Variable Precision Rough Set Model Based on Set Pair Homeopathic Similarity Relationship

Yong Liu[1] and Yi Lin[2]

[1]*School of Business, Jiangnan University, Wuxi 214122, China.*
[2]*School of Business,Slippery Rock University, Slippery Rock, U.S.A.*

Abstract
Due to the complexity and uncertainty of the objective world, as well as the limitation of human understanding, it has been difficult for the classical rough set method to deal with incomplete information system with noise data, ambiguity and other informational uncertainties. In view of this, the thought and method of the set pair theory and variable precision rough set are used in this paper to construct a novel variable precision rough set model. After scrutinizing the advantages and limitations of the models previously established based on the set pair similarity relationship, we define the set pair homeopathic similarity relationship based on threshold α connection degree. Then we employ this new concept to substitute the equivalence relationship of the variable precision rough set so that a novel variable precision rough set model based on the set pair homeopathic similarity relationship is established. Upon developing the properties of this model, we construct an example to show the feasibility and effectiveness of our novel model.
Keywords Incomplete information system; Set pair analysis; Variable precision rough set; Equivalence relationship.

1 Introduction

As an useful mathematical tool to deal with information that involves inaccuracy, uncertainty and fuzziness, rough set theory was proposed by Pawlak [1]. This theory has been widely applied in such areas as the knowledge discovery, data mining, decision analysis, pattern recognition, and other fields [2–4]. The classical rough set theory is developed to deal with complete information systems with discrete attribute, where all the attribute values are known, thus the upper and lower approximation sets of the object sets based on equivalence relations, which satisfy the reflexive, symmetric and transitive properties, are defined. However, due to the complexity of the real world, uncertainty and limitation of human knowledge, not all attribute values acquired are known. So it is difficult for the traditional rough set theory to deal with such information systems and to materialize the desired knowledge discovery and extraction of relevant rules

In order to make rough set theory effectively deal with systems of incomplete information, many scholars have been involved in the relevant research, leading to some important results. In summary, there exist two main classes of related works. On one hand, each incomplete information system is pre-processed in order to make up the missing data; then the given system is investigated as a system with complete information. On the other hand, the theoretical equivalence relation among the objects of the system of incomplete information is directly weakened so that some new kind of relation, such as tolerance relationship [5, 6], similarity relationship [7], dominance relationship [8], etc., is introduced. For the former class of research, due to the information loss caused by pre-processing of the information system, the related works will not be discussed in this paper. For the latter class of works, when all the unknown attribute values of the system of incomplete information are of the omission type, Kryszkiewicz builds the tolerance relationship, which satisfies the properties of reflexivity and symmetry.[9] When the unknown attributes values of the incomplete information are of the lost type, Stefanowski and Tsoukias establish the similarity relationship, which satisfies the properties of reflexivity and transitivity.[10] In view of the characteristics of incomplete information systems, Wang puts forward the limited tolerance relationship, which satisfies the properties of reflexivity and symmetry.[11] Corresponding to the systems of incomplete information that contain preference information,Greco, Matarazzo and Slowiski[12, 13],Shao and Zhang [14] respectively put forward rough set model based on tolerance dominance relationship, which satisfies the property of reflexivity. For systems with incomplete order information, Yang, Yang, Chen, et al [15],Xie, Song, Chen, et al [16] construct a rough set model based on the similar dominance relationship, which satisfies the property of reflexivity. They then analyze the shortcomings of the tolerance dominance relationship and the similar dominance relationship, leading to the establishment of a rough set model based on the limited tolerance dominance relationship, which satisfies the property of reflexivity. However, when these rough set models are used to classify the domain with many condition attributes, the definition of rough sets still seems too loose. In order to deal with this problem, the set pair analysis method is introduced into the study of incomplete information systems [17], and additional rough set models, which are based on set pair analysis, are constructed [18–26]. By carefully analyzing these models, it can be seen that most of them are based on a local aspect and consider the classification effect of some specific objects without an overall classification performance of all objects by taking into account of unified approach, relevant uncertainty, oppositions, etc. It has been difficult for each of these established models to deal with incomplete information systems with preference information.

In view of the shortcomings of the existing literature on incomplete information systems and the fact that there exist noise data, fuzzy and grey information in the incomplete information of systems, in this paper we will employ the set pair analysis method, combined with the previous studies, to construct a novel variable precision rough set model based on the set pair homeopathic similarity relationship.

The organization of this paper is as follows. Section 2 compares the advantages and disadvantages of the currently available set pair similarity relationships. Section 3 introduces the concept of our set pair homeopathic similarity relationship. Section 4 establishes the variable precision rough set model based on set pair homeopathic similarity relationship. A case study is given in Section 5. Then this presentation is concluded in Section 6.

2 The Set Pair Analysis

As a useful tool of system analysis to characterize and to study a variety of certainty and uncertainty and their laws of transformation in systems, the methodology of set pair analysis is proposed by Chinese scholar Keqin Zhao in 1989. Its core idea is to holistically analyze and deal with the certainty relationship and the uncertainty relationship of the objective matter of concern. This methodology has been widely applied to the study of such areas as mathematics, physics, systems science, management science, decision-making science, theory of predictions, computer science, social science, artificial intelligence, and many others. It has shown theoretical significance and potential for success for applications[17].

Definition 1[17].Assume that two given sets A and B from a set pair $H(A, B)$,and that under the particular background (W) of the problem of concern, H have n attributes, where s attributes are shared by both of A and B ,p attributes are about the opposition of A and B,and f attributes are neither shared by A and B nor about the opposition of A and B.Then the ratio

s/n is called the identity degree of A and B under the background (W) ;

f/n the discrepancy degree of A and B under the background (W) ;and

p/n the contrary degree of A and B under the background (W) ;

And the connection degree of and under the background is defined as follows:

$$\mu_W(A, B) = \frac{s}{n} + \frac{f}{n}i + \frac{p}{n}j \tag{1}$$

The connection degree reflects the relationship between A and B under the background (W),denoted as $= a + bi + cj$, $0 \leq a, b, c \leq 1$ and $a + b + c = 1$, where the parameters a, b, c stand for some relationship between the sets A and B under the background (W) .

Definition 2[17].For the given connection degree $= a + bi + cj$,the rate of the identity degree a and the contrary degree c under some specific background (W) is referred to as the set pair situation, denoted as follows:

$$S(H) = \frac{a}{c} \tag{2}$$

If $c = 0$,then $S(H) \to \infty$,which is referred to as the infinite situation with respect to the specific background (W) .

3 Set Pair Homeopathic Similarity Relationship

3.1 Incomplete Information System

Definition 3. Suppose that $S = (U, A, V, f)$ is an information system, where $U = \{U_1, U_2, ..., U_n\}$ is a finite and nonempty set, known as the universe, $A = C \bigcup D$ a finite and nonempty attribute set, $C = \{a_1, a_2, ..., a_m\}$ the condition attribute set, $D = \{d_1, d_2, ..., d_p\}$ the decision attribute set, and $V = \bigcup V_{ij}$ the value range of U on A, where V_{ij} is the attribute value of U_i on $a_j \in A$, $f : U \times A \to V$ is an information function. For $\forall a_j \in A, \forall x_i, x_k \in U, f(x_i, a_j) \in V_{ij}, \forall b \in B, B \subseteq A$,the equivalence relationship of x_i, x_k with respect to the attritube b is defined as follows:

$$R = \{(x_i, x_k) \in U \times U : f(x_i, b) = f(x_k, b)\} \quad (3)$$

Definition 4. Suppose that $S = (U, A, V, f)$ is an information system. If there is at least one object U_i and one attribute $a_j \in A$ such that V_{ij} takes the null value, denoted by *, then information system $S = (U, A, A, f)$ is called incomplete information system, denoted as $S^* = (U, A, V, f)$; otherwise, it is called a complete information system.

3.2 Comparison of Different Kinds of Set Pair Similarity Relationships

Due to the incompleteness of information of an incomplete information system, it is difficult for the traditional method of rough sets based on equivalence relationship to solve problems involving such systems. In view of this fact, some scholars suggest to use the methodology of the set pair analysis to define the similarity relationship to investigate incomplete information systems. In short, there exist the following four categories of similarity relationships [18–26].

Definition 5.In the incomplete information system $S^* = (U, A, V, f)$, the general similarity relationship is defined as follows:

$$SIM(x, y) = \{(x, y) \in U \times U \mid \forall a \in A, a(x) \geq a(y) or\ a(x) = * or\ a(y) = *\} \quad (4)$$

For $\forall a \in A, \forall x, y \in U, B \subseteq A, |B| = n$, teh connection degree based on the incomplete information system is denoted as

$$\mu_B(x, y) = \frac{s}{n} + \frac{f}{n} i + \frac{p}{n} j$$

where $s = |\{a \in B \mid a(x) = a(y)\ and\ a(x) \neq *\ and\ a(y) \neq *\}|$ stands for the number of the attributes that are clear and equal for x and y with respect to the attribute set B; while $p = |\{a \in B \mid a(x) \neq a(y)\ and\ a(x) \neq *\ and\ a(y) \neq *\}|$ stands for the number of the attributes that are clear but not equal for x and y

with respect to the attributes set B;

$f = |\{a \in B \mid a(x) \neq * \ or \ a(y) \neq *\}|$ stands for the number of the attributes that are not clear and equal for x and y with respect to the attributes set B.

Write $\mu_B(x,y) = \frac{s}{n} + \frac{f}{n}i + \frac{p}{n}j = a + bi + cj$ where $a = \frac{s}{n}, b = \frac{f}{n}, and \ c = \frac{p}{n}$

(1)Set Pair Similarity Relationship and Similarity Classes Based on the Connection Degree

Definition 6.[18]. For the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A$,the set pair similarity relationship of x and y is defined as follows:

$$SMH(B) = \{(x,y) \in U \times U \mid u_B(x,y) = a + bi + cj\} \tag{5}$$

Definition 7.[18].For the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A$,the set pair similarity classes of x and y with respect to the attribute set B are defined as follows:

$$SMH(B) = \{y \in U \mid u_B(x,y) = a + bi + cj\} \tag{6}$$

Obviously, this set pair similarity relationship satisfies the property of symmetry, but does not satisfy the properties of reflexivity and transitivity. In particular, in practical applications, it can be found that when the information system of concern contains a large amount of missing values, the result of model analysis tends to be not satisfactory, and the operational properties of the lower and upper approximations of rough sets are imperfect.

(2) Set Pair Similarity Relationship and Similarity Classes Based on a Threshold Value α

Definition 8. [19].For the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A, 0 \leq \alpha \leq 1$,the set pair similarity relationship of x and y based on the threshold value α is defined as follows:

$$SMH^{\alpha}(B) = \{(x,y) \in U \times U \mid u_B(x,y) = a + bi + cj, a + b \geq \alpha\} \tag{7}$$

Definition 9. In the incomplete information system $S^* = (U, A, V, f)$,for $\forall a \in A, \forall x, y \in U, B \subseteq A, 0 \leq \alpha \leq 1$,the set pair similarity classes of x and y with respect to the attribuye set B based on the threshold value α are defined as follows:

$$SMH_B^{\alpha}(B) = \{y \in U \mid u_B(x,y) = a + bi + cj, a + b \geq \alpha\} \tag{8}$$

Evidently, this set pair similarity relationship satisfies the property of symmetry, but not that of transitivity.

The rough set model based on the set pair similarity relationship only considers and limits the known identity degree and the proportion of the uncertain attributes among the objects. It ignores the influence of the discrepancy degree on the similarity, which will affect the performance of classification. In addition, identifying the influence of the identity degree and that of the discrepancy degree on the similarity will also to a certain extent increase the classification error. For the rough set based on the set pair similarity relationship, there exist two types of errors. On one hand, it ignores the difference between uncertain attributes and plays too much emphasis on their similarity. On the other hand, it ignores the interfering effect of the discrepancy degree on the decision-making system in which requirements of relatively higher accuracy are imposed, leading to major differences in the decision-making rules.

(3) Set Pair Similarity Relationship and Similarity Classes based on identity and contrary degrees (α, λ)

Definition 10. [22].In the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A, 0.5 \leq \alpha \leq 1, 0 \leq \lambda \leq 0.5$,the set pair similarity relationship of x and y based on identity degree α and contrary degree λ is defined as follows:

$$SMH^{\alpha}(B) = \{(x, y) \in U \times U \mid u_B(x, y) = a + bi + cj, a \geq \alpha, c \leq \lambda\} \quad (9)$$

Obviously, this set pair similarity relationship satisfies the properties of reflexivity and symmetry, but not that of transitivity.

Definition 11.In the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A, 0.5 \leq \alpha \leq 1, 0 \leq \lambda \leq 0.5$,the set pair similarity relationship of x and y with respect to the attribute set B based on identity degree α and contrary degree λ are defined as follows:

$$SMH_B^{\alpha,\lambda} = \{y \in U \mid u_B(x, y) = a+bi+cj, a \geq \alpha, c \leq \lambda, 0.5 \leq \alpha \leq 1, 0 \leq \lambda \leq 0.5\} \quad (10)$$

(4) Set Pair Similarity Relationship and Similarity Classes Based on Threshold Values (α, λ)

Definition 12[23].In the incomplete information system $S^* = (U, A, V, f)$, $\forall a \in A, \forall x, y \in U, B \subseteq A, 0 \leq \alpha \leq 1, 0 \leq \lambda \leq 1$,the set pair similarity relationship of x and y based on the threshold values (α, λ) is defined as follows:

$$SMH^{\alpha}(B) = \{(x, y) \in U \times U \mid u_B(x, y) = a + bi + cj, a + \lambda b - c \geq \alpha\} \quad (11)$$

Definition 13[23].In the incomplete information system $S^* = (U, A, V, f)$, for$\forall a \in A, \forall x, y \in U, B \subseteq A, 0 \leq \alpha \leq 1, 0 \leq \lambda \leq 1$,the set pair similarity classes of x and y with respect to the attribute set B based on the threshold values (α, λ) are defined as follows:

$$SMH_B^{\alpha}(x) = \{y \in U \mid u_B(x, y) = a + bi + cj, a + b\lambda - c \geq \alpha\} \quad (12)$$

The concept of set pair similarity classes reflects the object set that the similarity degree of x with respect to the attribute set B is greater or equal to(α, λ) .

In view of the existing two kinds of errors the above set pair similarity relationship suffers from, Tao, Dai and Zhang put forward the rough set model based on the threshold value (α, λ) .[23] Here the threshold value α is introduced into the model by separating the contributions of both uncertain and certainty attributes to the relationship of similarity and by paying much more attention to the disturbance of the contrary degree. The relationship of similarities between things is considered and characterized in three aspects: the identity degree, discrepancy degree and contrary degree. Due to the fact that the discrepancy degree possesses both positive and negative incentives on the connection degree, while the threshold value satisfies $0 \leq \lambda \leq 1$, the model only considers the positive incentives of the uncertain attributes on the similarity while neglecting their punitive effects. At the same time, because of the introduction of the threshold value λ , the subjectivity of the model is increased. Additionally, the model does not satisfy the property of reflexive, which affects the quality of the operation of the model.

3.3 Set Pair Homeopathic Similarity Relationship and Similarity Classes

In view of the existing problems in connection degrees, we will use the homeopathic value method to make improvement on the effect of connection degrees. The idea of the homeopathic value method is that along with the identity degree, the value is determined while keeping the same potential class status. For example, for the connection degree $\mu = ai + bj + ck$, the homeopathic value method is used to acquire the novel connection degree $\mu = a(1+b)i + b^2 j + c(1+b)k$. From comparing the two connection degrees, we can see that by using the homeopathic value method the uncertainty of the original connection degree $\mu = ai + bj + ck$ is divided into three parts in the novel connection degree ab. One part is classified into the identity degree b^2 ; one part into the discrepancy degree cb; and the third part into the contrary degree . At the same time, the set pair potentials of these two connection degrees are the same as $\frac{a}{c} = \frac{a(1+b)}{c(1+b)}$.

For the connection degree $a + \lambda b - c$ based on the fourth set pair similarity relationship, the homeopathic value method can be used to deal with the threshold λ so that a novel connection degree $a + (a - c)b - c + b^2\lambda$ is obtained. Because the impact from the part of the uncertainty that is classified into the discrepancy degree b^2 on the relationship of similarity is very little, b^2 can be ignored. Because of this reason, we define the set pair homeopathic similarity relationship as follows.

Definition 14.Suppose that $S^* = (U, A, V, f)$is an incomplete information system.For $\forall a \in A, \forall x \in U, B \subseteq A, 0 \leq \alpha \leq 1$, the set pair homeopathic similarity relationship of x and y based on the threshold value α is defined as follows:

$$SMH^{\alpha}(B) = \{(x, y) \in U \times U \mid u_B(x, y) = a + bi + cj, a + (a - c)b - c \geq \alpha\} \quad (13)$$

The set pair homeopathic similarity relationship satisfies the properties of reflexivity and symmetry, but not that of transitivity.

Definition 15.In the incomplete information system $S^* = (U, A, V, f)$.For $\forall a \in A, \forall x, y \in U, B \subseteq A, 0 \leq \alpha \leq 1$, the set pair homeopathic similarity classes of x and y with respect to the attribute set B based on the threshold value α are defined as follows:

$$SMH_B^{\alpha}(x) = \{y \in U \mid u_B(x, y) = a + bi + cj, a + (a - c)b - c \geq \alpha\} \quad (14)$$

The set pair homeopathic similarity classes reflect such object set that the similarity degree of with respect to the attribute set B is greater or equal to α.

4 Variable Precision Rough Set Based on Set Pair Homeopathic Similar Relationship

4.1 Construction of the Model

In any real-life information system, there always exist a variety of noise data, which is difficult for the traditional rough set method to deal with. In order to deal with such problem, Ziarko [27–29] put forward the variable precision rough set model by introducing the threshold value β and approximation space to reflect this kind of restrictions. Based on the thought and methodology of variable precision rough set model, we introduce the set pair homeopathic similarity relationship to substitute the equivalence relationship of rough set so that a novel variable precision rough set model can be constructed. Accordingly, the lower and upper approximations of the variable precision rough set model based on the set pair homeopathic similarity relationship can be defined as follows.

Definition 16.Suppose that $S^* = (U, A, V, f)$ is an incomplete information sysytem.For $X \subseteq U, B \subseteq C, 0 \leq \alpha \leq 10.5 \leq \beta \leq 1$, the β-lower and β-upper approximations of Xbased on the set pair homeopathic similarity relationship are defined respectively by:

$$\overline{apr}_B^{\alpha,\beta}(X) = \cup\{x \in U \mid \frac{|SMH^{\alpha}(x) \cap X|}{|SMH^{\alpha}(x)|} \geq \beta\} \quad (15)$$

$$\underline{apr}_B^{\alpha,\beta}(X) = \cup\{x \in U \mid \frac{|SMH^{\alpha}(x) \cap X|}{|SMH^{\alpha}(x)|} \geq 1 - \beta\} \quad (16)$$

the β-lower approximation $\underline{apr}_B^{\alpha,\beta}(X)$ of the set X based on the set pair homeopathic similarity relationship can also be called as the positive region of the variable precision rough set model, denoted as $pos_B^{\alpha,\beta}(X)$.In other words, for given confidence threshold values α and β,$\underline{apr}_B^{\alpha,\beta}(X)$ is the set in which the universe

U can be classified definitely into all the element sets of the set X based on the set pair homeopathic similarity relationship. The β-upper approximation $\overline{apr}_B^{\alpha,\beta}(X)$ of the set X reflects that for a given threshold value β ,the universe U can probably be classified into all the element sets of X

Definition 17.Suppose that $S^* = (U, A, V, f)$ is an incomplete information sysytem.For $0 \le \alpha \le 1, 0.5 \le \beta \le 1$,the β-negative region and β- -boundary of a subset X based on the set pair homeopathic similarity relationship are defined as follows:

$$neg_B^{\alpha,\beta}(X) = \cup\{x \in U : \frac{|SMH^\alpha(x) \cap X|}{|SMH^\alpha(x)|} \le 1 - \beta\} \tag{17}$$

$$bnd_B^{\alpha,\beta}(X) = \cup\{x \in U : 1 - \beta < \frac{|SMH^\alpha(x) \cap X|}{|SMH^\alpha(x)|} < \beta\} \tag{18}$$

The β-negative region $neg_B^{\alpha,\beta}(X)$ of X based on the set pair homeopathic similarity relationship reflects that for a given confidence threshold β , the universe certainly cannot be classified into all the elements of the collection set X. The boundary $bnd_B^{\alpha,\beta}(X)$ of X based on the set pair homeopathic similarity relationship reflects that for the given confidence threshold β , the universe U certainly cannot be classified into all the element sets of either the set X or the set -X .

Definition 18.Suppose that $S^* = (U, A, V, f)$ is an incomplete information sysytem.For $0 \le \alpha \le 1, 0.5 \le \beta \le 1$,the β classification quality of X based on the set pair homeopathic similarity relationship is defined as follows:

$$r^{\alpha,\beta}(B, D) = \frac{\left|\cup\{x \in U : \frac{|SMH^\alpha(x) \cap R(x)|}{|SMH^\alpha(x)|} \ge \beta\}\right|}{|U|} \tag{19}$$

The quantity $\gamma_{\alpha,\beta}(B, D)$ measures the proportion that the possible correct classification knowledge is in the existing knowledge for the given values α, β in the universe.

4.2 Properties of the Model

Theorem 1.For the given incomplete information system $S^* = (U, A, V, f)$, let $X \subseteq U, Y \subseteq U, \forall B \subseteq C, 0 \le \alpha \le 1, 0.5 < \beta \le 1$, the following hold true:

(1)$\overline{apr}_B^{\alpha,\beta}(X \cup Y) \supseteq \overline{apr}_B^{\alpha,\beta}(X) \cup \overline{apr}_B^{\alpha,\beta}(Y)$;
(2)$\underline{apr}_B^{\alpha,\beta}(X \cap Y) \subseteq \underline{apr}_B^{\alpha,\beta}(X) \cap \underline{apr}_B^{\alpha,\beta}(Y)$;
(3)$\underline{apr}_B^{\alpha,\beta}(X \cup Y) \supseteq \underline{apr}_B^{\alpha,\beta}(X) \cup \underline{apr}_B^{\alpha,\beta}(Y)$;and
(4)$\overline{apr}_B^{\alpha,\beta}(X \cap Y) \subseteq \overline{apr}_B^{\alpha,\beta}(X) \cap \overline{apr}_B^{\alpha,\beta}(Y)$;

Proof.(1)For $X \subseteq U, Y \subseteq U, 0 \le \alpha \le 1, 0.5 < \beta \le 1$, we have

$$\frac{|SMH^\alpha(x) \cap (X \cup Y)|}{|SMH^\alpha(x)|} \ge \frac{|SMH^\alpha(x) \cap X|}{|SMH^\alpha(x)|}$$

and

$$\frac{|SMH^{\alpha}(x)\cap(X\cup Y)|}{|SMH^{\alpha}(x)|}\geq\frac{|SMH^{\alpha}(x)\cap Y|}{|SMH^{\alpha}(x)|}$$

Therefore, we obtain $\overline{apr}_B^{\alpha,\beta}(X\cup Y)\supseteq\overline{apr}_B^{\alpha,\beta}(X)\cup\overline{apr}_B^{\alpha,\beta}(Y)$;

(2)For $X\subseteq U, Y\subseteq U, 0\leq\alpha\leq 1, 0.5<\beta\leq 1$, we have

$$\frac{|SMH^{\alpha}(x)\cap(X\cap Y)|}{|SMH^{\alpha}(x)|}\leq\frac{|SMH^{\alpha}(x)\cap X|}{|SMH^{\alpha}(x)|}$$

and

$$\frac{|SMH^{\alpha}(x)\cap(X\cap Y)|}{|SMH^{\alpha}(x)|}\leq\frac{|SMH^{\alpha}(x)\cap Y|}{|SMH^{\alpha}(x)|}$$

Therefore, we obtain $\underline{apr}_B^{\alpha,\beta}(X\cap Y)\subseteq\underline{apr}_B^{\alpha,\beta}(X)\cap\underline{apr}_B^{\alpha,\beta}(Y)$;

Similar to the argument of statement (1), statement (3) can be proved; and statement (4) can be proved in the same way as that of statement (2). QED

Theorem 2.For the given incomplete information system $S^*=(U,A,V,f)$, let $X\subseteq Y\subseteq U, \forall B\subseteq C, 0\leq\alpha\leq 1, 0.5<\beta\leq 1$, the following hold true:

(1)$\underline{apr}_B^{\alpha,\beta}(X)\subseteq\underline{apr}_B^{\alpha,\beta}(Y)$;and

(2)$\overline{apr}_B^{\alpha,\beta}(X)\subseteq\overline{apr}_B^{\alpha,\beta}(Y)$.

Proof.(1)Due to $X\subseteq Y$,it follows that $X\cap Y=X$ and $X\cup Y=Y$. So we obtain

$$\underline{apr}_B^{\alpha,\beta}(X)=\underline{apr}_B^{\alpha,\beta}(X\cap Y)=\underline{apr}_B^{\alpha,\beta}(X)\cap\underline{apr}_B^{\alpha,\beta}(Y)$$

Therefore,we have $\underline{apr}_B^{\alpha,\beta}(X)\subseteq\underline{apr}_B^{\alpha,\beta}(Y)$.

(2)From $X\subseteq Y$,it follows that $X\cup Y=Y$ so that we have

$$\overline{apr}_B^{\alpha,\beta}(Y)=\overline{apr}_B^{\alpha,\beta}(X\cup Y)=\overline{apr}_B^{\alpha,\beta}(X)\cup\overline{apr}_B^{\alpha,\beta}(Y)$$

Therefore,we obtain $\overline{apr}_B^{\alpha,\beta}(x)\subseteq\overline{apr}_B^{\alpha,\beta}(Y)$.QED

Theorem 3.For the given incomplete information system $S^*=(U,A,V,f)$, let $X\subseteq U, B_1\subseteq B_2\subseteq C, 0\leq\alpha\leq 1, 0.5<\beta\leq 1$, the following hold true:

(1)$\underline{apr}_{B_1}^{\alpha,\beta}(X)\subseteq\underline{apr}_{B_2}^{\alpha,\beta}(X)$;and

(2)$\overline{apr}_{B_1}^{\alpha,\beta}(X)\supseteq\overline{apr}_{B_2}^{\alpha,\beta}(X)$.

Proof.(1)For $\forall x\in\underline{apr}_{B_1}^{\alpha,\beta}(X)$,from $B_1\subseteq B_2\subseteq A$ it follows that $[x]_{B_1}^{\alpha,\beta}\subseteq X$ such that $[x]_{B_2}^{\alpha,\beta}\subseteq[x]_{B_1}^{\alpha,\beta}$.So we obtain $[x]_{B_1}^{\alpha,\beta}\subseteq X$ such that $x\in\underline{apr}_{B_2}^{\alpha,\beta}(X)$ Therefore,we have $\underline{apr}_{B_1}^{\alpha,\beta}(X)\subseteq\underline{apr}_{B_2}^{\alpha,\beta}(X)$

(2)For $\forall y\in\overline{apr}_{B_2}^{\alpha,\beta}(X)$,we have $[y]_{B_2}^{\alpha,\beta}\cap X\neq\emptyset$,satisfying that $[y]_{B_1}^{\alpha,\beta}\cap X\neq\emptyset$.So, we get $y\in\overline{apr}_{B_1}^{\alpha,\beta}(X)$.Therefore,$\overline{apr}_{B_1}^{\alpha,\beta}\supseteq\overline{apr}_{B_2}^{\alpha,\beta}(X)$ follows.QED

Theorem 4.For the given incomplete information system $S^* = (U, A, V, f)$, let $X \subseteq U, B \subseteq C, 0 \le \alpha_1 \le \alpha_2 \le 1, and\ 0.5 < \beta_1 \le \beta_2 \le 1$. Then following hold true:

(1)$\underline{apr}_B^{\alpha,\beta_1}(X) \subseteq \underline{apr}_B^{\alpha,\beta_2}(X)$;

(2)$\overline{apr}_B^{\alpha,\beta_1}(X) \supseteq \overline{apr}_B^{\alpha,\beta_2}(X)$.

(3)$\underline{apr}_B^{\alpha_1,\beta}(X) \subseteq \underline{apr}_B^{\alpha_2,\beta}(X)$; and

(4)$\overline{apr}_B^{\alpha_1,\beta}(X) \supseteq \overline{apr}_B^{\alpha_2,\beta}(X)$.

Proof.(1)For $\forall x \in \underline{apr}_B^{\alpha,\beta_1}(X)$,from $0.5 < \beta_1 \le \beta_2 \le 2$ it follows that $[x]_B^{\alpha,\beta_1} \subseteq X$. So, we have $[x]_B^{\alpha,\beta_2} \subseteq [x]_B^{\alpha,\beta_1}$ from which we get $x \in \underline{apr}_B^{\alpha,\beta_2}(X)$. And consequently we obtain $\underline{apr}_B^{\alpha,\beta_1}(X) \subseteq \underline{apr}_B^{\alpha,\beta_2}(X)$.

(2)For $\forall y \in \overline{apr}_B^{\alpha,\beta_2}(X)$, $[y]_P^{\alpha,\beta_2} \cap X \ne \emptyset$ holds true so that $[y]_B^{\alpha,\beta_1} \cap X \ne \emptyset$ follows. Hence, we get $y \in \overline{apr}_B^{\alpha,\beta_1}(X)$, and consequently $\overline{apr}_B^{\alpha,\beta_1}(X) \supseteq \overline{apr}_B^{\alpha,\beta_2}(X)$

Similar to the proof of statement (1), statement (3) follows; and statement (4) can be shown in the same way as that of statement (2). QED

The results in Theorem 4 indicate that β value is negatively related to the classification quality. In particular, when β value increases, the classification quality decreases and the positive and negative regions of the set X based on the variable precision rough set model, as proposed in this paper, will become narrower, while the boundary region of the set X will become wider. It means that only a small number of objects are classified. On the other hand, as the value of β decreases, the classification precision increases, and the positive and negative regions of the set X based on the variable precision rough set will widen, while the boundary region narrows. That means that most of the objects are classified, but possibly misclassified.

4.3 Attribute Reduction Algorithm

According to the similarity degree of the positive region, as constructed based on the proposed variable precision rough set model, the following attribute reduction algorithm is established for solving the reduction problem. The special steps are given as follows.

Input: The incomplete information system $S^* = (U, A, V, f)$, the set pair homeopathic threshold value α and confidence threshold value β .

. Output: A reduction B of the incomplete information system $S^* = (U, A, V, f)$

Step 1: Suppose $B = C$;

Step 2: Compute the system classification quality $\gamma^{\alpha,\beta}(C, D)$;

Step 3: For any condition attribute $a \in B$, compute $\gamma^{\alpha,\beta}(B - a, D)$ and the dependence degree $Sig^{\alpha,\beta}(a)$ of the attribute a;

Step 4: If $\gamma^{\alpha,\beta}(B - a, D) \ge \gamma^{\alpha,\beta}(C, D)$ and $Sig^{\gamma}(a)$ is the least, then $B = B - \{a\}$. If for some $a \in B, \gamma^{\alpha,\beta}(B - a, D) \le \gamma^{\alpha,\beta}(C, D)$ and all $Sig^{\gamma}(a)$ are equal, then the attribute a with the missing values for all objects is selected and let $B = B - \{a\}$;

Step 5: If for some $a \in B, \gamma^{\alpha,\beta}(B - a, D) \leq \gamma^{\alpha,\beta}(C, D)$ and all $Sig^{\gamma}(a)$ are equal, then go to Step 6; otherwise go to Step 3; and

Step6:Output a reduction of the initial incomplete information system.

5 Case analysis

Suppose that the decision-making information system $S = (U, V, A, f)$ is about the initial accessories of the military aircraft engine system, where $U = \{U_1, U_2, ..., U_{10}\}$ is the set of initial accessories, $C = \{a_1, a_2, a_3, a_4, a_5\}$ the set of condition attributes, where ,a_1, a_2, a_3, a_4, and a_5 stand respectively for importance, consumable, convertibility, procurability, and economical efficiency, and $D = \{d\}$. For $V_{a_1} = \{2, 1, 0\}$, 2 represents key, 1 represents more important, 0 represents general; for $V_{a_2} = \{2, 1, 0\}$, 2 represents consumable, 1 represents vulnerable, 0 represents general; for $V_{a_3} = \{1, 0\}$, 1 represents substitution, 0 represents cannot be replaced;, for $V_{a_4} = \{2, 1, 0\}$, 2 represents longer for the order cycle, 1 represents general for the order cycle, 0 represents immediate purchase; for $V_{a_5} = \{2, 1, 0\}$, 2 represents excellent, 1 represents general, 0 represents bad; for $V_d = \{1, 0\}$, 1 represents Yes, 0 represents no.

Due to the fact that there exist missing attribute values, the decision-making information system $S = (U, A, V, f)$ is an incomplete decision-making information system $S^* = (U, A, V, f)$, the details of shown are given in Table 1.

(1)When $\alpha = 0.2$, based on the set pair homeopathic similarity relationship,

Table 1 The decision-making information system for the initial accessories selection of the military aircraft engine system

order#	a_1	a_2	a_3	a_4	a_5	d
1	2	1	0	*	1	1
2	1	0	1	2	2	1
3	2	*	0	1	1	1
4	0	2	*	2	0	0
5	*	1	0	1	1	0
6	2	*	1	0	2	1
7	0	2	1	1	*	0
8	0	1	0	0	1	0
9	2	0	1	2	2	1
10	2	*	1	1	2	0

the universe on the condition attribute set C can be divided into:

$$U/C = \{X_1^{0.2}, X_2^{0.2}, X_3^{0.2}, X_4^{0.2}\}$$

where $X_1^{0.2} = \{U_1, U_3, U_5, U_8\}, X_2^{0.2} = \{U_2, U_6, U_9, U_{10}\}$ and $X_3^{0.2} = \{U_4, U_7\}$.

According to the decision attribute set D , the universe can be divided into:

$U/D = \{Y_1, Y_2\}$

where $Y_1 = \{U_1, U_2, U_3, U_6, U_9\}$ and $Y_2 = \{U_4, U_5, U_7, U_8, U_{10}\}$.

When $\beta = 1$, the lower and upper approximations of the decision-making classes Y_1 and Y_2 can be computed as follows:

$\underline{apr}_B^{\alpha,\beta}(Y_1) = \emptyset, \overline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_2, U_3, U_5, U_6, U_8, U_9, U_{10}\}$,

$\underline{apr}_B^{\alpha,\beta}(Y_2) = \{U_4, U_7\}, \overline{apr}_B^{\alpha,\beta}(Y_2) = \{U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8, U_9, U_{10}\}$,

and

$\gamma^{\alpha,\beta}(B, D) = \frac{|\{U_4,U_7\}|}{|\{U_1,U_2,U_3,U_4,U_5,U_6,U_7,U_8,U_9,U_{10}\}|} = \frac{2}{10} = 0.2$

When $\beta = 0.7$, the lower and upper approximations of the decision-making classes Y_1 and Y_2 can be produced as follows:

$\underline{apr}_B^{\alpha,\beta}(Y_1) = \{U_2, U_6, U_9, U_{10}\}, \overline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_2, U_3, U_5, U_6, U_8, U_9, U_{10}\}$,

$\underline{apr}_B^{\alpha,\beta}(Y_2) = \{U_4, U_7\}, \overline{apr}_B^{\alpha,\beta}(Y_2) = \{U_1, U_3, U_4, U_5, U_7, U_8\}$,

and

$\gamma^{\alpha,\beta}(B, D) = \frac{||\{U_4,U_7\}|\cup|\{U_2,U_6,U_9,U_{10}\}||}{|\{U_1,U_2,U_3,U_4,U_5,U_6,U_7,U_8,U_9,U_{10}\}|} = \frac{6}{10} = 0.6$

(2) When $\alpha = 0.4$, based on the set pair homeopathic similarity relationship, the universe on the condition attribute set C can be divided into:

$$U/C = \{X_1^{0.4}, X_2^{0.4}, X_3^{0.4}, X_4^{0.4}, X_5^{0.4}, X_6^{0.4}\}$$

where $X_1^{0.4} = \{U_1\}, X_2^{0.4} = \{U_2, U_9, U_{10}\}, X_3^{0.4} = \{U_3, U_5\}, X_4^{0.4} = \{U_4\}, X_4^{0.4} = \{U_6\}, X_5^{0.4} = \{U_7\}$ and $X_6^{0.4} = \{U_8\}$.

According to the decision attribute set D , the universe can be divided into:

$U/D = \{Y_1, Y_2\}$

where $Y_1 = \{U_1, U_2, U_3, U_6, U_9\}$ and $Y_2 = \{U_4, U_5, U_7, U_8, U_{10}\}$.

When $\beta = 1$, the lower and upper approximations of the decision-making classes Y_1 and Y_2 can be produced as follows:

$\underline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_6\}, \overline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_2, U_3, U_5, U_6, U_9, U_{10}\}$,

$\underline{apr}_B^{\alpha,\beta}(Y_2) = \{U_4, U_7, U_8\}, \overline{apr}_B^{\alpha,\beta}(Y_2) = \{U_2, U_3, U_4, U_5, U_7, U_8, U_9, U_{10}\}$,

and

$\gamma^{\alpha,\beta}(B, D) = \frac{||\{U_1,U_6\}|\cup|\{U_4,U_7,U_8\}||}{|\{U_1,U_2,U_3,U_4,U_5,U_6,U_7,U_8,U_9,U_{10}\}|} = \frac{5}{10} = 0.5$

When $\beta = 0.65$, the lower and upper approximations of the decision-making classes Y_1 and Y_2 are given as follows:

$\underline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_2, U_6, U_9, U_{10}\}, \overline{apr}_B^{\alpha,\beta}(Y_1) = \{U_1, U_2, U_3, U_5, U_6, U_9, U_{10}\}$,

$\underline{apr}_B^{\alpha,\beta}(Y_2) = \{U_4, U_7, U_8\}, \overline{apr}_B^{\alpha,\beta}(Y_2) = \{U_3, U_4, U_5, U_7, U_8\}$,

and $\gamma^{\alpha,\beta}(B, D) = \frac{||\{U_1,U_2,U_6,U_9,U_{10}\}|\cup|\{U_4,U_7,U_8\}||}{|\{U_1,U_2,U_3,U_4,U_5,U_6,U_7,U_8,U_9,U_{10}\}|} = \frac{8}{10} = 0.8$

Based on the attribute reduction algorithm in Section 4.3, when $\alpha = 0.4$ and $\beta = 0.65$, a reduction $\{a_1, a_5\}$ can be acquired. From the reduction $\{a_1, a_5\}$, some probabilistic decision-making rules set can be generated as shown in Table2.

Table 2 Probabilistic decision-making rules generated by the reduction $\{a_1, a_5\}$ with respect to the threshold values ($\alpha = 0.4, \beta = 0.65$)

Rules#	**Supprot number**	**Confidence(%)**
$a_1 = 2$ and $a_5 = 2 \xrightarrow{66.7\%} d = 1$	3	66.7
$a_1 = 1$ and $a_5 = 2 \xrightarrow{100\%} d = 1$	1	100
$a_1 = 0$ and $a_5 = 0 \xrightarrow{100\%} d = 1$	1	100
$a_1 = 0$ and $a_5 = 1 \xrightarrow{100\%} d = 1$	1	100
$a_1 = 0$ and $a_5 = 2 \xrightarrow{100\%} d = 1$	1	100

According to Table 2, we can see that eight objects in the object set can be correctly classified. That is, the classification quality is 80%. From the reduction $\{a_1, a_5\}$, it follows that the company, which manufactures military aircraft engine systems, pays more attention to importance and economical efficiency of the initial accessories. From the set of probabilistic decision-making rules, it follows that if the importance of the initial accessories is general, then the initial accessories should not be selected; if the importance is important and the economical efficiency is excellent, then the initial accessories should be selected. Based on the previous calculation and analysis, by adjusting the set pair homeopathic parameter α and the confidence threshold parameter β , the classification ability of the model cam be improved; and the model can be made to possess some fault tolerant ability so that it can materialize the correct classification, and effectively extract the rules for decision making.

6 Conclusions

Due to the complexity and uncertainty of the objective world, as well as the limitation of human understanding, available information systems often contain incomplete information, and the omission of attribute values in the systems increases the uncertainty and the noise of the system. It has been difficult for the traditional rough set method, developed for handling complete information systems, to deal with incomplete information systems. By taking advantage of the set pair analysis methodology, in this paper we constructed the variable precision rough set based on set pair homeopathic similar relationship. By adjusting the confidence threshold parameters (α, β) , we can make the model possess the fault tolerant ability so that it can effectively deal with information systems with noise data, ambiguity, and other kinds of uncertain information. It is shown that our model can materialize the extraction of decision-making rules and knowledge discovery out of a given incomplete information system. For any incomplete information system that contains preference information, grey information, noise data, and other kinds of uncertain information, how to construct the grey dominance variable precision rough set model based on the set pair homeopathic similar relationship needs to be investigated, while its potential scope of applications needs to be explored.

Acknowledgement

This work is partially funded by a Marie Curie International Incoming Fellowship within the 7^{th} European Community Framework Programme (Grant No. FP7-PIIF-GA-2013-629051); National Natural Science Foundation of China (71301061; 71503103), National Social Science Fund Project (12AZD111); Ministry of Education Humanities and Social Sciences Youth Fund(13YJC630120); Natural Science Foundation of Jiangsu Province (BK20150157); Jiangsu Province Social Science Fund Project(14GLC008); The research base of Chinese IOT development strategy(133930), the Fundamental Research Funds for the Central Universities(JUSRP11583; JUSRP1507ZD).

References

[1] Pawlak, Z. (1998), "Rough sets", *International Journal of Information and Computer Sciences*, Vol. 49, No. 5, pp. 415-422.

[2] Pawlak, Z., and Skowron, A. (2007a), "Rudiments of rough sets", *Information Sciences*, Vol. 1, pp. 3-27.

[3] Pawlak, Z., and Skowron, A. (2007b), "Rough sets: some extensions", *Information Sciences*, Vol. 1, pp. 28-40.

[4] Pawlak, Z., and Skowron, A. (2007c), "Rough sets and Boolean reasoning", *Information Sciences*, Vol. 1, pp. 41-73.

[5] Zhao S., Zhang L., Xu X. S.(2014). "Hierarchical description of uncertain information", *Information Sciences*, vol. 261, No. 1, pp. 133-146.

[6] Xu W. H., Wang Q. R. Zhang X. T. (2014). "Multi-granulation rough sets based on tolerance relations", *Soft Computing*, vol. 17, No. 7, pp. 1241-1252.

[7] Millo S., Reinier G.C., Deborah C.C..(2014), "Aggregation of Similarity Measures for Ortholog Detection: Validation with Measures Based on Rough Set Theory", *Computation Systems*, Vol. 18, No. 1, pp. 19-35.

[8] Zhang H. Y., Leung Y, Zhou L..(2013). "Variable-precision-dominance-based rough set approach to interval-valued information systems", *Information Sciences*, vol. 244, No. 2, pp. 75-91.

[9] Kryszkiewicz, M. (1998),"Rough set approach to incomplete information systems" , *Information Sciences*, Vol. 112, pp. 39-49.

[10] Stefanowski, J., and Tsoukias, A.. (2001), "Incomplete information tables and rough classification", *Computational Intelligence*, Vol. 17, pp. 545-566.

[11] Wang G. Y. (2002). "The extensions of rough set theory in incomplete information system", *Researcher and development of computer*, Vol. 39, No. 10, pp. 1238-1243.

[12] Greco, S., Matarazzo, B., and Slowiski, R. (2002a), "Rough approximation by dominance relations", *International Journal of Intelligent Systems*, Vol. 17, pp. 153-171.

[13] Greco, S., Matarazzo, B., and Slowiski, R. (2002b), "Rough set s theory f or multi-criteria decision analysis", *European Journal of Operational Research*, Vol. 129, pp. 1-47.

[14] Shao, M. W., and Zhang, W. X.. (2005), "Dominance relation and rules in an incomplete ordered information system", *International journal of intelligent systems*, Vol. 20, pp. 13-27.

[15] Yang, X. B., Yang, J. Y, and Chen, W., et al. (2008)." Dominance- based rough set approach and knowledge reductions in incomplete ordered information system", *Information Sciences*, 2008, vol. 178, No. 4, pp. 1219-1234.

[16] Xie, J., Song, Y. Q., Chen, J. M., et al. (2008)." Extensions of rough set model and set pair analysis in incomplete ordered decision system", *Computer science*, vol. 35, No. 12, pp. 154-157.

[17] Zhao K. Q.. (2000). "Set pair analysis and preliminary application", *Hangzhou: Zhejiang science and technology press*, pp. 68-91.

[18] Huang, B., and Zhou, X. Z. (2002), "Rough set model based on set pair analysis in incomplete system" *Computer science*, Vol. 29, pp. 1-3.

[19] Huang, B., and Zhou, X Z (2004), "The extensions of rough set theory in incomplete information system based on connection degree", *System engineering theory and practice*, Vol. 1, pp. 88-92.

[20] Xu, Y., Li, L. S., and Li, X. J.. (2008). "generalized rough set model based on set pair situation", *Journal of System Simulation*, vol. 20, No. 6, pp. 1515-1518.

[21] Xu, Y., and Li, L. S.. (2010). "Variable precision rough set model based on set pair situation", *Control and decision*, vol. 25, No. 11, pp. 1732-1736.

[22] Xu Y., and Li, L. S.. (2011). "Variable precision rough set model based on (α, β) connection degree tolerance relation", *Acta Automatica Sinica*, Vol. 37, No. 3, pp. 303-308.

[23] Tao, Z., Dai, H. J., and Zhang, Y.. (2008). "Set pair rough set in incomplete system", *Computer application*, Vol. 28, No. 7, pp. 1684-1685,1691.

[24] Liu, F. C.. (2005), "Set pair rough set model based on limited tolerance relations", *Computer science*, Vol. 32, No. 6, pp. 124-128.

[25] Liu, F. C.. (2006), "An algorithm for attributes reduction in variable precision rough set model based on set pair analysis", *Computer science*, Vol. 33, No. 3, pp. 185-187.

[26] Zhou L., and Shu, L.. (2006)." Rough set model based on new set pair analysis", *Fuzzy Systems and Mathematics*, vol. 20, No. 4, pp. 111-116.

[27] Ziarko, W. (1993a). "Variable precision rough set model", *Journal of Computer and System Sciences*, vol. 46, No. 1, pp. 39-59.

[28] Ziarko, W. (1993b)." Analysis of uncertain information in the framework of variable precision rough sets", *Foundations of Computing and Decision Sciences*, vol. 18, pp. 381-396.

[29] Ziarko, W. (1999). "Rough Sets, Fuzzy Sets and Knowledge Discovery", *Singapore: Springer*, pp. 1-98.

Corresponding author
Yong Liu can be contacted at:clly1985528@163.com