Hindawi Publishing Corporation
Journal of Applied Mathematics and Decision Sciences
Volume 2007, Article ID 58248, 13 pages
doi:10.1155/2007/58248

Research Article

A Modified Rough Set Approach to Incomplete Information Systems

E. A. Rady, M. M. E. Abd El-Monsef, and W. A. Abd El-Latif

Received 30 October 2006; Revised 27 January 2007; Accepted 12 March 2007

Recommended by James Moffat

The key point of the tolerance relation or similarity relation presented in the literature is to assign a "null" value to all missing attribute values. In other words, a "null" value may be equal to any value in the domain of the attribute values. This may cause a serious effect in data analysis and decision analysis because the missing values are just "missed" but they do exist and have an influence on the decision. In this paper, we will introduce the modified similarity relation denoted by MSIM that is dependent on the number of missing values with respect to the number of the whole defined attributes for each object. According to the definition of MSIM, many problems concerning the generalized decisions are solved. This point may be used in scaling in statistics in a wide range. Also, a new definition of the discernibility matrix, deduction of the decision rules, and reducts in the present of the missing values are obtained.

1. Introduction

Sooner or later, anyone who does statistical analysis runs into problems with missing data in which information for some variables is missing for some objects.

Missing data are questions without answers or variables without observation. Even a small percentage of missing data can cause serious problems with the analysis leading to draw wrong conclusions and imperfect knowledge. There are many techniques to manipulate the imperfect knowledge and manage data with missing items, but no one is absolutely better than the others. Different situations require different solutions as Allison [1] says: "The only really good solution to the missing data problem is not to have any."

Rough set theory proposed by Pawlak [2] is an effective approach to imprecision, vagueness, and uncertainty. Rough set theory overlaps with many other theories such that fuzzy sets, evidence theory, and statistics. From a practical point of view, it is a good tool for data analysis. The main goal of the rough set analysis is to synthesize approximation of concepts from acquired data.

The starting point of Rough set theory is an observation that the objects having the same description are indiscernible (similar) with respect to the available information. Determination of the similar objects with respect to the defined attributes values is very hard and sensible when some attribute values are missing. This problem must be handled very carefully. In this work, we tried to deal with this problem to restrict the conditions of similarity in the presence of missing values.

The indiscernibility relation is a fundamental concept of the rough set theory which used in the complete information systems. In order to process incomplete information systems, the indiscernibility relation needs to be extended to some equivalent relations.

In the literature there are several extensions such as tolerance relation, nonsymmetric relation, and valued tolerance relation. The problem of missing values handling within the rough set framework has been already discussed in the literature, for example, [3, 4]. These approaches consider alternative definitions of the discernibility which reflect various semantics of missing attribute values.

In [5, 6], the authors performed computational studies on the medical data, where unknown values of the attributes were replaced using probabilistic techniques. Recently, Greco et al. used a specific definition of the discernibility relation to analyze unknown attribute values for multicretria decision problems [7, 8]. In [9] two different semantics for incomplete information "missing values" and "absent values" were discussed also; they introduced two generalizations of the rough set theory to handle these situations. In [10] the author examined methods of valued tolerance relations. They proposed a correctness criterion to the extension of the conventional methods which is based on rough sets for handling missing values.

A new tolerance relation to handle incomplete information tables was introduced in [11]. A "null" value may be any observed value in the domain of the attribute values without any restriction. Hence, as well as any object has a lot of missing values, it can be similar to a lot of other objects in despite of their decisions. So the obtaining generalized decisions may be a combination of some decisions far from each other. Accordingly, the concluding results concerning the reducts, the set approximations, and the decision rules will be unreasonable and unacceptable.

This work enhances and develops Krysckiewiczs' work by introducing some essential restrictions and conditions on the similarity between objects and introducing the modified similarity relation (MSIM).

2. Complete information systems and decision tables

The concept of rough set has to be an effective tool for analyzing vague information and knowledge discovery. The starting point of rough set theory which is based on data analysis is a data set called an information system (IS). IS is a data table, whose columns are

labeled by attributes, rows are labeled by objects or cases, and the entire of the table are the attribute values.

Formally, IS is a pair IS = (U, AT), where U and AT are nonempty finite sets called "the universe" and "the set of attributes," respectively. For example, in Table 4.1, $U = \{1,2,3,4,5,6\}$, $AT = \{\text{price, mileage, size, max speed}\}$ and with every attribute $a \in AT$, a set V_a of its values called the "domain of a" is associated. If V_a contains missing values for at least one attribute, then S is called an incomplete information system, otherwise it is complete.

Any information table defines a function ρ that maps the direct product $U \times AT$ into the set of all values in Table 4.1 $\rho(1, \text{price}) = \text{high}$.

The concept of the indiscernibility relation is an essential concept in rough set theory which is used to distinguish objects described by a set of attributes in complete information systems. Each subset A of AT defines an indiscernibility relation as follows:

$$\text{IND}(A) = \{(x,y) \in U \times U : \rho(x,a) = \rho(y,a)\ \forall a \in A,\ A \subset AT\}. \tag{2.1}$$

Obviously, IND(A) is an equivalence relation. The family of all equivalence classes of IND(A), for example, a partition determined by A which is denoted by $U/\text{IND}(A)$ or U/A.

If we distinguish in an information system between two disjoint classes of attributes called conditions and decision attributes, the first one is an independent variable and the other one is dependent, respectively, then the system will be called a decision table (DT).

3. Incomplete information systems and the similarity relation

It may happen that some of the attribute values for some objects are missing (denoted by "$*$" in Table 4.1). In order to process incomplete information systems (IIS), the indiscernibility relation has been extended to some equivalent relations, for example, tolerance relation, similarity relation, valued tolerance relation, and so forth.

Following [11], similarity relation SIM(A) denotes a binary relation between objects that are possibly indiscernible in terms of values of attributes, for example, we cannot say that these objects are different and their similarity relation was defined by

$$\text{SIM}(A) = \{(x,y) \in U \times U,\ a \in A,\ \rho(x,a) = \rho(y,a) \text{ or } \rho(x,a) = * \text{ or } \rho(y,a) = *\} \tag{3.1}$$

and $S_A(x)$ denotes that the set $\{y \in U : (x,y) \in \text{SIM}(A),\ A \subset AT\}$ is the maximal set of objects which are possibly indiscernible by A with x.

The minimal reducts for IIS are as follows.

A set $A \subset AT$ is a reduct of IIS if and only if

$$\text{SIM}(A) = \text{SIM}(AT), \quad \forall B \subset A, \qquad \text{SIM}(B) \neq \text{SIM}(AT). \tag{3.2}$$

The (incomplete) decision table (DT) is an (incomplete) information system DT = $(U, AT \cup \{d\})$, where d is a distinguished attribute called decision, and AT are called conditions where $d \notin AT$ and $* \notin V_d$.

Let us define function $\partial_A : U \to P(V_d)$, $A \subseteq AT$, as follows:

$$\partial_A(x) = \{i : i = d(y),\ y \in S_A(x)\}, \tag{3.3}$$

where ∂_A is the generalized decision in DT.

Accordingly, the minimal reduct for DT was defined as follows: a set $A \subset AT$ is reduct of DT if and only if

$$\partial_A(x) = \partial_{AT}(x), \quad \forall B \subset A \partial_B(x) \neq \partial_{AT}(x). \tag{3.4}$$

Any decision table may be regarded as a set of generalized decision rules of the form

$$\wedge(c,v) \longrightarrow \vee(d,w), \quad \text{where } c \in A,\ v \in V_c,\ w \in V_d,\ d \notin A. \tag{3.5}$$

4. Motivations

Any missing value in the classical definition introduced in [11] can be similar to any other value in the domain of the attribute values. No restriction on the decision or on the values of other attributes for such case. Also when the number of missing values exceeds, a lot of vagueness and uncertainty appear in the information system; hence, the concluding results will not be reasonable. Let us show all these problems by the following example.

Example 4.1. Given the following descriptions of several cars according to price, mileage, size, and max speed, the decision attribute D is the acceleration as shown in Table 4.1:

(1) the set of objects or the universe set is $U = \{1,2,3,4,5,6\}$;

(2) the set of the attributes is $AT = \{\text{price, mileage, size, max speed}\}$.

According to [6], the set $S_{AT}(x)$, for all $x \in U$, can be calculated as follows:

$$\begin{aligned} S_{AT}(1) &= \{1\}, \\ S_{AT}(2) &= \{2,6\}, \\ S_{AT}(3) = S_A(5) &= \{3,4,5,6\}, \\ S_{AT}(4) &= \{3,4,5\}, \\ S_{AT}(6) &= \{2,3,5,6\}. \end{aligned} \tag{4.1}$$

Hence, the generalized decisions will be as in Table 4.2.

As shown in Table 4.2, the generalized decisions over the set of all attributes AT for the objects 3, 4, 5, 6 are a combination of completely far or extreme decisions "poor," "good," "excellent," and this is not an acceptable result. Accordingly, the calculated reducts for the objects 3, 4, 5, 6 and the reduced decision rules could not be determined.

In our point of view, object "3" is one of the important reasons to this disturbance which happens in the generalized decisions because it has a lot of missing attribute values, only the max speed attribute is known. This makes this object similar to a lot of other attributes, so we conclude that the number of missing values for each object must be taken into account during the analysis.

Table 4.1

Car	Price	Mileage	Size	Max speed	D
1	High	High	Full	Low	Good
2	Low	$*$	Full	Low	Good
3	$*$	$*$	$*$	High	Poor
4	High	$*$	Full	High	Good
5	$*$	$*$	Full	High	Excellent
6	Low	High	Full	$*$	Good

To introduce the new definition of similarity relation, we will use the following definitions of ρ_x^a or $\rho(x,a)$, where $\rho_x^a = v$ means that the object x has a value v for the attribute a where $x \in U$, $a \in AT$.

Definition 4.2. ρ_x^a is called "a defined value" if and only if $\rho_x^a \neq *$ where $*$ is the symbol of any missing value in Table 4.1.

Definition 4.3. x is called "a completely defined object" if and only if $\rho_x^a \neq *$ for all $a \in AT$.

Let (ρ_x^a, ρ_y^a) denote a pair of values of the attribute "a" for the objects "x" and "y," respectively. For example, in Table 4.1, $(\rho_1^{\text{price}}, \rho_2^{\text{price}}) = (\text{high}, \text{low})$.

Definition 4.4. The number "EP". $EP = |(\rho_x^a, \rho_y^a)|$ for all $a \in A$, $A \subseteq AT$ is the number of equal pairs for the attribute "a" for all $a \in A$ for the objects "x," "y," respectively, where ρ_x^a, ρ_y^a are defined values.

For example, in Table 4.1, if $A = AT$, then for the objects 1, 2, $EP = 2$.

Definition 4.5. The number $|m_x|$. $|m_x|$ denotes the number of missing values for the object x.

Definition 4.6. An object $x \in U$ is called

$$\begin{aligned} &\text{well defined object of } A \subset AT \quad \text{iff} \\ &\qquad |m_x| \leq \frac{N}{2} \text{ if } N \text{ even or } |m_x| \leq \frac{N+1}{2} \text{ if } N \text{ odd}, \\ &\text{poorly defined object of } A \subset AT, \quad \text{iff} \\ &\qquad |m_x| > \frac{N}{2} \text{ if } N \text{ even or } |m_x| > \frac{N+1}{2} \text{ if } N \text{ odd}, \end{aligned} \tag{4.2}$$

where N is the number of the attributes in the set A, for example, $|A| = N$.

Definition 4.7. The number $|nm_x|$. $|nm_x|$ denotes the number of not missing values for the object x.

Table 4.2

Car	∂_{AT}
1	{Good}
2	{Good}
3	{Poor, good, excellent}
4	{Poor, good, excellent}
5	{Poor, good, excellent}
6	{Poor, good, excellent}

The above definitions can be formulated also by using $|nm_x|$ as follows.
An object x where $x \in U$ is called

$$\begin{aligned} &\text{a well defined object of } A \subset AT \quad \text{iff } |nm_x| \geq \frac{N}{2} \text{ if } N \text{ even or } |nm_x| \geq \frac{N+1}{2} \text{ if } N \text{ odd} \\ &\text{a poorly defined object of } A \subset AT \text{ iff } |nm_x| < \frac{N}{2} \text{ if } N \text{ even or } |nm_x| < \frac{N+1}{2} \text{if } N \text{ odd,} \end{aligned} \tag{4.3}$$

where N is the number of the attributes in the set A, for example, $|A| = N = |m_x| + |nm_x|$.

Now we are in a position to give the notion of *modified similarity relation (MSIM)* as follows.

5. The modified similarity relation

The modified similarity relation (MSIM) can be defined as follows:

(i) $(x,x) \in \text{MSIM}(A)$ where $A \subset AT$, for all $x \in U$;

(ii) $(x,y) \in \text{MSIM}(A)$ where $A \subset AT$, $N = |A| \geq 2$ if and only if

(a)

$$\rho_x^a = \rho_y^a \quad \forall a \in A \text{ where } \rho_x^a,\ \rho_y^a \text{ are defined values,} \tag{5.1}$$

(b)

$$EP \geq \begin{cases} \dfrac{N}{2} & \text{if } N \text{ even,} \\ \dfrac{N+1}{2} & \text{if } N \text{ odd.} \end{cases} \tag{5.2}$$

To make use of the MSIM, we need to add more definitions as follows.

Definition 5.1. $MS_A(x)$ denotes that the set $\{y \in U : (x,y) \in \text{MSIM}(A),\ A \subset AT\}$ is the maximal set of objects which are possibly indiscernible by A with x.

Table 5.1

Car	∂_{AT}
1	{Good}
2	{Good}
3	{Poor}
4	{Good, excellent}
5	{Good, excellent}
6	{Good}

Definition 5.2. The generalized decision in DT is defined by

$$\partial_A(x) = \{i : i = d(y),\ y \in MS_A(x),\ A \subset AT\}. \tag{5.3}$$

Definition 5.3. The modified discernibility matrix. The elements in the discernibilty matrix can be defined as follows:

$$\alpha_A(x,y) = \begin{Bmatrix} a \in A' & \text{where } A' = A - \{a \in A \text{ s.t. } \rho_x^a = \rho_y^a\} \\ \phi & \text{if } (x,y) \in \text{MSIM}(A),\ A \subset AT \end{Bmatrix}. \tag{5.4}$$

Remarks.

Remark 5.4. If $N = 1$, then

$$\begin{aligned} (x,y) \in\ & \text{MSIM}(A) \Longleftrightarrow \rho_x^a = \rho_y^a, \quad \rho_x^a, \rho_y^a \text{ are defined values,} \\ & \text{if } \rho_x^a = *, \text{ then the object } x \text{ will be similar to itself only.} \end{aligned} \tag{5.5}$$

Remark 5.5. If all condition attribute values for any object are missing, then, according to definition of MSIM, it will be similar to itself only. This is sensible because we cannot determine which object in the universe is similar to the object which has all its attributes missing except the object itself.

Remark 5.6. From the second condition of MSIM, we can conclude that the number of defined values must be greater than or equal to the number of missing values for each object in any ordered pair belongs to MSIM(A), $A \subset AT$.

Formally $|nm_x| \geq |m_x|$ and $|nm_y| \geq |m_y|$ for all $(x,y) \in \text{MSIM}(A)$, $A \subset AT$.

Remark 5.7. Any pair of objects which does not satisfy conditions 1 or 2 in the definition of MSIM will not be similar to each other. For example, in Table 4.1

$$\begin{aligned} (5,6) &\notin \text{MSIM}(AT) \quad \text{where } EP = 1; \\ (1,2) &\notin \text{MSIM}(AT) \quad \text{where } \rho_1^{\text{price}} \neq \rho_2^{\text{price}}. \end{aligned} \tag{5.6}$$

To achieve the new results, one applies the above definitions on Table 4.1 and the results are as follows:

$$\begin{aligned}
\mathrm{MSIM}(AT) = \{&(1,1),(2,2),(2,6),(3,3),(4,5),(5,4),(4,4),(5,5),(6,2),(6,6)\},\\
&S_{AT}(1) = \{1\},\\
&S_{AT}(2) = \{2,6\},\\
&S_{AT}(3) = \{3\},\\
&S_{AT}(4) = S_{AT}(5) = \{4,5\},\\
&S_{AT}(6) = \{2,6\}.
\end{aligned} \tag{5.7}$$

The generalized decisions will be as in Table 5.1.

We can then list all decision rules for DT before reduction as follows:

(r_1) $(p,\mathrm{high}) \wedge (M,\mathrm{high}) \wedge (S,\mathrm{full}) \wedge (X,\mathrm{low}) \to (d,\mathrm{good})$,

(r_2) $(p,\mathrm{low}) \wedge (M,*) \wedge (S,\mathrm{full}) \wedge (X,\mathrm{low}) \to (d,\mathrm{good})$,

(r_3) $(p,*) \wedge (M,*) \wedge (S,*) \wedge (X,\mathrm{high}) \to (d,\mathrm{poor})$,

(r_4) $(p,\mathrm{high}) \wedge (M,*) \wedge (S,\mathrm{full}) \wedge (X,\mathrm{high}) \to (d,\mathrm{good}) \vee (d,\mathrm{excellent})$,

(r_5) $(p,*) \wedge (M,*) \wedge (S,\mathrm{full}) \wedge (X,\mathrm{high}) \to (d,\mathrm{good}) \vee (d,\mathrm{excellent})$,

(r_6) $(p,\mathrm{low}) \wedge (M,\mathrm{high}) \wedge (S,\mathrm{full}) \wedge (X,*) \to (d,\mathrm{good})$.

6. Reduction

The minimal reducts for the incomplete information system IIS are as follows: a set $A \subset AT$ is reduct of IIS if and only if

$$\mathrm{MSIM}(A) = \mathrm{MSIM}(AT), \quad \forall B \subset A, \qquad \mathrm{MSIM}(B) \neq \mathrm{MSIM}(AT). \tag{6.1}$$

The minimal reduct for DT (the decision table) was defined as follows: a set $A \subset AT$ is reduct of DT if and only if

$$\partial_A(x) = \partial_{AT}(x), \quad \forall B \subset A, \qquad \partial_B(x) \neq \partial_{AT}(x). \tag{6.2}$$

Also, the idea of the so-called discernibility functions can be used. The discernibility function is a Boolean function representation of discernibility matrix (5.4). The reducts can be determined uniquely as follows:

Δ is a discernibility function for IIS if and only if

$$\Delta = \wedge\{\vee \alpha_{AT}(x,y),\ (x,y) \in U \times U,\ \alpha_{AT}(x,y) \neq \phi\}; \tag{6.3}$$

$\Delta(x)$ is a discernibility function for the object x in IIS if and only if

$$\Delta(x) = \wedge\{\vee \alpha_{AT}(x,y),\ y \in U,\ \alpha_{AT}(x,y) \neq \phi\}; \tag{6.4}$$

Δ^* is a discernibility function for DT if and only if

$$\Delta^* = \wedge\{\vee \alpha_{AT}(x,y) : (x,y) \in U \times \{z \in U,\ d(z) \notin \partial_{AT}(x)\},\ \alpha_{AT}(x,y) \neq \phi\}; \tag{6.5}$$

Table 6.1

	1	2	3	4	5	6
1	ϕ	$\{P,M\}$	$\{P,M,S,X\}$	$\{M,X\}$	$\{P,M,X\}$	$\{P,X\}$
2	$\{P,M\}$	ϕ	$\{P,M,S,X\}$	$\{P,M,X\}$	$\{P,M,X\}$	ϕ
3	$\{P,M,S,X\}$	$\{P,M,S,X\}$	ϕ	$\{P,M,S\}$	$\{P,M,S\}$	$\{P,M,S,X\}$
4	$\{M,X\}$	$\{P,M,X\}$	$\{P,M,S\}$	ϕ	ϕ	$\{P,M,X\}$
5	$\{P,M,X\}$	$\{P,M,X\}$	$\{P,M,S\}$	ϕ	ϕ	$\{P,M,X\}$
6	$\{P,X\}$	ϕ	$\{P,M,S,X\}$	$\{P,M,X\}$	$\{P,M,X\}$	ϕ

Table 6.2

	1	2	3	4	5	6
1	—	—	$\{P,M,S,X\}$	—	$\{P,M,X\}$	—
2	—	—	$\{P,M,S,X\}$	—	$\{P,M,X\}$	—
3	$\{P,M,S,X\}$	$\{P,M,S,X\}$	—	$\{P,M,S\}$	$\{P,M,S\}$	$\{P,M,S,X\}$
4	—	—	$\{P,M,S\}$	—	—	—
5	—	—	$\{P,M,S\}$	—	—	—
6	—	—	$\{P,M,S,X\}$	—	$\{P,M,X\}$	—

$\Delta^*(x)$ is a discernibility function for the object x in DT if and only if

$$\Delta^*(x) = \wedge\{\vee \alpha_{AT}(x,y) : y \in \{z \in U,\ d(z) \notin \partial_{AT}(x)\},\ \alpha_{AT}(x,y) \neq \phi\}. \tag{6.6}$$

To construct a discernibility function for the mentioned example, we will use Table 6.1, in which the elements of discernibility matrix $\alpha_{AT}(x,y)$ for any pair (x,y) of objects from U are placed.

From Table 6.1, the reducts for IIS are

$$\begin{aligned}
\Delta &= M \wedge (P \vee X),\\
\Delta_1 &= M \wedge (P \vee X),\\
\Delta_2 &= P \vee M,\\
\Delta_3 &= P \vee M \vee S,\\
\Delta_4 &= (M \vee X) \wedge (P \vee M \vee S),\\
\Delta_5 &= (P \vee M) \vee (X \wedge S),\\
\Delta_6 &= P \vee X.
\end{aligned} \tag{6.7}$$

The reducts for DT can be calculated from Table 6.2 in which values of $\alpha_{AT}(x,y)$ for any pair (x,y) of objects such that

$$x \in U, \quad y \in \{z \in U,\ d(z) \notin \partial_{AT}(x)\} \tag{6.8}$$

are shown.

Table 6.3

SIM						
Objects	1	2	3	4	5	6
The generalized decisions	{g}	{g}	{p,g,e}	{p,g,e}	{p,g,e}	{p,g,e}
Reducts for DT by discernibility function	x	x	Cannot be determined	Cannot be determined	Cannot be determined	Cannot be determined
Reducts for DT by generalized decisions	x	x	$p_r \vee m \vee s \vee x$	$p_r \vee m \vee s \vee x$	$p_r \vee m \vee s \vee x$	$p_r \vee m \vee s \vee x$
MSIM						
Objects	1	2	3	4	5	6
The generalized decisions	{g}	{g}	{p}	{g,e}	{g,e}	{g}
Reducts for DT by discernibility function	$p_r \vee m \vee x$	$p_r \vee m \vee x$	$p_r \vee m \vee s$	$p_r \vee m \vee s$	$p_r \vee m \vee s$	$p_r \vee m \vee x$
Reducts for DT by generalized decisions	$p_r \vee m \vee x$	$p_r \vee m \vee x$	$p_r \vee m \vee s$	s	s	$p_r \vee m \vee x$

Hence we have

$$\begin{aligned} \Delta^* &= (P \vee M \vee X) \wedge (P \vee M \vee S) = (P \vee M) \vee (X \wedge S), \\ \Delta_1^* = \Delta_2^* = \Delta_6^* &= P \vee M \vee X, \qquad \Delta_3^* = \Delta_4^* = \Delta_5^* = P \vee M \vee S. \end{aligned} \tag{6.9}$$

The results between the classical similarity SIM and the modified similarity relation MSIM are compared and seen in Table 6.3 for simplicity as follows:

(i) *for decisions*: "g" for good,"p" for poor, "e" for excellent,

(ii) *for attributes*: "p_r" for price, "m" for mileage, "s" for size, and "x" for max speed.

Reduction of knowledge that preserves generalized decisions for all objects in DT is less from decision making standpoint.

Accordingly, the decision rules become

(r_1') $(p,\text{high}) \wedge (M,\text{high}) \wedge (X,\text{low}) \to (d,\text{good})$,

(r_2') $(p,\text{low}) \wedge (M,*) \wedge (X,\text{low}) \to (d,\text{good})$,

(r_3') $(p,*) \wedge (M,*) \wedge (S,*) \to (d,\text{poor})$,

(r_4') $(p,\text{high}) \wedge (M,*) \wedge (S,\text{full}) \to (d,\text{good}) \vee (d,\text{excellent})$,

(r_5') $(p,*) \wedge (M,*) \wedge (S,\text{full}) \to (d,\text{good}) \vee (d,\text{excellent})$,

(r_6') $(p,\text{low}) \wedge (M,\text{high}) \wedge (X,*) \to (d,\text{good})$.

We can combine the rules which have the same generalized decisions as follows:

(r_1'') $(p,\text{high} \vee \text{low}) \wedge (M,\text{high}^*) \wedge (X,\text{low}^*) \to (d,\text{good})$,

Table 7.1

Decision class	Lower approximation	Upper approximation
$X_{\text{good}} = \{1,2,4,6\}$	$\{1,2,6\}$	$\{1,2,4,5,6\}$
$X_{\text{poor}} = \{3\}$	$\{3\}$	$\{3\}$
$X_{\text{excellent}} = \{5\}$	ϕ	$\{4,5\}$

Table 7.2

Generalized decision class	Lower approximation	Upper approximation
$X_{\{\text{good}\}} = \{1,2,6\}$	$\{1,2,6\}$	$\{1,2,6\}$
$X_{\{\text{poor}\}} = \{3\}$	$\{3\}$	$\{3\}$
$X_{\{\text{good,excellent}\}} = \{4,5\}$	$\{4,5\}$	$\{4,5\}$

(r_2'') $(p,\text{high}^*) \wedge (M,^*) \wedge (S,\text{full}) \rightarrow (d,\text{good}) \vee (d,\text{excellent})$,
(r_3'') $(p,^*) \wedge (M,^*) \wedge (S,^*) \rightarrow (d,\text{poor})$,
where the symbol "$*$" above the attribute value means that the attribute value $\vee$ is missing.

The third rule does not determine the values of the attributes P or M or X, then we can conclude that the poorly objects could not give a good result with respect to the decision rule; at least we can exclude the attributes which do not affect the decision.

7. Set approximations

The above procedure using the new approach can be used to deduce the set approximations as follows.

Let $X \subseteq U$ and $A \subset AT$. $\underline{A}X$ is the lower approximation of X if and only if

$$\underline{A}X = \{x \in U,\ MS_A(x) \subseteq X\} = \{x \in X,\ MS_A(x) \subseteq X\}. \tag{7.1}$$

$\overline{A}X$ is the upper approximation of X if and only if

$$\overline{A}X = \{x \in U,\ MS_A(x) \cap X \neq \phi\} = \bigcup \{MS_A(x),\ x \in X\}. \tag{7.2}$$

$\underline{A}X$ is a set of objects that belongs to X, while $\overline{A}X$ is a set of objects that possibly belongs to X.

From Table 4.1, we have $U/\text{IND}(d) = \{X_{\text{good}}, X_{\text{poor}}, X_{\text{excellent}}\}$, where $X_{\text{good}} = \{1,2,4,6\}$, $X_{\text{poor}} = \{3\}$, $X_{\text{excellent}} = \{5\}$, and the lower and upper approximations for each decision class for AT as in Table 7.1.

From Table 4.2, we have $U/\text{IND}(\partial_{AT}) = \{X_{\{\text{good}\}}, X_{\{\text{poor}\}}, X_{\{\text{good, excellent}\}}\}$, where $X_{\{\text{good}\}} = \{1,2,6\}$, $X_{\{\text{poor}\}} = \{3\}$, $X_{\{\text{good, excellent}\}} = \{4,5\}$, and the lower and upper approximations for the generalized decision class of AT as in Table 7.2.

8. Conclusion

In this paper, we remarked that a rough set theory can be an effective tool to deal with the incomplete information system. It can clarify a very important notation in statistics, which is the scaling. The big difference between the results can be seen from Tables 4.2 and 5.1. This leads to the importance of rough set theory in case of the incomplete information systems and solving some problems in scaling in statistics. Also, reducing the decision rules and the decisions can be introduced in a general form. In the literature, similarity relation handled the missing values as null value; it can be similar to any value in the domain of the attribute values. This point of view causes many problems in the analysis.

Our work is introducing a new definition of similarity relation MSIM which depends on some important conditions concerning the number of missing values. The essential point of MSIM is making the generalized decisions which have taken over the whole set of attributes a reasonable combination of decisions, hence, the reducts can be computed easily and the generalized decision has a valuable meaning. Also, a discernibility matrix is defined, reducts for IIS and for DT are derived, and the decision rules are introduced. Finally, the set approximations are defined.

References

[1] P. Allison, *Missing Data*, Sage, Thousand Oaks, Calif, USA, 2001.

[2] Z. Pawlak, "Rough sets," *International Journal of Computer and Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[3] J. W. Grzymala-Busse, "On the unknown attribute values in learning from examples," in *Proceedings of the 6th International Symposium on Methodologies for Intelligent Systems (ISMIS '91)*, pp. 368–377, Springer, Charlotte, NC, USA, October 1991.

[4] R. Slowinski and J. Stefanowski, "Rough classification in incomplete information systems," *Mathematical and Computer Modelling*, vol. 12, no. 10-11, pp. 1347–1357, 1989.

[5] J. W. Grzymala-Busse and L. K. Goodwin, "A closest fit approach to missing attribute values in preterm birth data," in *Proceedings of the 7th Workshop on New Directions in Rough Sets, Data Mining and Granular-Soft Computing (RSFDGrC '99)*, A. Skowon and N. Zhong, Eds., vol. 1711 of *Lecture Notes in Artificial Intelligence*, pp. 405–413, Springer, Yamaguchi, Japan, November 1999.

[6] J. W. Grzymala-Busse and M. Hu, "A comparison of several approaches to missing attribute values in data mining," in *Proceedings of the 2nd International Conference on Rough Sets and New Trends in Computing (RSCTC '00)*, pp. 378–385, Springer, Banff, Canada, October 2000.

[7] S. Greco, B. Matarazzo, and R. Slowinski, "Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems," in *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC '99)*, A. Skowron and N. Zhong, Eds., vol. 1711 of *Lecture Notes in Artificial Intelligence*, pp. 146–157, Springer, Yamaguchi, Japan, November 1999.

[8] S. Greco, B. Matarazzo, and R. Slowinski, "Rough set processing of vague information using fuzzy similarity relations," in *Finite Versus Infinite: Contributions to an Eternal Dilemma*, C. S. Calude and G. Păun, Eds., Discrete Mathematics and Theoretical Computer Science (London), pp. 149–173, Springer, London, UK, 2000.

[9] J. Stefanowski and A. Tsoukiàs, "Incomplete information tables and rough classification," *Computational Intelligence*, vol. 17, no. 3, pp. 545–566, 2001.

[10] M. Nakata and H. Sakai, "Rough sets handling missing values probabilistically interpreted," in *Proceedings of the 10th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing (RSFDGrC '05)*, D. S'lezak, G. Wang, M. S. Szczuka, I. Düntsch, and Y. Yao, Eds., vol. 3641 of *Lecture Notes in Artificial Intelligence*, pp. 325–334, Springer, Regina, Canada, August-September 2005.

[11] M. Kryszkiewicz, "Rough set approach to incomplete information systems," *Information Sciences*, vol. 112, no. 1–4, pp. 39–49, 1998.

E. A. Rady: Institute of Statistical Studies and Research, Cairo University, Cairo 12613, Egypt
Email address: narady@asu-pharmacy.edu.eg

M. M. E. Abd El-Monsef: Department of Mathematics, Faculty of Science, Tanta University, Tanta 31527, Egypt
Email address: mme@dr.com

W. A. Abd El-Latif: Department of Mathematics, Faculty of Science, Tanta University, Tanta 31527, Egypt
Email address: wfanwar@yahoo.com