Applied Mathematics & Information Sciences
An International Journal

Rough Set Model Based on the Dual-limited Symmetric Similarity Relation

Yu-Ming Zhai[1], Rui-Xia Yan[2], Zhen-Yu Huang[3] and Bin Guo[4]

[1]School of Economics and Management, Shanghai Institute of Technology, Shanghai 201418 P. R. China
[2]School of Management, Shanghai University of Engineering Science, Shanghai 201620, P. R. China
[3]Shanghai International Studies University, Shanghai 200083, P. R. China
[4]School of Software, Jiangxi Normal University, Nanchang 330022, P. R. China

Received: 14 Nov. 2012, Revised: 30 Dec. 2012, Accepted: 15 Jan. 2013
Published online: 1 Feb. 2013

Abstract: Information systems are often incomplete in object world. This paper puts forward a new rough set model in incomplete information system. The concepts of comparable degree and reliability are proposed, and a dual-limited symmetric similarity relation is constructed. Then a new rough set model based on dual-limited symmetric similarity relation is designed, which determines the upper approximation, lower approximation and boundary. Meanwhile, classification granularity and accuracy of knowledge are studied. An example presented illustrates the effectiveness and practicality of the rough set model based on the dual-limited symmetric similarity relation.

Keywords: dual-limited symmetric similarity relation, rough set, decision making, incomplete information system

1 Introduction

Since rough set theory[1] was proposed, the theoretical models of rough set have gone through continuous improvement and development, and penetrated into many disciplines. Rough set has been the basis for the data mining, knowledge reduction and granular computing theory. As a mathematical tool of dealing with uncertain, imprecise and incomplete information, it has been widely used in the artificial intelligence and cognitive science, especially in the intelligent information processing and other fields.

However, the classical rough set theory is based on equivalent classification to research information systems. If the object' value in some attribute is not determined (i.e., a null value), rough set cannot be utilized to process these objects. The classical rough set theory cannot deal with null values in incomplete information system. In order to improve the capacity of rough set data processing, some classification and proposed similarity relation and tolerance relation instead of equivalence relation as the basis of rough set. Slowinski and Vanderpooten[2,3] put forward similarity relation, which meet the reflexive similarity. Similarity relation ignores minor differences in attribute values. Greco[4] put forward binary relation to meet transitivity for the analysis of incomplete information table. The extensions of the method maintained all characteristics of the classical model. When the information table has no null value, it is equivalent to the classical model. Kryszkiewicz[5,6] presented a similarity relation to satisfy both reflexivity and symmetry to acquire decision rules in completely information table. Skowron and Stepaniuk[7] presented a tolerance relation to satisfy reflexivity and symmetry. These models extend the application of rough set theory in the uncertain information processing.

Among them, rough set based on symmetric similarity relation instead of equivalence relation can effectively solve the problem of null value in incomplete information system and make the classification of knowledge in incomplete information system. There are some defects of symmetric similarity relation in the processing in incomplete information system. The knowledge partition granularity from rough set based on symmetric similarity relation is too coarse. Precision and accuracy of symmetric similarity classification are declined.

* Corresponding author e-mail: huangzy@shisu.edu.cn

Symmetric similarity relation regards that a null value of attribute can be equal to any known attribute value, which possibly causes the defects of less similarity between objects and low partition granularity. The defects are prone to the miscarriage of justice to a similar class in a false condition, which is bound to affect the application of rough set model based on symmetric similarity relation. For example, there are two objects $X = \{1,1,1,1,1,1\}$ and $Y = \{0,0,0,0,0,0\}$. Obviously, any attribute values of X and Y are not equal. Therefore, X and y do not satisfy the equivalence relation; neither do the symmetric similarity relation. However, for some reason, some attribute value of X and Y is missing, $X = \{1,1,1,*,*,*\}$ and $Y = \{*,*,*,0,0,0\}$ ("*" indicates a null value). According to Kryszkiewicz's [8] definition of similarity relation, then the symmetry X and Y is symmetric similarity and belong to the same symmetrical similarity class. This is obviously a miscarriage of justice. In some cases, the similarity degrees between objects vary greatly. Maybe some completely equal (i.e. equivalence relation). Maybe while some individual attribute values are equal, and most of the rest attributes are null, such as $X = \{1,2,1,0,1,0\}$, $Y0 = \{1,2,1,0,1,0\}$, $Y1 = \{1,2,1,0,1,*\}$, $Y2 = \{1,2,1,*,*,*\}$, $Y3 = \{1,*,*,*,*,*\}$, $Y4 = \{*,*,*,*,*,*\}$. Based on the definition of symmetric similarity relation, Y0, Y1, Y2, Y3 and Y4 similar symmetrically to X. But the similarity degrees are greatly different. Y0 and X are similar and satisfy the equivalence relation. Y4 and X have similarity relation for all attributes values of Y4 are null.

In order to make the knowledge classification in incomplete information systems based on rough set more accurate, the reference [8] studied the tolerance relation and non- symmetric similarity relation. And [8] pointed out that the requirement of tolerance relation (symmetric similarity relation) was too loose. In order to solve the problem, limited tolerance relation [8] was proposed and its related models were discussed. However, limited tolerance relation has not classified the unknown attribute values. To deal with two unknown attributes values in incomplete information system by utilizing rough set theory, Grzymala-Buse presented characteristic relation [9]. At present, the extensions of rough set model in incomplete information system have studied much and their applications have been popular in real life[10, 11].

When we classify knowledge utilizing symmetric similarity relations, we often face with these problems.

(1) If the objects compared contain of null value, the comparability declines, even not comparable.

(2) If some attribute value range is large, and the attribute values of the objects compared contain a null value in these properties comparison of the object (when at least one of the attribute values is a null value), it is necessary to study the reliability of the comparison.

This paper presents the concept of comparability and credibility to limit the symmetric similarity relations. In order to refine the granularity of knowledge classification and improve the accuracy of the classification, rough set model based on dual-limited symmetric similarity relation is built to determine the upper approximate, lower approximation and boundary. In order to gain a higher quality of classification, examples are put forward to verify the above theory. From two aspects, the theoretical proof and practice results verify the validity and practicality of the rough set model based on dual-limited symmetric similarity relation.

2 Basic Theory of Rough Set Theory

2.1 Rough Set Theory

Definition 2.1 [1] An information system is defined as $S = (U,A,V,F)$, where

(1) $U = \{x_1, x_2, \cdots, x_n\}$ is a set of objects of finite nonempty set, which is also called universe.

(2) A is a finite nonempty set attributes, subset C and D respectively are called condition attribute set and decision attribute set. $A = C \cup D$ and $C \cap D = \emptyset$.

(3) $V = \bigcup_{a \in A} V_a$ indicates a set of the attribute values, where value scope of attribute.

(4) $f : U \times R \to V$ represents a information function, and it specifies the attribute value of each object of x .

Each attribute subset $P \subseteq C$determines a binary indiscernibility relation: $IND(P) = \{(x,y) \in U \times U | \forall a \in P, f(x,a) = f(y,a)\}$ Relation $IND(P)$is a division of U, denoted as $U/IND(P)$. $U/IND(P)$is denoted as U/P for short. $U/P = \{P_1, P_2, \cdots, P_k\}$, where

$P_i = [x]_{p_i} = \{y | \forall a \in P_i, f(x,a) = f(y,a)\}$is an equivalence class.

Definition 2.2 [1] In information system $S = (U,A,V, f)$, $A = C \cup D$and $\forall P \subseteq C$. $U/P = \{P_1, P_2, \cdots, P_k\}$ is partition of U based on P. $\forall X \subseteq U$, define

$$\underline{P}(X) = \cup \{Y \in U/P | Y \subseteq X\} \tag{1}$$

and

$$\overline{P}(X) = \cup \{Y \in U/P | Y \cap X \neq \emptyset\} \tag{2}$$

$\underline{P}(X)$ and $\overline{P}(X)$ are lower and upper approximation of X respectively.

Definition 2.3 [12] In information system $S = (U,A,V,F)$, $A = C \cup D$. Let $P \subseteq C$ and $Q \subseteq D$. The P dependency based on Q is defined as $K = \gamma_P(Q) = \frac{POS_P(Q)}{|U|}$, where 0≤k<1.

We use $P \underset{K}{\Rightarrow} Q$ to show Q is induced from P with K dependency. When k = 1, Q is dependent on P completely. when $0 < K < 1$, Q is in secondary dependency to P. When k = 0, Q is completely induced from the P.

2.2 Similarity Relation

Pawlak rough set classifies objects through equivalence relation. Many scholars extend rough set model by replacing the equivalence relation with similarity relation. So the rough set can deal with incomplete information systems. Similarity relation is divided into symmetric similarity relation and non symmetric similarity relation.

Definition 2.4 [12] In incomplete information system$S = (U, C \cup D, V, F)$, $B \subseteq C$ and $a \in B$. Define symmetric similarity relation as follows.

$$T(B) = \{(x,y) \in U \times U | (f(x,a) = * \vee f(y,a) = * \vee f(x,a) = f(y,a)\} \quad (3)$$

For any$X \subseteq U$, define

$$\underline{T_B}(X) = \cup\{Y \in U/T(B) | Y \subseteq X\} \quad (4)$$

and

$$\overline{T_B}(X) = \cup\{Y \in U/T(B) | Y \cap X \neq \emptyset\} \quad (5)$$

where $[x]_{T_B} = \{y \in U | (x,y) \in T(B)\}$.

$\underline{T_B}(X)$ and $\overline{T_B}(X)$ called the lower approximation and upper approximation of X in incomplete information system $S = (U, C \cup D, V, F)$.

Definition 2.5 [13,14] In incomplete information system$S = (U, C \cup D, V, F)$, $B \subseteq C$. Define non-symmetric similarity relation as follows.

$$\forall x, y \in U (S_B(x,y) \Leftrightarrow \forall c_j \in B(c_j(x) = * \vee c_j(x) = c_j(y))) \quad (6)$$

Symmetric similarity class (referred as similarity classes, i.e. symmetric similarity resembles to the set of object x) is $S_B(x) = [x]_S = (y \in U | (y,x) \in S(B))$, while the set of object symmetrically similar with x is $S_B^{-1}(x) = [x]_S^{-1} = (y \in U | (x,y) \in S(B))$.

For any $X \subseteq U$, define

$$\underline{S_B}(X) = \cup\{Y \in U/S(B) | Y \subseteq X\} \quad (7)$$

and

$$\overline{S_B}(X) = \cup\{Y \in U/S(B) | Y \cap X \neq \emptyset\} \quad (8)$$

where $[x]_{S_B} = \{y \in U | (x,y) \in S(B)\}$.

$\underline{S_B}(X)$ and $\overline{S_B}(X)$ are called the lower approximation and upper approximation of X in incomplete information system $S = (U, C \cup D, V, F)$.

2.3 Limited Tolerance Relation

Compared with tolerance relation and non symmetric similarity relation, limited tolerance relation is proposed by scholars, which inherits the merits of tolerance relation and non symmetric similarity relation.

Missing value in information system is denoted as '*'. Limited tolerance relation is defined as follows.

Definition 2.6 In incomplete information system $S = (U, C \cup D, V, F)$, $L \subseteq U * U$ is a limited tolerance relation. Let $B \subseteq A$, and $P_B(x) = \{b | b \in B \wedge b(x) \neq *\}$.

$$\begin{aligned} \forall x, y \in (L_B(x,y) \Leftrightarrow \forall b \in B(b(x) = b(y) = *) \vee \\ ((P_B(x) \cap P(y) \neq \emptyset) \wedge \forall b \in B((b(x) \neq *) \\ \wedge (b(y) \neq *) \rightarrow (b(x) = b(y))))) \end{aligned} \quad (9)$$

Obviously, limited tolerance relation L contains reflexivity and symmetry. However, limited tolerance relation(L) does not contain transitivity.

Definition 2.7 In incomplete information system $S = (U, C \cup D, V, F)$, $I_B^L = \{y | y \in U \wedge L_B(x,y)\}$ is a class based on limited tolerance relation L. Approximations based on limited tolerance relation L are defined as follows.

$$D_L^B = \{x | x \in U \wedge I_B^L(x) \cap D \neq \emptyset\} \quad (10)$$

$$D_B^L = \{x | x \in U \wedge I_B^L(x) \subseteq D\} \quad (11)$$

In the view of symmetric similarity relation, a null value of attribute can be equal to any known attribute value and cause the defects of less similarity between objects and low partition granularity possibly. The defects are prone to the miscarriage of justice to a similar class in a false condition, which is bound to affect the application of rough set model based on symmetric similarity relation. For example, there are two objects $X = \{1,1,1,1,1,1\}$ and $Y = \{0,0,0,0,0,0\}$. Obviously, any attribute values of X and Y are not equal. Therefore, X and Y do not satisfy the equivalence relation; neither do the symmetric similarity relation. However, for some reason, some attribute value of X and Y is missing, $X = \{1,1,1,*,*,*\}$ and $Y = \{*,*,*,0,0,0\}$ ("*" indicates a null value). According to Kryszkiewicz's definition of symmetric similarity relation[8], the symmetry X and Y is similar and belong to the same symmetrical similarity class. This is obviously a miscarriage of justice. In some cases, the degrees of similarity between objects maybe vary greatly. For example, some completely equal (i.e. equivalence relation), while some individual attribute values are equal, and most of the rest attributes are nullable. Such as $X = \{1,2,1,0,1,0\}$, $Y0 = \{1,2,1,0,1,0\}$, $Y1 = \{1,2,1,0,1,*\}$, $Y2 = \{1,2,1,*,*,*\}$, $Y3 = \{1,*,*,*,*,*\}$, $Y4 = \{*,*,*,*,*,*\}$. Based on the definition of symmetric similarity relation, Y0, Y1, Y2, Y3 and Y4 have similarly symmetry to X. But the similarity degrees are greatly different. Y0 and X are similar and satisfy the equivalence relation. Y4 and X similarity relation is based on the all attributes values of Y4 are null. Non-symmetric similarity relation classification is not only stricter than symmetric similarity relation, but also shares the defect. So we propose a new dual-limited symmetric similarity relation.

3 Rough Set Model Based on the Dual-limited Symmetric Similarity Relation

Definition 3.1 In incomplete information system $S=(U,C\cup D,V,F)$, an unknown attribute value is marked as $*$. Define

$$F(x,y,a,*)=\left\langle \begin{array}{l}1, f(x,a)=*\vee f(y,a)=*\\ 0, f(x,a)\neq *\wedge f(y,a)\neq *\end{array}\right. \quad (12)$$

$F(x,y,a,*)$ is used to determine whether there are at least a null concerning the attribute of the object x and y.

For$B\subseteq C$, $F(x,y,B,*)=\sum\limits_{a\in B}F(x,y,a,*)$ shows that there is at least a attribute number of an unknown value at set B of object x and y.

Definition 3.2 In incomplete information system$S=(U,C\cup D,V,F)$, $B\subseteq C$. Symmetric similarity relation reliability can be defined as

$$SR(x,y,a)=\frac{|B|-\sum\limits_{a\in B}F(x,y,a,*)}{|B|} \quad (13)$$

$SR(x,y,a)$ can be used to represent the similarity reliability degree between two objects at the attribute set B. When $SR=0$, the similar reliability degree is minimum. When $SR=1$, the similarity reliability degree is maximum.

Definition 3.3 In incomplete information system $S=(U,C\cup D,V,F)$, $B\subseteq AT$and $a\in B$. V_a is the value range of attribute a. Let $x\in U$ and $y\in U$. Define the equal probability of x and y as follows.

$$p(x,y,a,*)=\begin{cases}\frac{1}{|V_a|},\ f(x,a)=*\wedge f(y,a)\neq *\\ \quad\vee f(x,a)\neq *\wedge f(y,a)=*\\ \frac{1}{|V_a|^2},\ f(x,a)=*\wedge f(y,a)=*\\ 1,\ \ f(x,a)\neq *\wedge f(y,a)\neq *\\ \quad\wedge f(x,a)\neq *\wedge f(y,a)\end{cases} \quad (14)$$

The similarity reliability of x and y can be defined as $P(x,y)=\min\limits_{a\in B}(p(x,y,a,*))$and 0< $P(x,y)$ ≤1.

Definition 3.4 In incomplete information system $S=(U,C\cup D,V,F)$, $B\subseteq AT$and $a\in B$. V_a is the value range of attribute. The dual-limited symmetric similarity relation is defined as

$$\begin{array}{l}T^{\alpha,\beta}(B)=\{(x,y)\in U\times U|((f(x,a)=*\\ \vee f(y,a)=*\vee f(x,a)=f(y,a))\\ \wedge SR\geq\alpha\wedge P(x,y)\geq\beta)\vee I_U,a\in B\}\end{array} \quad (15)$$

where

(1) $I_U=\{(x,x)|x\in U\}$.

(2) 0≤ α ≤1 is constants, which shows the minimum threshold to control the similar degree and reliability. 0≤ α ≤1 is constants, which increase the judgment of comparable degree and reliability in the definition refines the knowledge classification granularity.

(3) When $SR=1$, the similarity relation reaches its highest comparable degree. Each attribute value at attribute a in attribute set B of object x and y do not contain "*". At this time, the similarity relation is strengthened as equivalence relation. Therefore, equivalence relation is a special case of symmetric similarity relation.

(4) Symmetric equivalence relation $T(B)$ is α reliable. Otherwise, $T(B)$ is not reliable, and then x is not similar to y.

It is possible to enhance the ability of class by introducing similarity comparative and similarity reliability. From Definition 3.4, x and y are similarity and must satisfy a condition:

The proportion of attributes under which x and y are equivalence from all attributes considered is no less than α.

Remark 3.1 Dual-symmetric similarity relation satisfies reflexivity and symmetric. However, dual-symmetric similarity relation does not satisfy transitive.

From Definition 3.4, if x is symmetrically similar to y, the following conditions must be met.

(1) Attribute set of $a\in B$, the number of the attributes of the equal value(and not null value) for object x and y in all attributes must be no less than the proportion of α constant.

(2) The minimum probability of equal value of x and y at each attribute is not less than β.

(3) And based on this, if x has is of null value or y has null value, or value of x and y is equal but there is not null value, then x and y has symmetric similarity relation.

The symmetric similarity relation has reflectivity and symmetry , but not transitivity.

Definition 3.5 In incomplete information system $S=(U,C\cup D,V,F)$, symmetric similarity class (similarity class for short, i.e. a set of objects symmetrically similar to x) is $T_B(x)=[x]_{TB}=(y\in U|(y,x)\in T(B))$. In incomplete information system $S=(U,C\cup D,V,F)$, $B\subseteq C$. The lower approximation set $\underline{T_B}(X)$of X and the upper approximation set $\overline{T_B}(X)$are defined as follows:

$$\underline{T_B}(X)=\cup\{Y\in U/T(B)|Y\subseteq X\}=\{x\in U|[x]_{T_B}\subseteq X\}$$
$$\overline{T_B}(X)=\cup\{Y\in U/T(B)|Y\cap X\neq\emptyset\}=\{x\in U|[x]_{T_B}\cap X\neq\emptyset\}$$

where $[x]_{T_B}=\{y\in U|(x,y)\in T(B)\}$shows a set of objects of TB with symmetric similarity relation in U, i.e. a symmetric similarity class defined by X. $U/T(B)$ indicates a U division by binary relation.

In fact $\underline{T_B}(X)$ is the maximum set of the definite objects of X based on existing knowledge, but $\overline{T_B}(X)$ is the minimum set of the possible objects of X based on existing knowledge could belong to the object of X of the set.

$POS_{T_B}(X)=\underline{T_B}(X)$ is X's positive domain, which is a set of all members belonging entirely to X.

$NEG_{T_B}(X)=U-\overline{T_B}(X)$ is X's negative domain, which is a set of all members definitely not belonging entirely to X.

$BN_{T_B}(X) = \overline{T_B}(X) - \underline{T_B}(X)$ is X's boundary domain, which is a set of members neither being to X according to the symmetric similarity relation TB nor being classified toX^C. Here X^C is complement of C.

Example 3.1 In incomplete information systems $S1 = (U, C \cup D, V, F)$ let $U = \{x_1, x_2, x_3, x_4, x_5\}andC = \{a_1, a_2, a_3, a_4, a_5\}$. The value of each attribute domain is$\{0,1\}$, $\{0,1,2\}$, $\{0,1,2,3\}$, $\{0,1\}$ and $\{0,1,2\}$ respectively. Let x1={1,1,0,*,2}, x2={*,1,*,*,2}, x3={*,1,0,1,2}, x4={1,0,*,1,0}, x5={*,0,*,1,0}.

Suppose α=0.6 and α=0.25 then $p(x_4,x_5,a_3,*) = \frac{1}{|a_3|^2} = \frac{1}{16}$, $P(x_4,x_5) = \min\limits_{i=1\sim5}(p(x_4,x_5,a_i,*)) = \frac{1}{16} < 0.25$, $F(x_1,x_2,AT,*) = \sum\limits_{a\in AT} F(x,y,a,*) = 3$, $SR(x_1,x_2) = \frac{|B|-F(x,y,AT,*)}{|B|} = \frac{5-3}{5} = 0.4. U/TB = \{\{x1,x2,x3\},\{x4,x5\}\}$.

If neither comparable degree nor reliability is considered, then $U/T(B) = \{\{x1,x2,x3\},\{x4,x5\}\}$.

If not reliability, but comparable degree is considered, then $U/T(B) = \{\{x1,x3\},\{x2\},\{x4,x5\}\}$.

If both comparable degree and reliability are considered, then $U/T(B) = \{\{x1,x3\},\{x2\},\{x4\},\{x5\}\}$.

From Example 3.1, the indiscernible x1 and x2 can be distinguished considering the comparability and reliability. So do x4 and x5. Example 3.1 indicates that the introduction of reliability and its concept makes knowledge classification granularity finer and more accurate in the symmetric similarity relation.

4 Acquisition of Attribute Weight

In the decision-making process, the importance of each different attribute is different, which should be considered by us. To obtain the degree of importance, rough set theory follows the basic principle.

(1) On the basis of the classification from the symmetric similarity relation, some attributes need to be eliminated.

(2) The change of attribute dependence needs to be analyzed without certain attributes.

(3) If the removal of an attribute brings great change to the corresponding attribute dependency, the importance of this attribute is high.

(4) Otherwise, the importance of this attribute is low.

Definition 4.1 In the incomplete information system $S = (U, C \cup D, V, F) P \subseteq C$ and $Q \subseteq D$. Define

$POS_P(Q) = \bigcup\limits_{X \subseteq U/Q} \underline{P}(X)$, where $\underline{P}(X) = \underline{T_P}(X)$.

$POS_P(Q)$means the a set of objects from U assigned accurately by the similarity classification $U/T(P)$ to the symmetric similarity set Q.

Definition 4.2 In the incomplete information system $S = (U, C \cup D, V, F)$ $P \subseteq C$ and $Q \subseteq D$. Define

$$K = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}.$$

Kshows that knowledge Q is dependent on knowledge P to degree K, denoted as $P \underset{K}{\Rightarrow} Q$.

Definition 4.3[12] (The relative importance of attributes) In the incomplete information system $S = (U, C \cup D, V, F)$the importance of attribute subset $A_i \subseteq A$ to D is defined as $\sigma_D(A_i) = \gamma_A(D) - \gamma_{A-A_i}(D)$, while $\sigma_D(A_i)$ means the importance of attribute A_i to decision attribute set D. The higher the value is, the more importance of its corresponding attribute is and vice versa.

In the incomplete information system $S = (U, C \cup D, V, F)$, $U = \{x_i | i = 1, 2, \cdots, n\}$ and $A = \{a_1, a_2, \cdots, a_m\}$. Then |AT|=m-0, |D|=m_1 and |U|=n.

5 A Case of Application

In incomplete decision information system like table 1, there are condition attributes A1, A2, A3, A4, A5 and decision attribute D. "*" denotes missing value in Table 5.1. This decision table gets a chore through the attribute reduction. The method of last section is used to calculate attributes weights of all the conditions in order to obtain the importance of each the condition attributes for decision-making. Suppose α=0.6 and β=0.25, use limit symmetric similarity relation T to classify the above data domain according to the condition attribute and the decision attribute.

Table.5.1 IIS decision-making after reduced

No.	A1	A2	A3	A4	A5	D
1	1	1	1	0	1	1
2	0	*	1	0	1	1
3	*	*	0	1	0	0
4	1	*	1	1	1	1
5	*	*	1	1	1	2
6	0	1	1	*	*	1
7	*	1	1	1	0	0

$U/AT = \{\{1\},\{2\},\{3\},\{4,5\},\{6\},\{7\}\}$,
$U/D = \{\{1,2,4,6\},\{3,7\},\{5\}\}$,

And each gets rid of a condition attribute to compute the distribution and get:

$U/(AT - A_1) = \{\{1,2\},\{3\},\{4,5\},\{6,7\}\}$,
$U/(AT - A_2) = \{\{1\},\{2,6\},\{3\},\{4,5\},\{7\}\}$,
$U/(AT - A_3) = \{\{1\},\{2\},\{3,7\},\{4,5\},\{6\}\}$,
$U/(AT - A_4) = \{\{1,4,5\},\{2,5\},\{3\},\{6,7\}\}$,
$U/(AT - A_5) = \{\{1\},\{2,6\},\{3\},\{4,5,7\}\}$.

Analysis the importance of each condition attribute to the decision making:

$POS_A(D) = \{1,2,3,6,7\}$,
$POS_{AT-A_1}(D) = \{1,2,3\}$,
$POS_{AT-A_2}(D) = \{1,2,3,6,7\}$,
$POS_{AT-A_3}(D) = \{1,2,3,6,7\}$,

$POS_{AT-A_4}(D) = \{3\}$,
$POS_{AT-A_5}(D) = \{1,2,3,6\}$,
$\gamma_{AT}(D) = \frac{POS_{AT}(D)}{|U|} = 5/7$,
$\gamma_{AT-A_1}(D) = \frac{POS_{AT-A_1}(D)}{|U|} = 3/7$,
$\gamma_{AT-A_2}(D) = 5/7$,
$\gamma_{AT-A_3}(D) = 5/7$,
$\gamma_{AT-A_4}(D) = 1/7$,
$\gamma_{AT-A_5}(D) = 4/7$,
$\sigma_D(A_1) = \gamma_{AT}(D) - \gamma_{AT-A_1}(D) = \quad 2/7$,
$\sigma_D(A_2) = 0/7$,
$\sigma_D(A_3) = 0/7$,
$\sigma_D(A_4) = 4/7$,
$\sigma_D(A_5) = 1/7$.

After the standardized treatment, the weights of the five condition attribute are 0.2857, 0.0000, 0.0000, 0.5714 and 0.1429. The result shows that the order of attribute weight is A4, A1, A5, A3 and A2. That is the attribute A4 has the greatest influence. A3 and A2 have the least influence.

6 Conclusion

In order to solve the problem of lack of accuracy, large granularity knowledge as well as the increase of error in knowledge classification in symmetric similarity relation, this paper puts forward a comparable degree and reliability to limit the symmetrical relationship. A rough set model based on dual-limited symmetric similarity relation is set up. The model proposed in this paper inherits the advantages of dealing with incomplete information system utilize symmetric similarity relation, which especially limits the comparison of null attribute value ”*” and make up for the defects of the symmetric similarity relation. This method makes sample object classification in the incomplete information system more reasonable, reduces the error rate of classification in the original model and improves the accuracy and granularity of knowledge classification.

7 Acknowledgments

This work was supported by National Social Science Foundation of China(No.12BGL088), Doctorate Foundation of Shanghai (No. 11692191400) and Ministry Social Science Foundation of China(No. 11YJA790165).
The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Z. Pawlak, Rough sets, International Journal of Parallel Programming. **1411**, 341-356 (1982).
[2] R. Slowinski, D. Vsnderpooten, Similarity relations as a basis for rough approximations, ICS Research Report, Warsaw Univ. Technology, 53-95 (1995).
[3] R. Slowinski, and D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Transactions on Knowledge and Data Engineering, **12**(2) , 331-336 (2000).
[4] S. Greco, B. Matarazzo, R. Slowinski, Handling missing values in rough set analysis of multi-attribute and multi-criteria decision problems, New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 146-157 (1999).
[5] M. Kryszkiewcz, Rough set approach to incomplete information systems, Information Sciences, **112**(124), 39-49 (1998).
[6] M. Kryszkiewcz, Rules in incomplete information systems, Information Sciences, **113**(324), 271-292 (1999).
[7] A. Skowron, K. J. Stepaniu, Tolerance approximation spaces, Fundamenta Informaticae, **27**(2/3), 45-53 (1996).
[8] G. Wang, Extension of rough set under incomplete information systems, Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, **2**, 1098-1103 (2002).
[9] J. W. Grzymala-Busse, Characteristic relations for incomplete data generalization of the indiscernibility relation, Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing, 244-253 (2004).
[10] L. Wang, X. Yang, J. Yang, and C. Wu, Incomplete multigranulation rough sets in incomplete ordered decision system, Bio-Inspired Computing and Applications, 323-330 (2012).
[11] X. B. Yang, M. Zhang, H. L. Dou and J. Y. Yang, Neighborhood systems-based rough sets in incomplete information system, Knowledge-Based Systems, **24**(8), 858-867 323-330 (2012).
[12] D. Q. Miao, D.G. Li, The Computing and Application of Rough Set Theory, Tsinghua University Press, (2008).
[13] G. Y. Wang, Rough set theory and knowledge acquisition”, Xi’an: Xi’an Jiaotong university Press, (2001).
[14] J. Stefanowski, A. Tsoukias, On the extension of rough sets under incomplete information, New Directions in Rough Sets, Data Mining, and Granular-Soft Computing, 73-81 (1999).

Yuming Zhai received the MS degree in Software Engineering from East China Normal University in 2006, and the PhD degree in Management Science and Engineering from the Glorious Sun school of Business and Management, Donghua University in 2012. He is currently a Lecturer in Shanghai Institute of Technology. His research interests are in the areas of Intelligent Decision, Dominance Rough Set.