value are all in the target set or all outside the target set.

The conventional rough set draws on the equivalence relation which states that objects are inevitably belong to the target set if all objects equivalent to them belong to the target set. Oppositely, resting on the relative dissimilarity relation, we believe that objects are certainly exterior to the target set, i.e., they belong to the complement of the target set, if all the objects relatively dissimilar to them, are also exterior to the target set. Instead of a target set \( dX_i \), we will thus define rough sets for its complement, i.e., \( -dX_i \).

Let \( I = (U, R) \) be an IDS. It consists of a non-empty, finite universe of discourse \( U \) and of the relative dissimilarity relation \( R \). Given \( A \subseteq AT \), the \( \alpha \) lower and upper approximations of \( -dX_i \) w.r.t. \( A \) is defined by:

\[
\overline{A}_\alpha(-dX_i) = \bigcup \{DIS^\alpha_A(x)|x \in U, DIS^\alpha_A(x) \subseteq -dX_i\},
\]

\[
\overline{A}_\alpha(-dX_i) = \bigcup \{DIS^\alpha_A(x)|x \in U, DIS^\alpha_A(x) \cap -dX_i \neq \emptyset\}.
\]

The boundary region and \( \alpha \) outside region of \( -dX_i \) w.r.t. \( A \) are defined by:

\[
BND^\alpha_A(-dX_i) = \overline{A}_\alpha(-dX_i) - \overline{A}_\alpha(-dX_i),
\]

\[
OUT^\alpha_A(-dX_i) = U - \overline{A}_\alpha(-dX_i).
\]

In the above example, let \( dX_i = \{x \in U|f_2(x) = excel\} = \{1, 2\} \). Thus, \( -dX_i = \{3, 4, 5, 6, 7, 8\} \), \( \overline{A}_\alpha(-dX_i) = \{3, 4, 5, 6, 7, 8\} \), and \( \overline{A}_\alpha^{0.1}(-dX_i) = \{1, 2, 3, 4, 5, 6, 7, 8\} \).

We can prove the following properties: (See Appendix)

1. \( \overline{A}_\alpha(X) \subseteq X \subseteq \overline{A}_\alpha(X) \).
2. \( \overline{A}_\alpha(\emptyset) = \emptyset, \overline{A}_\alpha(\emptyset) = \emptyset \).
3. \( \overline{A}_\alpha(U) = U, \overline{A}_\alpha(U) = U \).
4. \( \overline{A}_\alpha(X \cup Y) = \overline{A}_\alpha(X) \cup \overline{A}_\alpha(Y) \).
5. \( \overline{A}_\alpha(X \cap Y) = \overline{A}_\alpha(X) \cap \overline{A}_\alpha(Y) \).
6. \( \overline{A}_\alpha(X \cup Y) = \overline{A}_\alpha(X) \cup \overline{A}_\alpha(Y) \).
follows:

(7) $\overline{R}_A^\alpha (X \cap Y) = R_A^\alpha (X) \cap R_A^\beta (Y)$.

(8) If $X \subseteq Y$ then $R_A^\alpha (X) \subseteq R_A^\alpha (Y)$, $\overline{R}_A^\alpha (X) \subseteq \overline{R}_A^\alpha (Y)$.

(9) If $\alpha \geq \beta$, then $R_A^\beta (X) \subseteq R_A^\alpha (X)$, $\overline{R}_A^\beta (X) \subseteq \overline{R}_A^\alpha (X)$.

IV. REDUCT IN IDS

In this section, we attempt to acquire knowledge, i.e., relative reducts, from an IDS. Relative reducts in CIS were extracted based on the notion of the generalized decision and discernibility matrix [16], [17]. However, these notions can not be defined in the IDS because we have introduced the missing value in the decision part. Therefore, we have to find a new approach to extract relative reduct for the IDS.

Using the relative dissimilarity relation, we define a relative dissimilarity function between objects in terms of $A \subseteq AT$ as follows:

\[
\mu_A^\alpha (x,y) = \begin{cases} 
1 & \text{if } (x,y) \in DIS^\alpha (A) \forall x, y \in U \\
-1 & \text{otherwise}.
\end{cases}
\]  

(14)

The function classifies pairs of objects into two groups: 1 is assigned to a pair of objects that are relatively dissimilar, otherwise $-1$. Then we extend this function for two sets of attributes by the next equation. Formally, a relative comparability function between two attribute sets $A, B \subseteq AT$ is determined by:

\[
\varepsilon^\alpha (A, B) = \begin{cases} 
1 & \text{if } \mu_A^\alpha (x,y) = 1 \Leftrightarrow \mu_B^\alpha (x,y) = 1, \forall x, y \in U \\
0 & \text{otherwise}.
\end{cases}
\]  

(15)

de$^\alpha (A, B) = 1$ iff the relative dissimilarity relation between objects in terms of $A$ is equivalent to that in terms of $B$ for every pair of objects. Two sets $A, B$ are called relatively comparable. In the case of $A$ and $AT$, the set $A$ is said to preserve the relative dissimilarity relation of the IDS. This recognition result is in the concept of a reduct in IDS in the following. A reduct of IDS is a minimal set of attribute $A$ that satisfies the two following conditions:

(1) $DIS^\alpha (A) = DIS^\alpha (AT)$,

(2) $\forall B \subseteq A, DIS^\alpha (B) \neq DIS^\alpha (AT)$.

(16)

The first condition shows that the attributes in the set $A$ jointly preserve the relative dissimilarity relation while the second condition shows each attribute in $A$ is individually necessary for the preservation, i.e., any proper subset $B \subset A$ does not maintain the relation.

We can prove that the reduct definition above is correspondent to the following conditions. $A \subseteq AT$ is a reduct of IDS iff it satisfies

Theorem 5.

\[
(1) \varepsilon^\alpha (A, AT) = 1,
\]

\[
(2) \varepsilon^\alpha (A - \{a\}, A) = 0 \text{ } \forall a \in A.
\]  

(17)

Proof: See Appendix.

The first condition indicates that if two sets $A$ and $AT$ are relatively comparable, they also preserve the relative dissimilarity relation. In addition, if an attribute $a$ is removed from $AT$, the new set $\{AT - \{a\}\}$ is not relatively comparable to $AT$ anymore, thus the relative dissimilarity relation between them will not be preserved.

V. EXAMPLE

In this section, we will compare two approaches of the knowledge acquisition: deletion of objects with missing decision values and the proposed approach.

In the case of the former approach, the information system turns to be IIS (Table IV). Following the method mentioned in [16], the relative reduct is extracted by constructing the discernibility matrix shown in Table V. The acquired relative reducts are $\{P, M, X\}$ and $\{P, S, X\}$.

In our approach, the original IDS is kept unchanged as shown in Table II. With different values of $\alpha$, we compare the relative dissimilarity functions of the full set of attributes, i.e., $\mu_{\{AT\}}^\alpha (x,y)$, and of its subsets, i.e., $\mu_{\{P, M, X\}}^\alpha (x,y)$, $\mu_{\{P, S, X\}}^\alpha (x,y)$, $\mu_{\{M, S, X\}}^\alpha (x,y)$ respectively.

Suppose $\alpha = 0.05$, the three sets $\{P, S, X\}$, $\{P, M, X\}$, and $\{M, S, X\}$ are the reducts in this IDS since they satisfy all the conditions of (17). The relative dissimilarity function on $AT$ and these three sets coincides as shown in Table VI. On the other hand, suppose $\alpha = 0.1$, the set $\{M, S, X\}$ is excluded, just only two sets $\{P, S, X\}$, $\{P, M, X\}$ remain reducts. Notice that the knowledge acquisition, i.e., the relative reducts are different with different $\alpha$ values, and one of these $\alpha$ values give the same reducts as extracted by the object removal approach. In fact, deleting objects with missing decision values is nothing but abandoning valuable information included in condition attributes, which has led to the smaller sets of
reduce \{P, M, X\} and \{P, S, X\}. On the other hand, our approach has found that one more set \{M, S, X\} should also be included in the reduct sets when we do not know the real missing decision values. The proposed method thus offers a more generalized approach to deal with missing decision data.

### VI. CONCLUSION

It is not rare that condition as well as decision values are missing because of many reasons. The necessity to study on the issue seems clear, however there are few studies on this topic. This paper introduces a new rough set model to cope with the issue. Recognizing the similarity to the semi-supervised approach, the method outlined in this paper utilizes objects with missing decision values even if they have only partial information available. After clarifying which are and which are not neighbor objects, a threshold \(\alpha\) is employed to filter possible candidates for the missing decision values with the help of the local and universal decision beliefs. Unlike other methods, to believe that the dissimilarity measurement among objects would be more promising rather than their similarity measurement, we created the relative dissimilarity relation and base the approximation space on this notion. The paper has compared the knowledge induction gained by both methods: the object removal and the proposed method. While the former may lead to the deficiency in knowledge acquisition, the latter provides a more generalized approach to handle missing decision data without having to remove them, hence minimizing the threat of data loss. This parameter-based method also offers a subtle advantage: the flexibility. With different \(\alpha\) value adopted on each system, variants of approximation and knowledge acquisition can be generated.

### REFERENCES


Theorem 1: If $Q(t) = 0$ then $P(x,t) = 0 \forall x \in U$, but the opposite does not hold. 

Proof: $\Rightarrow Q(t) = 0$ means $\exists y \in U$ such that $f_d(y) = t$. 
So, there is no $y \in S_A(x)$ such that $f_d(y) = t$, or $P(x,t) = 0 \forall x \in U$. 

$\Leftarrow$ Let $t \in V_d$. There may exist an object $y$ such that $y \notin S_A(x)$ and $f_d(y) = t$. In this case, $P(x,t) = 0$ but $Q(t) \neq 0$.

Theorem 2: If $\exists x \in U.s.t.P(x,t) > 0$ then $Q(t)$>0, but the opposite does not hold. 

Proof: $\Rightarrow$ Suppose $\exists x \in U.s.t.P(x,t) > 0$. This implies $\exists y \in S_A(x)$ such that $f_d(y) = t$. So, $Q(t)$ > 0. 

$\Leftarrow$ Let $t \in V_d$. There may exist an object $y$ such that $y \notin S_A(x)$ and $f_d(y) = t$. In this case, $Q(t)$>0 but $P(x,t) = 0$.

Theorem 3: If $\alpha > \beta$, then $\omega^A_\beta(x) \subseteq \omega^A_\alpha(x) \forall x \in U$. 

Proof: If $f_d(x) \neq *$, $\omega^A_\beta(x) = \omega^A_\alpha(x) \forall x$. If $f_d(x) = *$, suppose $m \in \omega^A_\beta(x)$, then $m \in \omega^A_\alpha(x) = \{t|P(x,t),Q(t) > \alpha, y \in S_A(x), t = f_d(y) \neq *\}$. Because $\alpha > \beta$, thus $m \in \{t|P(x,t),Q(t) > \beta, y \in S_A(x), t = f_d(y) \neq *\}$ or $m \in \omega^A_\alpha(x)$.

Theorem 4: If $\alpha > \beta$, $DIS^A_\beta(x) \subseteq DIS^A_\alpha(x) \forall x \in U$. 

Proof: Suppose $y \in DIS^A_\beta(x)$, which means $[(x,y) \notin TOL(A)] \land [\omega^A_\beta(x) \cap \omega^A_\alpha(y) = \emptyset]$. Because $\omega^A_\alpha(x) \subseteq \omega^A_\beta(x)$ according to the theorem 3, we have $[(x,y) \notin TOL(A)] \land [\omega^A_\alpha(x) \cap \omega^A_\alpha(y) = \emptyset]$. Thus, $y \in DIS^A_\alpha(x)$.

Properties:

(1) $R^A_\alpha(X) \subseteq X \subseteq \overline{R^A_\alpha(X)}$.

Proof: Suppose $y \in R^A_\alpha(X)$, then $\exists z \in U$ such that $y \in DIS^A_\alpha(z) \subseteq X$. Thus $y \in X$, and $\overline{R^A_\alpha(X)} \subseteq X$.

Suppose $y \in X$, then $\exists z \in U$ such that $y \in DIS^A_\alpha(z)$. Because $y = DIS^A_\alpha(z) \cap X \neq \emptyset$. $DIS^A_\alpha(z) \subseteq \bigcup\{DIS^A_\beta(t)\mid t \in U,DIS^A_\beta(t) \cap X \neq \emptyset\}$. So, $y \in R^A_\alpha(X)$, and $X \subseteq \overline{R^A_\alpha(X)}$.

(2) If $X = \emptyset$, $R^A_\alpha(\emptyset) = \overline{R^A_\alpha(\emptyset)} = \emptyset$.

We assume that the target set must have regular decision values. And $X$ is the complement of the target set. $X = \emptyset$ implies that the target set is $U$, thus this is not an IDS.

Proof: From (1), if $X = \emptyset$, then $R^A_\alpha(X) = \emptyset$. We must prove $\overline{R^A_\alpha(X)} = \emptyset$. Suppose that $\overline{R^A_\alpha(X)} \neq \emptyset$. This means $\exists z \in U$ s.t. $DIS^A_\alpha(z) \cap X \neq \emptyset$, a contradiction because $X = \emptyset$. Hence $\overline{R^A_\alpha(X)} = \emptyset$.

(3) $\overline{R^A_\alpha(U)} = U, \overline{\overline{R^A_\alpha(U)}} = U$.

Proof: From (1), it follows $R^A_\alpha(X) = U$ if $X = U$. On the other hand, $\forall x \in U,DIS^A_\alpha(x) \subseteq U$, thus $\bigcup DIS^A_\alpha(x) = U$. For that reason, $\overline{R^A_\alpha(X)} = \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \subseteq X\} = U$.

(4) $\overline{R^A_\alpha(X \cup Y)} = \overline{R^A_\alpha(X)} \cup \overline{R^A_\alpha(Y)}$.

Proof: If $X = \emptyset$ or $Y = \emptyset$, the equation is certain. In case $X \neq \emptyset$ and $Y \neq \emptyset$, $\forall y \in \overline{R^A_\alpha(X \cup Y)}$.

$\Rightarrow y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \cap (X \cup Y) \neq \emptyset\}$

$\Rightarrow \{y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \cap X \neq \emptyset\} \cup \{y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \cap Y \neq \emptyset\}\}$

$\ni y \in \overline{R^A_\alpha(X)} \cup \overline{R^A_\alpha(Y)}$.

(5) $R^A_\alpha(X \cap Y) = R^A_\alpha(X) \cap R^A_\alpha(Y)$

(6) $\overline{R^A_\alpha(X \cup Y)} = \overline{R^A_\alpha(X)} \cup \overline{R^A_\alpha(Y)}$

(7) $\overline{R^A_\alpha(X \cap Y)} = \overline{R^A_\alpha(X)} \cap \overline{R^A_\alpha(Y)}$.

Proof: Similar to (4).

(8) If $X \subseteq Y$ then $\overline{R^A_\alpha(X)} \subseteq \overline{R^A_\alpha(Y)}$.

$\Rightarrow$ Assume $y \in \overline{R^A_\alpha(X)}$, then $y \in X$. This is a contradiction because $y \in Y$ but $y \notin \overline{R^A_\alpha(Y)}$.

Assume $y \in \overline{R^A_\alpha(X)}$, $y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \cap X \neq \emptyset\}$. Because $X \subseteq Y$, then $y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \cap Y \neq \emptyset\}$. Thus $y \notin \overline{R^A_\alpha(Y)}$.

(9) If $\alpha > \beta$, then $\overline{R^A_\alpha(X)} \subseteq \overline{R^A_\beta(X)}$.

Proof: Suppose $y \in \overline{R^A_\alpha(X)}$, $y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \subseteq X\}$. Because $DIS^A_\alpha(x) \subseteq DIS^A_\beta(x)$ according to the theorem 4, we have $y \in \bigcup\{DIS^A_\beta(x)\mid x \in U,DIS^A_\beta(x) \subseteq X\}$, or $y \notin \overline{R^A_\alpha(X)}$.

Similarly, we have $\overline{R^A_\beta(X)} \subseteq \overline{R^A_\alpha(X)}$.

Theorem 5: $A \subseteq AT$ is a reduct in IDS iff it holds the two following conditions:

(1) $\varepsilon^\alpha(A,AT) = 1$

(2) $\varepsilon^\alpha(A - \{a\}, A) = 0 \forall a \in A$

Proof: $\Rightarrow$ Suppose $A$ is a reduct, $DIS^\alpha(A) = DIS^\alpha(AT)$, implying $\mu^\alpha_A(x,y) = 1 = \forall x,y \in U$, then $\varepsilon^\alpha(A,AT) = 1$.

On the other hand, a reduct $A$ demonstrates that $\forall B \subseteq A,DIS^\alpha(B) \neq DIS^\alpha(AT)$. Since $B$ can be represented as $B = \{A - \{a\}\}, a \in A$, thus $DIS^\alpha(A - \{a\}) \neq DIS^\alpha(AT)$ or $\varepsilon^\alpha(A - \{a\}, A) = 0$.

$\Leftarrow$ The first condition requires that the relative dissimilarity function between two arbitrary objects produced on $A$ is equivalent to that produced on $AT$, i.e., preserving the relative dissimilarity relation. This condition directly collaborates with the first requirement of reduct. The second condition $\varepsilon^\alpha(A - \{a\}) = 0 \forall a \in A$ indicates that if $\mu^\alpha_{AT - \{a\}}(x,y) = 1$, then $\mu^\alpha_{AT}(x,y) = -1$ and vice versa, or the relative dissimilarity relation can not be preserved by removing $a$. Since we can not remove any attribute from $A$, it is the minimal subset which can preserve the relative dissimilarity relation.