Rough Set Extension under Incomplete Information System with "?" Values

Ahmed Hamed Attia

Computer Science Department, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt, ahmed_hamed@ci.suez.edu.eg

Abstract: Classical rough set theory (RST) can't process incomplete information system (IIS) because it is based on an indiscernibility relation which is a kind of equivalent relation. In the literature a non-symmetric similarity relation based rough set model (NS-RSM) has been introduced as an extended model under IIS with "?" values directly. Unfortunately, in this model objects in the same similarity class are not necessarily similar to each other and may belong to different target classes. In this paper, a new inequivalent relation called Maximal Limited Consistent block relation (MLC) is proposed. The proposed MLC relation improves the lower approximation accuracy by finding the maximal limited blocks of indiscernible objects in IIS with "?" values. Maximal Limited Similarity rough set model (MLS) is introduced as an integration between our proposed relation (MLC) and NS-RSM. The resulted MLS model works efficiently under IIS with "?" values. Finally, an illustrative example is given to validate MLS model. Furthermore, approximation accuracy comparisons have been conducted among NS-RSM and MLS. The results of this work demonstrate that the MLS model outperform NS-RSM.

Keywords: Rough Set Theory; Incomplete information system; non-symmetric similarity relation; Tolerance relation; Maximal consistent block relation; limited tolerance relation

1. Introduction

Classical Rough Set Theory (RST) proposed by Pawlak[13] made a great success in processing and analyzing complete information systems characterized by uncertainty. The indiscernibility relation is reflexive, symmetric and transitive. However, the indiscernibility relation is considered rigid relation[7] because it is based on the assumption that all objects values on every attribute must be complete. Such an assumption, contrasts with several real-valued information systems situations where the information may be incomplete either because the attributes values are missing or because the attributes values are absent[4]. This limits the applicability of the classical RST in real-world applications where some of the attribute values are incomplete. The classical RST can deal with IIS using two strategies. The first is an indirect method that transforms an IIS into a complete information system according to some rules such as probability statistical methods; this is called data reparation[6,18]. However, this strategy changes the original information of IS. The second is a direct method that extends the basic concepts of the classical rough set theory in IIS by relaxing the requirement of indiscernibility relation of reflexivity, symmetry and transitivity[1,2,8-10,12,14,16,17,19-23].

The unknown values were categorized by Grzymala-Busse[5] as follows

1. The unknown values are "lost ?", this unknown value means that it is an absent value and can't be compared with any other values in the domain of the corresponding attribute.

2. The unknown values are "don't care *", this unknown value means that it is a missing value and can be compared with any other values in the domain of the corresponding attribute.

In recent years, many researchers working on the RST model extension by extending the indiscernibility relation to inequivalent relations that can process IISs directly. For example, Kryszkiewicz (1998); Kryszkiewicz (1999) proposed the tolerance relation that is reflexive and symmetric but not necessarily transitive to deal with "*" values. In this relation, objects that have no values in common are considered indistinguishable. This obviously is unreasonable case which limits the applicability of the tolerance relation. Stefanowski and Tsoukiàs (1999); Stefanowski and Tsoukiàs (2001) investigated the similarity relation of Skowron and Stepaniuk (1996) to be reflexive and transitive but not necessarily symmetric to deal with "?" values. This relation separates two objects that are very similar to each other but with little loss in the information, also objects in the same similarity class are not necessarily similar to each other and may belong to different target classes. This excludes some objects from the lower approximation of the target set. This decreases the cardinality of the lower approximation leading to unpromising results with respect to approximation accuracy. Wang (2002); Wang et al (2008) proposed the limited tolerance relation that is reflexive and symmetric but not necessarily transitive. However, limited tolerance relation has not differentiated the two types of unknown attribute values. Leung and Li (2003); Cheng et al (2007) proposed the maximal consistent block relation that is reflexive and symmetric but not necessarily transitive to deal with "*" values. Maximal consistent block relation describes the maximal collection of indistinguishable objects in the tolerance classes. Such relation achieves better approximation accuracy than that provided by tolerance relation, but it inherits the limitation of the tolerance relation where objects that have no values in common are considered indistinguishable.

Among the previous rough set extensions, the non-symmetric similarity relation can effectively locates the lower/upper approximations (Slowinski et al (2014)). Unfortunately, objects in the same similarity class are not necessarily similar to each other. The maximal consistent block relation provides promising approximation accuracy in IIS with "*" values (Liu and Shao (2014)) but unable to deal with IIS with "?" values because it depends on tolerance relation. In this paper, we extend the relative concept of maximal consistent block relation based on the limited tolerance relation instead of the tolerance relation to present the maximal limited consistent block relation (MLC). MLC relation can find the maximal limited blocks of indiscernible objects in IIS with "?" values. Furthermore, we integrate MLC relation with NS-RSM to introduce the Maximal Limited Similarity model (MLS). The resulted model improves the lower approximation accuracy than NS-RSM. An illustrative example is given to validate our MLS model. Moreover, two approximation accuracy comparisons have been conducted among NS-RSM and MLS. The results demonstrate that with regard to the approximation accuracy, MLS model outperform NS-RSM model

The rest of this paper is organized as follows. Some related RST extensions under IIS are reviewed in section 3. Section 4 presents the Maximal Limited Consistent block relation (MLC). In section 5, we propose the Maximal Limited Similarity (MLS) model as integration of NS-RSM relation with MLC relation. In section 6, two approximation accuracy comparisons among NS-RSM and MLS have been conducted. Section 7 concludes the paper.

2. Preliminaries

3. Some RST Extensions Under IIS

In this section, we review some related RST extensions under IIS with their issues.

3.1 Tolerance relation based rough set model

Kryszkiewicz (1998) proposed the tolerance relation to deal with "*" values as Definition 1.

Definition 1 Given *IIS* in which $A \subseteq AT$, the tolerance relation is given by

$$SIM(A) = \{(x,y) \in U \times U : \forall a \in A, fa(x) = fa(y) \text{ or } fa(x) = * \lor fa(y) = *\} \tag{1}$$

The tolerance classes given the tolerance relation are given by

$$S_A(x) = \{y \in U : (x,y) \in SIM(A)\} \tag{2}$$

The tolerance relation lower/upper approximation for the target set X are given respectively as follows

$$\underline{Apr}_A(X) = \{x \in U : S_A(x) \subseteq X\} = \{x \in X : S_A(x) \subseteq X\} \tag{3}$$

$$\overline{Apr}_A(X) = \{x \in U : S_A(x) \cup X \neq \varphi\} = \cap \{S_A(x) : x \in X\} \tag{4}$$

3.2 Maximal consistent block relation

Leung and Li (2003) proposed the maximal consistent block relation to improve the results of the tolerance relation. Maximal consistent block describes the maximal collection of the indistinguishable objects in the tolerance class as Definition 2.

Definition 2. Given *IIS* in which $A \subseteq AT$ and $X \subseteq U$, we say that X is consistent with respect to A if $(x,y) \in SIM(A)$ for all $x,y \in X$. We say X is maximal consistent block of A if there doesn't exist $Y \subseteq U$ where $X \subset Y$ and Y is consistent with respect to A. We denote the maximal consistent blocks with respect to A as $\zeta(A)$ and the maximal consistent blocks with respect to A that includes some object $x \in U$ as $\zeta_x(A)$.

The maximal consistent block lower/upper approximations for the target set X are given respectively as follows

$$\underline{Apr}_A(X) = \cap\{Y \in \zeta(A): Y \subseteq X\}, \tag{5}$$
$$\overline{Apr}_A(X) = \cap\{Y \in \zeta(A): Y \cup X \neq \varphi\}. \tag{6}$$

3.3 Non-symmetric similarity relation based rough set model (NS-RSM)

Definition 3. Given *IIS* in which $A \subseteq AT$, the non-symmetric similarity relation is given by

$$\forall_{x,y} S(x,y) \Leftrightarrow \forall_{a \in A} f_a(x) \neq ? , f_a(x) = f_a(y)\} \tag{7}$$

For each object, two similarity classes are defined as follows

$$R(x) = \{y \in U: S(y,x)\} \tag{8}$$
$$R^{-1}(x) = \{y \in U: S(x,y)\} \tag{9}$$

Where, $R(x)$ represents the set of objects similar to x and $R^{-1}(x)$ represents the set of objects to which x is similar. The lower/upper approximations of set X are given respectively as follows

$$\underline{Apr}_A(X) = \{x \in U: R^{-1}(x) \subseteq X\} \tag{10}$$
$$\overline{Apr}_A(X) = \cap\{R(x): x \in X\} \tag{11}$$

3.4 Limited tolerance relation

The tolerance relation requirement is too weak as it regards two objects with no common values as indistinguishable, also, the requirement of the non-symmetric similarity relation is too strict as it separates two objects that are very similar to each other but with little loss in the information. This makes the process too extreme. Consequently, Wang (2002); Wang et al (2008) proposed the limited tolerance relation that tries to relax the requirements of both tolerance relation and non-symmetrical similarity relation as Definition 4.

Definition 4. Given *IIS* in which $A \subseteq AT$, the limited tolerance relation is given by

$$\forall_{x,y \in U \times U} (LTA(x,y) \Leftrightarrow \forall_{a \in A}(f_a(x) = f_a(y) = \text{unknown})$$
$$\vee \ ((P_A(x) \cap P_A(y) \neq \varnothing) \tag{12}$$
$$\wedge \ \forall_{a \in A} \ (f_a(x) \neq unknown$$
$$\wedge \ f_a(y) \ ) \neq unknown)$$
$$(f_a(x) = f_a(y))))$$

where $P_A(x) = \{a \in A: f_a(x) \text{ is known value}\}$ and the unknown values can be interpreted as * or ? values. The limited tolerance classes of x is denoted by $LT_A(x)$ where

$$I_A(x) = \{y: y \in U \wedge LT_A(x,y)\} \tag{13}$$

The lower and upper approximations of set X are given respectively as follows

$$\underline{Apr}_A(X) = \{x: x \in U \wedge I_A(x) \subseteq X\} \tag{14}$$
$$\overline{Apr}_A(X) = \{x: x \in U \wedge I_A(x) \cup X \neq \varphi\} \tag{15}$$

4 MAXIMAL LIMITED CONSISTENT BLOCK RELATION (MLC)

Maximal consistent block relation deals only with IIS with "*" values in which it depends only on tolerance relation as shown in section 3.2. In this section we propose the maximal limited consistent block relation (MLC) that extends the maximal consistent block relation based on the limited tolerance relation instead of tolerance relation. This is because limited tolerance relation can deal with IIS with either "*" or "?" values. MLC describes the maximal collection of

indistinguishable objects in the limited tolerance classes. This allows MLC to provide a promising approximation accuracy than the one obtained by limited tolerance relation. The form of MLC relation is given in Definition 5.

Definition 5. Given *IIS* in which $A\subseteq AT$ and $X\subseteq U$, we say that X is limited consistent with respect to A if $(x,y)\in LT_A(x,y)$ $\forall x,y\in X$. We say that X is maximal limited consistent block of A if there doesn't exist $Y\subseteq U$ where $X\subset Y$ and Y is limited consistent with respect to A. We denote the maximal limited consistent blocks with respect to A as $\zeta^{LT}(A)$ and the maximal limited consistent blocks with respect to A that includes some objects $x\in U$ as $\zeta_x^{LT}(A)$

Such relation is reflexive but not necessarily transitive and symmetric.

In the following, to obtain the lower/upper approximations of MLC, we will redefine the lower/upper approximation obtained by equations (5) and (6) based on Definition 6 as follows

$$\underline{Apr}_A(X)=\cap\{Y\in\zeta^{LT}(A):Y\subseteq X\} \tag{16}$$

$$\overline{Apr}_A(X)=\cap\{Y\in\zeta^{LT}(A):Y\cup X\neq\varphi\} \tag{17}$$

The positive, negative and the boundary regions are given respectively as follows

$$POS(X)=\underline{Apr}_A(X), \tag{18}$$

$$NEG(X)=U-\overline{Apr}_A(X), \tag{19}$$

$$BND(X)=\overline{Apr}_A(X)-\underline{Apr}_A(X) \tag{20}$$

5 MAXIMAL LIMITED SIMILARITY ROUGH SET MODEL (MLS)

As we discussed in section 3.3, objects in the same similarity classes are not necessarily similar to each other and may belong to different target classes which increases uncertainty in the data. This excludes some objects from the lower approximation of the target set in spite the fact they could be classified in the lower approximation leading to unpromising results with respect to approximation accuracy. Our proposed solution for the previous problem is to find the maximal limited blocks of similar objects in similarity classes to ensure that objects in the same similarity class are similar to each other. This reduces the uncertainty from the data allowing accurate computation to be done for the lower approximation leading to promising results with respect to approximation accuracy. In this section we introduce the Maximal Limited Similarity (MLS) model as an integration between NS-RSM and our proposed MLC relation. The resulted model able to provide a promising approximation accuracy than NS-RSM. The form of MLST is given in Definition 7 and Definition 8.

Definition 7. Given *IIS* and $R(x)$ be set of objects similar to x in which $A\subseteq AT$ and $X\subseteq R(x)$, we say that X is limited consistent block of objects similar to x with respect to A if $(x,y)\in S(y,x)$ and $\forall y,z\in X,(y,z)\in LTA(y,z)$. We say that X is maximal limited consistent block of objects similar to x if there doesn't exist $Y\subseteq R(x)$ where $X\subset Y$ and Y is limited consistent with respect to A. We denote the maximal limited consistent blocks of all $R(x)$ as $\zeta_R^{LT}(A)$ and the maximal limited consistent blocks of objects similar to x with respect to A as $\zeta_{R(x)}^{LT}(A)$.

Definition 8. Given IIS and R−1(x) be set of objects to which x is similar in which $A\subseteq AT$ and $X\subseteq R-1(x)$, we say that X is limited consistent block of objects to which x is similar with respect to A if $(x,y)\in S(x,y)$ and $\forall y,z\in X,(y,z)\in LTA(y,z)$. We say that X is maximal limited consistent block of objects to which x is similar if there doesn't exist $Y\subseteq R-1(x)$ where $X\subset Y$ and Y is limited consistent with respect to A. We denote the maximal limited consistent blocks of all R^{-1} as $\zeta_{R-1}^{LT}(A)$ and the maximal limited consistent blocks of objects to which x is similar with respect to A as $\zeta_{R-1(x)}^{LT}(A)$.

Such relation is reflexive but not necessarily transitive and symmetric.

In the following, to obtain the lower/upper approximations of MLS, we will redefine the lower/upper approximation

obtained by equations 10 and 11 based on Definitions 7 and 8 as follows

$$\underline{Apr}_A(X)=\cap\{Y\in\zeta_{R^{-1}}^{LT}(A):Y\subseteq X\} \tag{21}$$

$$\overline{Apr}_A(X)=\cap\{Z\in\zeta_{R(x)}^{LT}(A):x\in X\} \tag{22}$$

Similar to the classical RST, the positive, negative and the boundary regions are given respectively as follows

$$POS(X)=\underline{Apr}_A(X), \tag{23}$$

$$NEG(X)=U-\overline{Apr}_A(X), \tag{24}$$

$$BND(X)=\overline{Apr}_A(X)-\underline{Apr}_A(X). \tag{25}$$

5.1 Properties of MLS

The tolerance relation requirement is too weak as it regards Property 1. Any similarity class $R^{-1}(x)$ of attributes subset A can be represented as the union of maximal limited consistent blocks included in it. In other words

$$R^{-1}(x)=\cap\{Y\in\zeta_{R^{-1}}^{LT}(A):Y\subseteq R^{-1}(x)\}=\cap\{Y\in\zeta_{R^{-1}(x)}^{LT}(A)\}.$$

Property 2. Any similarity class $R(x)$ of attributes subset A can be represented as the union of maximal limited consistent blocks included in it. In other words

$$R(x)=\cap\{Z\in\zeta_{R}^{LT}(A):Z\subseteq R(x)\}=\cap\{Z\in\zeta_{R(x)}^{LT}(A)\}.$$

Property 3. Given *IIS* in which $A,B\subseteq AT$ and $X\subseteq U$, then

1. $\underline{Apr}_A(X)\subseteq X\subseteq\overline{Apr}_A(X).$

2. $A\subseteq B$ $\overline{Apr}_B(X)\subseteq\overline{Apr}_A(X)$. But $\underline{Apr}_A(X)\subseteq\underline{Apr}_B(X)$ doesn't hold.

Theorem 1 Given IIS, the lower approximation of $X\subseteq U$ obtained using MLS model is a refinement of the one obtained using NS-RSM.

Proof. We have to clarify that the lower approximation of NS-RSM obtained using equation 21 is subset of lower approximation of MLS obtained using equation 10.

Suppose that $x\in$ lower approximation of the NS-RSM obtained using equation 10, then $R^{-1}(x)\subseteq X$ holds. By property 1, $R^{-1}(x)=\cap\{Y\in\zeta_{R^{-1}}^{LT}(A):Y\subseteq R^{-1}(x)\}=\cap\{Y\in\zeta_{R^{-1}(x)}^{LT}(A)\}$. Then there exist $Y\in\zeta_{R^{-1}(x)}^{LT}$ that contain x and $Y\subseteq X$. Therefore, $x\in$ lower approximation of MLS obtained using equation 21. This means that $\forall x\in$ lower approximation of NS-RSM, $x\in$ lower approximation of MLS. The inverse isn't necessarily true. Consequently, the lower approximation of X obtained using MLS is at least equal to the lower approximation of X obtained using NS-RSM.

The investigation of NS-RSM using MLC allowed accurate computation to be done for the lower approximation when the data set contains a lot of uncertainty, this is because MLC reduces the uncertainty from data. When the data set contains little uncertainty which is considered the worst case to MLS, the lower approximation obtained using MLS is at least equal to the lower approximation obtained using NS-RSM.

Theorem 2 Given IIS, the upper approximation of $X\subseteq U$ obtained using MLS relation is equal to the one obtained using NS-RSM.

Proof. The upper approximation of NS-RSM is given by

$\overline{Apr}_A(X)=\cap\{R(x):x\in X\}$. From property 2, $R(x)=\cap\{Z\in\zeta_{R}^{LT}(A):Z\subseteq R(x)\}=\cap\{Z\in\zeta_{R(x)}^{LT}(A)\}$ for any $x\in U$. So,

$$\overline{Apr}_A(X)=\cap\{\cap\{Z\in\zeta_{R(x)}^{LT}(A)\}:x\in X\}=\cap\{Z\in\zeta_{R(x)}^{LT}(A):x\in X\}=\text{ upper approximation of MLC}$$

6 EXPERIMENTAL RESULTS

The tolerance relation requirement is too weak as it regards This section presents an illustrative example to validate MLS model. Furthermore, approximation accuracy comparisons have been conducted among NS-RSM and MLS. We have implemented our proposed model and the related model using Matlab R12a on a PC with windows 8, Intel(R) Core(TM) i7 CPU 2.4 GHZ and 6GB memory. In this experiment we employ the incomplete data set shown in Table 1 Stefanowski and Tsoukiàs (1999) to illustrate and compare the approximation accuracy of MLS with NS-RSM.

The metric for evaluating the performance of different rough set models is the approximation accuracy (Dai and Xu (2012)) of a set that is a type of uncertainty measurements that measures the quality of the data (Slowinski et al (2014)). The approximation accuracy is an agreement measure of completeness of the knowledge of the approximated set and is given by

$$\tau=\frac{|Apr(A)|}{|\overline{Apr}(A)|}$$

where, $|\cdot|$ means cardinality and $\underline{Apr}(A)$, $\overline{Apr}(A)$ means the lower and upper approximations respectively. Obviously, $0\leq\tau\leq1$. This means, smaller the boundary region the higher the approximation accuracy.

U	a_1	a_2	a_3	a_4	d
x_1	3	2	1	?	1
x_2	2	3	2	?	1
x_3	2	3	2	0	2
x_4	?	2	?	1	1
x_5	?	2	?	1	2
x_6	2	3	2	1	1
x_7	3	?	?	3	1
x_8	2	3	2	2	2
x_9	3	2	1	3	2
x_{10}	1	?	?	?	1
x_{11}	3	2	?	?	2
x_{12}	3	2	1	1	1
x_{13}	3	2	2	3	2

Table 1: small illustrative example

This experiment uses the incomplete data set shown in Table 1 where, $U=\{x_1,x_2,...,x_{13}\}$, $A=\{a_1,a_2,a_3,a_4\}$ with 25% lost values and d is the decision attribute that determines a partition on the universe such that $U/d=\{d_1,d_2\}=\{\{x\in U:f_d(x)=1\},\{x\in U:f_d(x)=2\}\}$. Then

(1) We obtain the similarity classes in terms of NS-RSM as follows

$$R^{-1}(x_1)=\{x_1,x_9,x_{12}\} \qquad R(x_1)=\{x_1,x_{11}\}$$

$$R^{-1}(x_2)=\{x_2,x_3,x_6\} \qquad R(x_2)=\{x_2\}$$

$$R^{-1}(x_3)=\{x_3\} \qquad\qquad R(x_3)=\{x_2,x_3\}$$

$$R^{-1}(x_4)=\{x_4,x_5,x_{12}\} \qquad\qquad R(x_4)=\{x_4,x_5\}$$

$$R^{-1}(x_5)=\{x_4,x_5,x_{12}\} \qquad\qquad R(x_5)=\{x_4,x_5\}$$

$$R^{-1}(x_6)=\{x_6\} \qquad\qquad R(x_6)=\{x_2,x_6\}$$

$$R^{-1}(x_7)=\{x_7,x_9,x_{13}\} \qquad\qquad R(x_7)=\{x_7\}$$

$$R^{-1}(x_8)=\{x_8\} \qquad\qquad R(x_8)=\{x_2,x_8\}$$

$$R^{-1}(x_9)=\{x_9\} \qquad\qquad R(x_9)=\{x_1,x_7,x_9,x_{11}\}$$

$$R^{-1}(x_{10})=\{x_{10}\} \qquad\qquad R(x_{10})=\{x_{10}\}$$

$$R^{-1}(x_{11})=\{x_1,x_9,x_{11},x_{12},x_{13}\} \qquad\qquad R(x_{11})=\{x_{11}\}$$

$$R^{-1}(x_{12})=\{x_{12}\} \qquad\qquad R(x_{12})=\{x_1,x_4,x_5,x_{11},x_{12}\}$$

$$R^{-1}(x_{13})=\{x_{13}\} \qquad\qquad R(x_{13})=\{x_7,x_{11},x_{13}\}$$

Consequently,

$$\underline{S_A}(d_1)=\{x_6,x_{10},x_{12}\}$$

$$\overline{S_A}(d_1)=\{x_1,x_2,x_4,x_5,x_6,x_7,x_{10},x_{11},x_{12}\}$$

$$\underline{S_A}(d_2)=\{x_3,x_8,x_9,x_{13}\}$$

$$\overline{S_A}(d_2)=\{x_1,x_2,x_3,x_4,x_5,x_7,x_8,x_9,x_{11},x_{13}\}$$

Obviously, $\tau(d_1)=\dfrac{3}{9}=\dfrac{1}{3}$ and $\tau(d_2)=\dfrac{4}{10}=\dfrac{2}{5}$. Some objects are included in the same relation in spite the fact that they are not similar to each other. Such problem is a result of NS-RSM does not find the maximal block of indiscernible objects. For example, in $R^{-1}(x_1)=\{x_1,x_9,x_{12}\}$ object x_9 and object x_{12} are not similar, this lead to $R^{-1}(x_1)\nsubseteq d_1$ which prevents object x_1 from being included in the lower approximation of d_1. In $R^{-1}(x_2)=\{x_2,x_3,x_6\}$ object x_3 and object x_6 are not similar, this lead to $R^{-1}(x_2)\nsubseteq d_1$ which prevents object x_2 from being included in the lower approximation of d_1. This shrinks the lower approximation leading to unpromising results with respect to approximation accuracy.

(2) We obtain the similarity classes in terms of MLS as follows

$$\zeta^{LT}_{R(x_1)}(A) = \{Z_1 = \{x_1, x_{11}\}\}$$

$$\zeta^{LT}_{R^{-1}(x_1)}(A) = \{Y_1 = \{x_1, x_9\}, Y_2 = \{x_1, x_{12}\}\}$$

$$\zeta^{LT}_{R(x_2)}(A) = \{Z_2 = \{x_2\}\}$$

$$\zeta^{LT}_{R^{-1}(x_2)}(A) = \{Y_3 = \{x_2, x_3\}, Y_4 = \{x_2, x_6\}, Y_5 = \{x_2, x_8\}\}$$

$$\zeta^{LT}_{R(x_3)}(A) = \{Z_3 = \{x_2, x_3\}\}$$

$$\zeta^{LT}_{R^{-1}(x_3)}(A) = \{Y_6 = \{x_3\}\}$$

$$\zeta^{LT}_{R(x_4)}(A) = \{Z_4 = \{x_4, x_5\}\}$$

$$\zeta^{LT}_{R^{-1}(x_4)}(A) = \{Y_7 = \{x_4, x_5\ x_{12}\}\}$$

$$\zeta^{LT}_{R(x_5)}(A) = \{Z_4 = \{x_4, x_5\}\}$$

$$\zeta^{LT}_{R^{-1}(x_5)}(A) = \{Y_7 = \{x_4, x_5, x_{12}\}\}$$

$$\zeta^{LT}_{R(x_6)}(A) = \{Z_5 = \{x_2, x_6\}\}$$

$$\zeta^{LT}_{R^{-1}(x_6)}(A) = \{Y_8 = \{x_6\}\}$$

$$\zeta^{LT}_{R(x_7)}(A) = \{Z_6 = \{x_7\}\}$$

$$\zeta^{LT}_{R^{-1}(x_7)}(A) = \{Y_9 = \{x_7, x_9\}, Y_{10} = \{x_7, x_{13}\}\}$$

$$\zeta^{LT}_{R(x_8)}(A) = \{Z_7 = \{x_2, x_8\}\}$$

$$\zeta^{LT}_{R^{-1}(x_8)}(A) = \{Y_{11} = \{x_8\}\}$$

$$\zeta^{LT}_{R(x_9)}(A) = \{Z_8 = \{x_1, x_7, x_9, x_{11}\}\}$$

$$\zeta^{LT}_{R^{-1}(x_9)}(A) = \{Y_{12} = \{x_9\}\}$$

$$\zeta^{LT}_{R(x_{10})}(A) = \{Z_9 = \{x_{10}\}\}$$

$$\zeta^{LT}_{R^{-1}(x_{10})}(A) = \{Y_{13} = \{x_{10}\}\}$$

$$\zeta^{LT}_{R(x_{11})}(A) = \{Z_{10} = \{x_{11}\}\}$$

$$\zeta^{LT}_{R(x_{12})^{-1}}(A) = \{Y_{17} = \{x_{12}\}\}$$

$$\zeta^{LT}_{R(x_{13})}(A) = \{Z_{14} = \{x_7, x_{11}, x_{13}\}\}$$

$$\zeta^{LT}_{R(x_{13})^{-1}}(A) = \{Y_{18} = \{x_{13}\}\}$$

Consequently,

$$\underline{S}_A(d_1) = \{x_1, x_2, x_6, x_{10}, x_{12}\}$$

$$\overline{S_A}(d_1) = \{x_1, x_2, x_4, x_5, x_6, x_7, x_{10}, x_{11}, x_{12}\}$$

$$\underline{S}_A(d_2) = \{x_3, x_8, x_9, x_{11}, x_{13}\}$$

$$\overline{S_A}(d_2) = \{x_1, x_2, x_3, x_4, x_5, x_7, x_8, x_9, x_{11}, x_{13}\}$$

Obviously, $\tau(d_1) = \frac{5}{9}$ and $\tau(d_2) = \frac{5}{10} = \frac{1}{2}$. The results are more informative than NS-RSM model. This is because

MLS finds the maximal limited consistent blocks of indiscernible objects and now objects in the same relation are similar to each other. For example, in $\zeta_R^{LT^{-1}}(x_1)(A)=\{Y_1=\{x_1,x_9\},Y_2=\{x_1,x_{12}\}\}$ object x_9 and object x_{12} are not included in the same block as NS-RSM does. This lead to $Y_2\subseteq d_1$ which allow object x_1 to be included in the lower approximation of d_1. In $\zeta_R^{LT^{-1}}(x_2)(A)=\{Y_3=\{x_2,x_3\},Y_4=\{x_2,x_6\},Y_5=\{x_2,x_8\}\}$ object x_3 and object x_6 are not in the same block. This lead to $Y_4\subseteq d_2$ which allow object x_2 to be included in the lower approximation of d_1.

7 CONCLUSION

In most real-valued information systems, attributes are incomplete to some degree which affect the process of analyzing data. Acquiring knowledge from such an IIS is difficult. So, researchers introduced many approaches to address this problem. Many solutions based on extending RST have been proposed. In this paper, non-symmetric similarity relation based rough set model (NS-RSM) has been investigated and analyzed. It was found that NS-RSM gathers objects that are not similar to each other in the same relation which results in unpromising results with respect to approximation accuracy. To address this problem, we present the maximal limited similarity model (MLS) that finds the maximal limited blocks of similar objects in similarity classes to ensure that objects in the same similarity class are similar to each other. This reduces the uncertainty from the data allowing accurate computation to be done for the lower approximation leading to promising results with respect to approximation accuracy. To demonstrate the efficiency of the proposed MLS model, comparisons of the approximation accuracy of the proposed model with NS-RSM have been carried out on four benchmark data sets. The results of the comparisons show that MLS is superior than NS-RSM. Also the comparisons demonstrate that MLS can obtain higher approximation accuracy than NS-RSM and can approximate the target set more accurately than NS-RSM. For further study, MLS model can be used to improve RST reduction.

References

1. Attia AH, Sherif AS, El-Tawel GS. Maximal limited similarity-based rough set model. Soft Computing 2016; 20(8):3153-3161.
2. Cheng YS, Zhang YS, Hu XG, Zhang YZ. Uncertainty measure of knowledge and rough set based on maximal consistency block technique. Proceedings of the Sixth International Conference on Machine Learning and Cybernetics 2007; Hong Kong, 3069-3074
3. Dai J, Xu Q. Approximations and uncertainty measures in incomplete information systems. Information Sciences 2012; 198:62-80.
4. DU WF, ZI T. Properties of lower and upper approximation operators under various kinds of relations. International Conference on Mechanics and Civil Engineering 2014; 472-477.
5. Grzymala-Busse JW. Characteristic relations for incomplete data: A generalization of the indiscernibility relation. In: Rough Sets and Current Trends in Computing, Proceeding of the Third International Conference on Rough Sets and Current Trends in Computing 2004; 244-253.
6. Grzymala-Busse JW, Hu M. A comparison of several approaches to missing attribute values in data mining. In: Lecture Notes in Computer Science, Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing. Lecture Notes in Computer Science 2005; Berlin:Springer, 378-385.
7. Huang J, Guan Y, Shen J, Wang H. Rough approximations in tolerance rough set models. 11th International Conference on Fuzzy Systems and Knowledge Discovery 2014; 61-66.
8. Kryszkiewicz M. Rough set approach to incomplete information system. Information Sciences 1998; 112:39-49.
9. Kryszkiewicz M. Rules in incomplete information systems. Information Sciences 1999; 113:271-292.
10. Leung Y, Li D. Maximal consistent block technique for rule acquisition in incomplete information systems. Information Sciences 2003; 153:85-106.
11. Liu X, Shao M. Approachs to computing maximal consistent block. Machine Learning and Cybernetics Communications in Computer and Information Science 2014; 481:264-274.
12. Nguyen DV, Yamada K, Unehara M. Extended tolerance relation to define a new rough set model in incomplete information systems. Hindawi Publishing Corporation Advances in Fuzzy Systems 2013.
13. Pawlak Z. Rough sets. International Journal of Information and Computer Sciences 1982; 11:341-356
14. Skowron A, Stepaniuk J Tolerance approximation space. Fundamenta Informaticae 1996; 27:245-253
15. Slowinski R, Greco S, Matarazzo B. Rough-set-based decision support. Springer Science - Search Methodologies 2014; 557-609.
16. Stefanowski J, Tsoukiàs A. On the extension of rough sets under incomplete information. In: Lecture Notes in Computer Science, 7th Intemational Workshop on New Directions in Rough Sets, Data Mining, and Granular-

Soft Computing 1999; 1711: 73-81.

17. Stefanowski J, Tsoukiàs A. Incomplete information tables and rough classification. Computational Intelligence 2001; 17:545-566.

18. Wang G. Rough set theory and knowledge acquisition. Xi'an: Xi'an Jiaotong University Press, 2001; 38-39.

19. Wang G. Extension of rough set under incomplete information systems *. Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the IEEE International Conference 2002; 2: 1098-1103.

20. Wang G, Guan L, Hu F. Rough set extensions in incomplete information systems. Frontiers of Electrical and Electronic Engineering in China 2008; 3:399-405.

21. Yang X. Difference relation-based rough set and negative rules in incomplete information system. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 2009; 18:649-665.

22. Yang X, Yang J. Expansions of rough sets in incomplete information systems. Science Press Beijing and Springer-Verlag Berlin Heidelberg 2012; 45-99.

23. Yin X, XiuyiJia, Shang L. A new extension model of rough sets under incomplete information. Springer Berlin Heidelberg 2006; 4062:141-146.