Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

Information Granulation and Approximation in a Decision-theoretic Model of Rough Sets

Y.Y. Yao

Department of Computer Science
University of Regina
Regina, Saskatchewan
Canada S4S 0A2
E-mail: yyao@cs.uregina.ca
URL: http://www.cs.uregina.ca/∼yyao

Summary. Granulation of the universe and approximation of concepts in the granulated universe are two related fundamental issues in the theory of rough sets. Many proposals dealing with the two issues have been made and studied extensively. We present a critical review of results from existing studies that are relevant to a decision-theoretic modeling of rough sets. Two granulation structures are studied, one is a partition induced by an equivalence relation and the other is a covering induced by a reflexive relation. With respect to the two granulated views of the universe, element oriented and granule oriented definitions and interpretations of lower and upper approximation operators are examined. The structures of the families of fixed points of approximation operators are investigated. We start with the notions of rough membership functions and graded set inclusion defined by conditional probability. This enables us to examine different granulation structures and the induced approximations in a decision-theoretic setting. By reviewing and combining results from existing studies, we attempt to establish a solid foundation for rough sets and to provide a systematic way for determining the required parameters in defining approximation operators.

1 Introduction

The concept of information granulation was first introduced by Zadeh in the context of fuzzy sets in 1979 [44]. The basic ideas of information granulation have appeared in fields, such as interval analysis, quantization, rough set theory, the theory of belief functions, divide and conquer, cluster analysis, machine learning, databases, and many others [45]. There is a fast growing and renewed interest in the study of information granulation and computations under the umbrella term of *Granular Computing* (GrC), covering theories, methodologies, techniques, and tools that make use of granules in the process of problem solving [4,7,8,15,17,18,21,22,28,34,43,46].

Granulation of a universe involves the decomposition of the universe into parts, or the grouping of individual elements or objects into classes, based on available information and knowledge. Elements in a granule are drawn together by indistinguishability, similarity, proximity or functionality [45]. With

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

the granulation of a universe, a subset of a universe may be considered as a whole unit rather than individual elements. A set of indistinguishable objects is considered to be a granule of the universe. One can thus form a granulated view of the universe. A natural consequence of granulation is the problem of approximating concepts using granules. The theory of rough sets can be used for constructing a granulated view of the universe and for interpreting, representing, and processing concepts in the granulated universe. It offers a more concrete model of granular computing.

The basis of the theory of rough sets is the indiscernibility or indistinguishability of objects or elements in a universe of interest [11,12]. The standard approach for modeling indiscernibility of objects is through an equivalence relation defined based on their attribute values with reference to an information table [12]. Two objects are equivalent if they have exactly the same description. The induced granulation is a partition of the universe, i.e., a family of pair-wise disjoint subsets. It is a coarsening of the universe, and is studied extensively in mathematics under the name of quotient set. The notion of indiscernibility can be generalized into similarity defined by a reflexive binary relation [9,23,24,20,30,37]. The set of objects similar to an element can be viewed as an elementary granule or a neighborhood with the element as its center. Distinct elementary granules may have non-empty overlaps, and the family of elementary granules with respect to all elements of the universe form a covering of the universe.

Each equivalence class can be viewed as a basic building block or an elementary granule. All other sets are to be represented in terms of equivalence classes (granules). Two formulations are particularly relevant to decision-theoretic modeling of rough sets [29,30,38–40]. The element oriented method is based on the notion of rough membership [13]. One can define a rough membership function with respect to a subset of the universe based on its overlaps with equivalence classes [13]. Rough membership functions can be viewed as a special type of a fuzzy membership functions. The core and support of the membership function are defined as the subsets of objects with full and non-zero memberships, respectively. They produce the lower and upper approximations [41]. The granule oriented method is based on the set inclusion relation. An equivalence class is in the lower approximation of a set if it is contained in the set, and the equivalence class is in the upper approximation of a set if it has a non-empty overlap with the set. These formulations can be extended to cases where non-equivalence relations are used [23,20,30,37].

The lower and upper approximation operators as defined by core and support of a rough membership function represent only two extreme cases. One is characterized by full membership and the other by non-zero membership. They may be regarded as qualitative approximations of a set. The actual degree of membership is not taken into consideration. Likewise, the definition by set inclusion only considers full inclusion, without considering the degree

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

of set inclusion. This makes the rough set approach to be very sensitive to the accuracy of input data and not suitable to process noisy data.

To resolve the problem with qualitative approximations, extended rough set approximations have been suggested. For the element oriented method, Pawlak *et al.* [14] introduced the notion of probabilistic rough set approximations. The conditional probabilities used to define probabilistic rough sets are in fact the rough membership functions. The lower probabilistic approximation is defined as the set of elements with membership values being greater than 0.5, the upper probabilistic approximation is the set of elements with membership values being greater than or equal to 0.5. Yao *et al.* [38,40] proposed and studied a more general type of probabilistic rough set approximations based on the Bayesian decision theory. A pair of parameters (α, β) with $\alpha > \beta$ can be determined from a loss (cost) function. The lower probabilistic approximation is defined as the set of elements with membership values being greater than or equal to α and the upper probabilistic approximation as the set of elements with membership values being greater than β.

The main results of the decision-theoretic rough set model were later given and studied again by some authors based on graded set inclusion by extending the granule oriented definition. Ziarko [47] introduced a variable precision rough set model. A measure called the degree of misclassification is defined as the inverse of the conditional probabilities or the rough membership functions. A threshold value α is used. An equivalence class is in the lower approximation if its degree of misclassification is below or equals to α. Equivalently, this means that an element belongs to the lower approximation if its membership value is greater than or equals to $1 - \alpha$. Similarly, an equivalence class is in the upper approximation if its degree of misclassification is less than $1 - \alpha$, or equivalently an element belongs to the upper approximation if its membership value is greater than α. However, unlike the decision-theoretic model, there do not exist theoretic justification and a systematic way to determine the parameter α in the variable precision rough set model, except it was suggested that value of α must be in the range $[0, 0.5)$. The parameter of the variable precision rough set model can be easily interpreted in the decision-theoretic rough set model, as shown by Yao *et al.* [39].

Other extensions of granule oriented definition are based on different measures of the degree of, or graded, inclusion of two sets. Skowron and Stepaniuk [20] introduced an abstract notion of vague inclusion. The measure of vague inclusion is a function that maps every two subsets to a value in the unit interval $[0, 1]$ and is characterized the monotonicity with respect to the second argument. An example of vague inclusion can be defined by rough membership function [20]. By extending the notion of rough membership functions to the power set of the universe, Polkowski and Skowron [16,19] introduced the notion of rough inclusion which is characterized by additional properties. Bryniarski and Wybraniec-Skardowska [1] used a family of inclusion relations

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

in defining rough set approximation. By choosing a set of threshold values, one can easily obtain a set of inclusion relations from a measure of graded inclusion. All those proposals for graded or degree of inclusion can be used to define rough set approximations, as is done in the variable precision rough set model.

The introduction of parameterized rough set approximations offers more flexibilities to the theory and extends its domain of applications. In general, the existence of parameterized measures is very useful. As there may not exist universally good measures for all data, one can fine tune the parameters to search for relevant measures from a family of parameterized measures with respect to a given set of data [25,26]. When applying this general principle to parameterized rough set approximations, one can search for relevant approximations for a given set of data. In order to achieve this goal, we need to provide intuitive interpretations of the parameters and design a systematic way to fine tune the parameters.

From the brief summary of studies related to decision-theoretic model of rough sets, we can now state the objective of this paper. By reviewing and combining results from existing studies, we attempt to establish a solid foundation for rough sets and to provide a systematic way for determining the required parameters in defining rough set approximations. Intuitive arguments and experimental investigations are important, as reported and demonstrated by many studies on generalizing rough sets based on the degree of membership or the degree of set inclusion. A solid and sound decision-theoretic foundation may provide a convincing argument and guidelines for applying the theory.

Granular computing covers many more topics, such as fuzzy if-then rules and computing with words. Partitions and coverings represent very simple granulated views of the universe. For instance, objects of the universe can have complex structures. The indistinguishability and similarity of such objects should be defined by taking into consideration their structures. A related issue is the search for suitable similarity relations and granulations for a particular application. We will not deal with these advanced topics. Instead, our discussion is restricted to topics and simple granulation structures related to rough sets, and particularly related to a decision-theoretic model of rough sets [38,40]. More specifically, we only deal with two related fundamental issues, namely, granulation and approximation. Nevertheless, the argument can be applied to granular computing in general.

The rest of the paper is organized as follows. Section 2 gives a brief overview of two granulation structures on the universe. One is defined by an equivalence relation and the other by a reflexive relation. Section 3 focuses on two definitions of rough set approximations. One is based on rough membership functions and the other on the set inclusion relation between an equivalence class and the set to be approximated. Approximation structures are discussed. Section 4 discusses a decision-theoretic model of rough sets.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

2 Granulation of Universe

The notion of indiscernibility provides a formal way to describe the relationships between elements of a universe under consideration. In the theory of rough sets, indiscernibility is modeled by an equivalence relation. A granulated view of the universe can be obtained from the equivalence classes. By generalizing equivalence relations to similarity relations characterized only by reflexivity [23,37], one may obtain a different granulation of the universe.

2.1 Granulation by equivalence relations

Let $E \subseteq U \times U$ be an equivalence relation on a finite and non-empty universe U. That is, E is reflexive, symmetric, and transitive. The equivalence relation can be defined based on available knowledge. For example, in an information table, elements in the universe are described by a set of attributes. Two elements are said to be equivalent if they have the same values with respect to some attributes [12,42]. The equivalence class,

$$[x]_E = \{y \in U \mid yEx\}, \tag{1}$$

consists of all elements equivalent to x, and is also the equivalence class containing x. The relation E induces a partition of the universe U:

$$U/E = \{[x]_E \mid x \in U\}. \tag{2}$$

That is, U/E is a family of pair-wise disjoint subsets of the universe and $\bigcup_{x \in U} [x]_E = U$. The partition is commonly known as the quotient set and provides a granulated view of the universe under the equivalence of elements. Intuitively speaking, the available knowledge only allows us to talk about an equivalence class as a single unit. In other words, under the granulated view, we consider an equivalence class as a whole instead of individuals.

The pair $apr = (U, E)$ is referred to as an approximation space, indicating the intended application of the partition U/E for approximation [11]. Each equivalence class is called an elementary granule. The elementary granules, the empty set $\emptyset$ and unions of equivalence classes are called definable granules in the sense that they can be defined precisely in terms of equivalence classes of E. The meaning of definable sets will be clearer when we discuss about approximations in the next section. Let $\mathrm{Def}(U/E)$ denote the set of all definable granules. It is closed under set complement, intersection, and union. In fact, $\mathrm{Def}(U/E)$ is a sub-Boolean algebra of the Boolean algebra formed by the power set 2^U of U and an σ-algebra of subsets of U generated by the family of equivalence classes U/E. In addition, U/E is the basis of the σ-algebra $\sigma(U/E)$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

2.2 Granulation by similarity relations

Indiscernibility as defined by an equivalence relation, or equivalently a partition of the universe, represents a very restricted type of relationships between elements of the universe. In general, the notions of similarity, and coverings of the universe, may be used [9,10,20,23,30,37].

Suppose R is a binary relation on the universe U representing the similarity of elements in U. We assume that R is at least reflexive, i.e., an element must be similar to itself, but not necessarily symmetric and transitive [23]. For two elements $x, y \in U$, if xRy we say that x is similar to y. The relation R may be more conveniently represented using the set of elements similar to x, or the predecessor neighborhood [30], as follows:

$$(x)_R = \{y \in U \mid yRx\}. \tag{3}$$

The set $(x)_R$ consists of all elements similar to x. By the assumption of reflexivity, we have $x \in (x)_R$. When R is an equivalence relation, $(x)_R$ is the equivalence class containing x. The family of predecessor neighborhoods,

$$U/R = \{(x)_R \mid x \in U\}, \tag{4}$$

is a covering of the universe, namely, $\bigcup_{x \in U} (x)_R = U$. For two elements $x, y \in U$, $(x)_R$ and $(y)_R$ may be different and have a non-empty overlap. This offers another granulated view of the universe.

Through the similarity relation, an element x is viewed by the set of elements similar to it, namely, $(x)_R$. We define an equivalence relation $\equiv_R$ on U as follows [31,35]:

$$x \equiv_R y \iff (x)_R = (y)_R. \tag{5}$$

Two elements are equivalent if they have exactly the same neighborhood. If R is an equivalence relation, then $\equiv_R$ is the same as R.

The pair $apr = (U, R)$ is referred to as a generalized approximation space. The neighborhood $(x)_R$ is called an elementary granule. Elementary granules, the empty set $\emptyset$ and unions of elementary granules are called definable granules in $apr = (U, R)$. Let $\mathrm{Def}(U/R)$ denote the set of all definable granules. It is closed under set union, and may not necessarily be closed under set complement and intersection. The set $\mathrm{Def}(U/R)$ contains both the empty set $\emptyset$ and the entire set U. From $\mathrm{Def}(U/R)$, we define

$$\mathrm{Def}^c(U/R) = \{A^c \mid A \in \mathrm{Def}(U/R)\}, \tag{6}$$

where A^c denotes the complement of A. The new system $\mathrm{Def}^c(U/R)$ contains $\emptyset$ and U and is closed under set intersection. It is commonly known as a closure system [33]. If the relation R is an equivalence relation, both systems become the same one. In general, the two systems are not the same.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

3 Rough Set Approximations

In an approximation space or a generalized approximation space, a pair of rough set approximation operators, known as the lower and upper approximation operators, can be defined in many ways [11,12,29,31,30,41]. Two definitions are discussed in this section. The element oriented definition focuses on the belongingness of a particular element to the lower and upper approximations of a set. The granule oriented definition focuses on the belongingness of an entire granule to the lower and upper approximations [35]. While the two definitions produce the same results in an approximation space $apr = (U, E)$, they produce different results in a generalized approximation space $apr = (U, R)$.

We pay special attention to two families of subsets of the universe. One consists of those subsets whose lower approximations are the same as themselves, i.e., the fixed points of lower approximation operator. The other consists of those subsets whose upper approximations are the same as themselves, i.e., the fixed points of upper approximation operator. The structures of the two families show the structures and consequences of different granulation methods, and may provide more insights into our understanding of approximation operators.

3.1 Rough membership functions

In an approximation space $apr = (U, E)$, an element $x \in U$ belongs to one and only one equivalence class $[x]_E$. For a subset $A \subseteq U$, a rough membership function is defined by [13]:

$$\mu_A(x) = \frac{|[x]_E \cap A|}{|[x]_E|}, \tag{7}$$

where $|\cdot|$ denotes the cardinality of a set. The rough membership value $\mu_A(x)$ may be interpreted as the conditional probability that an arbitrary element belongs to A given that the element belongs to $[x]_E$. In fact, conditional probabilities were used earlier in the development of a probabilistic rough set model [14,27,38,40].

Rough membership functions may be interpreted as fuzzy membership functions interpretable in terms of probabilities defined simply by the cardinalities of sets [32,35,41]. With this interpretation, one can define at most $2^{|U|}$ fuzzy sets. Two distinct subsets of U may derive the same rough membership function. By definition, the membership values are all rational numbers.

The theory of fuzzy sets is typically developed as an uninterpreted mathematical theory of abstract membership functions without the above limitations [6]. In contrast, the theory of rough set provides a more specific and more concrete interpretation of fuzzy membership functions. The source of the fuzziness in describing a concept is the indiscernibility of elements. The

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

limitations and constraints of such an interpreted sub-theory should not be viewed as the disadvantages of the theory. In fact, such constraints suggest conditions that may be verified when applying the theory to real world problems. It might be more instructive and informative, if one knows that a certain theory cannot be applied. Explicit statements of conditions under which a particular model is applicable may prevent misuse of the theory.

When interpreting fuzzy membership functions in the theory of rough sets, we have the constraints:

$$
\begin{array}{ll}
\text{(m1)} & \mu_U(x) = 1, \\
\text{(m2)} & \mu_\emptyset(x) = 0, \\
\text{(m3)} & y \in [x]_E \Longrightarrow \mu_A(x) = \mu_A(y), \\
\text{(m4)} & x \in A \Longrightarrow \mu_A(x) \neq 0, \\
\text{(m5)} & x \notin A \Longrightarrow \mu_A(x) \neq 1, \\
\text{(m6)} & \mu_A(x) = 1 \Longleftrightarrow [x]_E \subseteq A, \\
\text{(m7)} & \mu_A(x) > 0 \Longleftrightarrow [x]_E \cap A \neq \emptyset, \\
\text{(m8)} & A \subseteq B \Longrightarrow \mu_A(x) \leq \mu_B(x).
\end{array}
$$

Property (m3) is particularly important. It shows that elements in the same equivalence class must have the same degree of membership. That is, indiscernible elements should have the same membership value. Such a constraint, which ties the membership values of individual elements according to their connections, is intuitively appealing. Although this topic has been investigated by some authors [2], there is still a lack of systematic study. Properties (m4) and (m5) states that an element in A cannot have a zero membership value, and an element not in A cannot have a full membership. They can be equivalently expressed as:

$$
\begin{array}{ll}
\text{(m4)} & \mu_A(x) = 0 \Longrightarrow x \notin A, \\
\text{(m5)} & \mu_A(x) = 1 \Longrightarrow x \in A,
\end{array}
$$

According to properties (m6) and (m7), $\mu_A(x) = 1$ if and only if for all $y \in U$, $x \in [y]_E$ implies $y \in A$, and $\mu_A(x) > 0$ if and only if there exists a $y \in U$ such that $y \in A$ and $x \in [y]_E$. Since $x \in [x]_E$, property (m5) is a special case of (m6). Property (m8) suggests that a rough membership function is monotonic with respect to set inclusion.

In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation, for a subset A of the universe, a rough membership function can be defined by substituting $[x]_E$ with $(x)_R$ in equation (7) as follows [31,32]:

$$
\mu_A(x) = \frac{|(x)_R \cap A|}{|(x)_R|}. \tag{8}
$$

By the reflexivity of R, one can verify that properties (m1), (m2), and (m4)-(m8) also hold, provided that $[x]_E$ is replaced by $(x)_R$. For (m3), we can have

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

the weak version:

$$\text{(m3a)} \qquad (y \in (x)_R, \mu_A(x) = 1) \Longrightarrow \mu_A(y) \neq 0,$$
$$\text{(m3b)} \qquad (y \in (x)_R, \mu_A(x) = 0) \Longrightarrow \mu_A(y) \neq 1.$$

These two properties are of qualitative nature. They state that membership values of related elements are related. If an element y is similar to another element x with full membership, then y cannot have a null membership. Likewise, if y is similar to an element x with null membership, then y cannot have a full membership. Properties (m3a) and (m3b) can also be expressed as:

$$\text{(m3a)} \qquad (y \in (x)_R, \mu_A(y) = 0) \Longrightarrow \mu_A(x) \neq 1,$$
$$\text{(m3b)} \qquad (y \in (x)_R, \mu_A(y) = 1) \Longrightarrow \mu_A(x) \neq 0.$$

They can be similarly interpreted. With respect to the equivalence relation $\equiv_R$, we have the property:

$$\text{(m3c)} \qquad x \equiv_R y \Longrightarrow \mu_A(x) = \mu_A(y).$$

It is closer to the original (m3) of the standard rough membership function. All these properties appear to be intuitively sound and meaningful.

A binary relation only defines a dichotomous relationship. Two elements are either related or not related. It is not surprising that we can only draw conclusions with respect to elements with null or full membership, as indicated by the previously stated properties.

The constraints on rough membership functions have significant implications on rough set operators. Rough membership functions corresponding to A^c, $A \cap B$, and $A \cup B$ must be defined using set operators and equation (7) or equation (8).

By laws of probability, we have:

$$\text{(o1)} \qquad \mu_{A^c}(x) = 1 - \mu_A(x),$$
$$\text{(o2)} \qquad \mu_{A \cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_{A \cap B}(x),$$
$$\text{(o3)} \qquad \max(0, \mu_A(x) + \mu_B(x) - 1) \leq \mu_{A \cap B}(x) \leq \min(\mu_A(x), \mu_B(x)),$$
$$\text{(o4)} \qquad \max(\mu_A(x), \mu_B(x)) \leq \mu_{A \cup B}(x) \leq \min(1, \mu_A(x) + \mu_B(x)),$$

Unlike the commonly used fuzzy set operators as typically defined by t-norms and t-conorms [6], the new intersection and union operators are non-truth-functional. That is, it is impossible to obtain rough membership functions of $A \cap B$ and $A \cup B$ based solely on the rough membership functions of A and B. One must also consider their overlaps and their relationships to the equivalence class $[x]_E$ or the predecessor neighborhood $(x)_R$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

One can verify the following additional properties corresponding to the properties of t-norms and t-conorms:

(t1) Boundary conditions
$(\mu_A(x) = 0, \mu_B(x) = 0) \Longrightarrow \mu_{A\cap B}(x) = 0,$
$(\mu_A(x) = 1, \mu_B(x) = a) \Longrightarrow \mu_{A\cap B}(x) = a,$
$(\mu_A(x) = a, \mu_B(x) = 1) \Longrightarrow \mu_{A\cap B}(x) = a,$

(t2) Monotonicity
$(A \subseteq C, B \subseteq D) \Longrightarrow \mu_{A\cap B}(x) \le \mu_{C\cap D}(x),$

(t3) Symmetry
$\mu_{A\cap B}(x) = \mu_{B\cap A}(x),$

(t4) Associativity
$\mu_{A\cap(B\cap C)}(x) = \mu_{(A\cap B)\cap C}(x);$

(s1) Boundary conditions
$(\mu_A(x) = 1, \mu_B(x) = 1) \Longrightarrow \mu_{A\cup B}(x) = 1,$
$(\mu_A(x) = 0, \mu_B(x) = a) \Longrightarrow \mu_{A\cup B}(x) = a,$
$(\mu_A(x) = a, \mu_B(x) = 0) \Longrightarrow \mu_{A\cup B}(x) = a,$

(s2) Monotonicity
$(A \subseteq C, B \subseteq D) \Longrightarrow \mu_{A\cup B}(x) \le \mu_{C\cup D}(x),$

(s3) Symmetry
$\mu_{A\cup B}(x) = \mu_{B\cup A}(x),$

(s4) Associativity
$\mu_{A\cup(B\cup C)}(x) = \mu_{(A\cup B)\cup C}(x).$

The boundary conditions follow from (o3) and (o4), and monotonicity follows from (m8). While other properties are very close to the properties of t-norms and t-conorms, the monotonicity property is much weaker than the monotonicity of a t-norm t and a t-conorm s, i.e., $(a \le c, b \le d) \Longrightarrow t(a, b) \le t(c, d)$ and $(a \le c, b \le d) \Longrightarrow s(a, b) \le s(c, d)$. For four arbitrary sets A, B, C, D with $\mu_A(x) = a$, $\mu_B(x) = b$, $\mu_C(x) = c$, $\mu_D(x) = d$, $a \le c$ and $b \le d$, $\mu_{A\cap B}(x) \le \mu_{C\cap D}(x)$ and $\mu_{A\cup B}(x) \le \mu_{C\cup D}(x)$ may not necessarily hold.

3.2 Element oriented approximations

In an approximation space $apr = (U, E)$, we define a rough membership function μ_A for a subset $A \subseteq U$. By collecting elements with full and non-zero memberships, respectively, we obtain a pair of lower and upper approximations of A as follows:

$$\underline{apr}(A) = \{x \in U \mid \mu_A(x) = 1\} = core(\mu_A),$$
$$\overline{apr}(A) = \{x \in U \mid \mu_A(x) > 0\} = support(\mu_A). \quad (9)$$

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

They are indeed the core and support of the fuzzy set μ_A. An equivalent and more convenient definition without using membership functions is given by:

$$\begin{aligned} \underline{apr}(A) &= \{x \in U \mid [x]_E \subseteq A\}, \\ \overline{apr}(A) &= \{x \in U \mid [x]_E \cap A \neq \emptyset\}. \end{aligned} \tag{10}$$

The lower and upper approximations can be interpreted as a pair of unary set-theoretic operators, $\underline{apr}, \overline{apr} : 2^U \longrightarrow 2^U$. They are dual operators in the sense that $\underline{apr}(A) = (\overline{apr}(A^c))^c$ and $\overline{apr}(A) = (\underline{apr}(A^c))^c$. Other properties of approximation operators can be found in many articles [5,11,12,31,37,39].

In this definition, we focus on whether a particular element is in the lower and upper approximations. It is thus referred to as the element oriented definition of rough set approximations. More specifically, an element $x \in U$ belongs to the lower approximation of A if *all* its equivalent elements belong to A. It belongs to the upper approximation of A if at least *one* of its equivalent elements belongs to A. The element oriented interpretation of approximation operators is related to the interpretation of the necessity and possibility operators in modal logic [29,37].

So far, we have shown that, as a consequence of granulation, a set A is viewed differently. The fuzzification of A leads to a rough membership function, and the approximation of A leads to a pair of sets. Moreover, approximations of a set can be viewed as a qualitative characterization of a rough membership function using the core and support. A study of families of sets that are invariant under fuzzification and approximation may bring more insights into the understanding of granulation structures.

A set A is said to be a lower exact set if $A = \underline{apr}(A)$, an upper exact set if $A = \overline{apr}(A)$, and a lower and an upper exact set if $\underline{apr}(A) = A = \overline{apr}(A)$. Lower exact sets are fixed points of the lower approximation operator $\underline{apr}$ and upper exact sets are fixed points of the upper approximation operator $\overline{apr}$. Let

$$\begin{aligned} \mathrm{E}(\underline{apr}) &= \{A \subseteq U \mid A = \underline{apr}(A)\}, \\ \mathrm{E}(\overline{apr}) &= \{A \subseteq U \mid A = \overline{apr}(A)\}, \end{aligned} \tag{11}$$

be the set of lower exact sets and the set of upper exact sets, respectively. By definition, we immediately have the following results.

Theorem 1. *In an approximation space $apr = (U, E)$, we have:*

$$\mathrm{E}(\underline{apr}) = \mathrm{E}(\overline{apr}) = \mathrm{Def}(U/E). \tag{12}$$

Theorem 2. *In an approximation space $apr = (U, E)$, we have:*

$$\mu_A(x) = \chi_A(x), \text{ for all } x \in U, \tag{13}$$

if and only if $A \in \mathrm{Def}(U/E)$, where χ_A is the characteristic function of A defined by $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

Theorem 1 shows that a set in $\mathrm{Def}(U/E)$ is both lower and upper exact, and only a set in $\mathrm{Def}(U/E)$ has such a property. For this reason, a set in $\mathrm{Def}(U/E)$ is called a definable set. Theorem 2 states that μ_A is a crisp set if and only if $A \in \mathrm{Def}(U/E)$. All other subsets of U will induce non-crisp fuzzy sets. The fuzziness is a natural consequence of indiscernibility of elements.

In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, rough set approximations can be defined by replacing $[x]_E$ with $(x)_R$ in equations (9) and (10) as follows:

$$\begin{aligned} \underline{apr}(A) &= \{x \in U \mid (x)_R \subseteq A\} \\ &= \{x \in U \mid \mu_A(x) = 1\} = core(\mu_A), \\ \overline{apr}(A) &= \{x \in U \mid (x)_R \cap A \neq \emptyset\} \\ &= \{x \in U \mid \mu_A(x) > 0\} = support(\mu_A). \end{aligned} \tag{14}$$

The results regarding fuzzification, as well as lower and upper exact sets, are summarized in the following theorems.

Theorem 3. *In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, we have:*

a. $A = \underline{apr}(A)$ *if and only if* $A = \bigcup_{x \in A} (x)_R$,

$A = \overline{apr}(A)$ *if and only if* $A = \bigcap_{x \notin A} ((x)_R)^c$,

b. $\mathrm{E}(\underline{apr}) \subseteq \mathrm{Def}(U/R)$, $\mathrm{E}(\overline{apr}) \subseteq \mathrm{Def}^c(U/R)$,

c. $\mathrm{E}(\underline{apr})$ *and* $\mathrm{E}(\overline{apr})$ *are closed under* $\cap$ *and* $\cup$,

d. $A \in \mathrm{E}(\underline{apr})$ *if and only if* $A^c \in \mathrm{E}(\overline{apr})$,

e. $\mathrm{E}(\underline{apr}) \cap \mathrm{E}(\overline{apr})$ *is a sub–Boolean algebra of* 2^U.

Theorem 4. *In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, we have:*

$$\mu_A(x) = \chi_A(x), \text{ for all } x \in U, \tag{15}$$

if and only if $A \in \mathrm{E}(\underline{apr}) \cap \mathrm{E}(\overline{apr})$.

The sets $\mathrm{E}(\underline{apr})$ and $\mathrm{E}(\overline{apr})$ are not necessarily the same and may not be closed under set complement. While $\mathrm{E}(\underline{apr})$ is a sub-family of $\mathrm{Def}(U/R)$ closed under both $\cap$ and $\cup$, $\mathrm{E}(\overline{apr})$ is a sub-family of $\mathrm{Def}^c(U/R)$ closed under both $\cap$ and $\cup$. A lower exact set must be expressed as a union of some elementary granules. However, not every union of elementary granules is a lower exact set. A set A is lower and upper exact if and only if μ_A is a crisp set.

In defining rough set approximations, only the two extreme points of the unit interval $[0, 1]$ are used, namely, 0 is used for upper approximations and 1

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

for lower approximations. In general, we can use a pair of values (α, β) with $\alpha > \beta$ to define a pair of graded lower and upper approximations:

$$\begin{aligned} \underline{apr}_{\alpha}(A) &= \{x \in U \mid \mu_A(x) \geq \alpha\} = (\mu_A)_{\alpha}, \\ \overline{apr}_{\beta}(A) &= \{x \in U \mid \mu_A(x) > \beta\} = (\mu_A)_{\beta+}, \end{aligned} \tag{16}$$

where $(\mu_A)_{\alpha}$ denotes the α-cut of the fuzzy set μ_A and $(\mu_A)_{\beta+}$ the strong β-cut of μ_A. The condition $\alpha > \beta$ implies that $\underline{apr}_{\alpha} \subseteq \overline{apr}_{\beta}(A)$. When $\alpha = \beta$, in order to keep this property, we define:

$$\begin{aligned} \underline{apr}_{\alpha}(A) &= \{x \in U \mid \mu_A(x) > \alpha\} = (\mu_A)_{\alpha+}, \\ \overline{apr}_{\alpha}(A) &= \{x \in U \mid \mu_A(x) \geq \alpha\} = (\mu_A)_{\alpha}, \end{aligned} \tag{17}$$

By imposing an additional condition $\alpha + \beta = 1$, we can obtain a pair of dual operators [47,39]. For standard approximation operators, we have:

$$\begin{aligned} \underline{apr}(A) &= \underline{apr}_{1}(A), \\ \overline{apr}(A) &= \overline{apr}_{0}(A). \end{aligned} \tag{18}$$

The probabilistic rough set approximation operators proposed by Pawlak *et al.*[14] is given by $(\underline{apr}_{0.5}, \overline{apr}_{0.5})$. Properties of rough set approximations under the pair of parameters (α, β) can be found in [37–40,47].

We can define the notions of lower exact sets and upper exact sets of graded approximation operators. In an approximation space $apr = (U, E)$, from property (m3) we can conclude that an entire equivalence class is either in or not in a lower or an upper approximation. This implies that $\underline{apr}_{\alpha}(A)$ and $\overline{apr}_{\beta}(A)$ must be in $\text{Def}(U/E)$. Conversely, if A is in $\text{Def}(U/E)$, $\underline{apr}_{\alpha}(A) = A$ and $\overline{apr}_{\beta}(A) = A$ for $\alpha \in (0, 1]$ and $\beta \in [0, 1)$. It therefore follows that graded approximations do not change families of lower and upper exact sets.

Theorem 5. *In an approximation space $apr = (U, E)$, for $\alpha \in (0, 1]$ and $\beta \in [0, 1)$ we have:*

$$\begin{aligned} \text{E}(\underline{apr}_{\alpha}) &= \{A \subseteq U \mid A = \underline{apr}_{\alpha}(A)\} = \text{E}(\underline{apr}) = \text{Def}(U/E), \\ \text{E}(\overline{apr}_{\beta}) &= \{A \subseteq U \mid A = \overline{apr}_{\beta}(A)\} = \text{E}(\overline{apr}) = \text{Def}(U/E). \end{aligned} \tag{19}$$

The result of Theorem 5 cannot be easily extended to a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R. The characterization of families of graded lower exact sets and graded upper exact sets in a generalized approximation space is an interesting problem.

Theorem 6. *In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, for $\alpha \in (0, 1]$ and $\beta \in [0, 1)$ we have:*

$$\begin{aligned} \text{E}(\underline{apr}_{\alpha}) &= \{A \subseteq U \mid A = \underline{apr}_{\alpha}(A)\} \subseteq \text{Def}(U/\equiv_R), \\ \text{E}(\overline{apr}_{\beta}) &= \{A \subseteq U \mid A = \overline{apr}_{\beta}(A)\} \subseteq \text{Def}(U/\equiv_R), \end{aligned} \tag{20}$$

where $\text{Def}(U/\equiv_R)$ is the family of definable sets defined by the equivalence relation $\equiv_R$.

The theorem easily follows from the property (m3c). The families $\text{E}(\underline{apr}_{\alpha})$ and $\text{E}(\overline{apr}_{\beta})$ may not necessarily be closed under $\cap$ and $\cup$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

3.3 Graded inclusion of sets

A rough membership function is defined based on the relationship between two sets, one is the equivalence class $[x]_R$ (or the neighborhood $(x)_R$) of an element x and the other is a set A. For the maximum membership value 1, we have $[x]_E \subseteq A$, namely, $[x]_E$ is a subset of A. For the minimum membership value 0, we have $[x]_E \cap A = \emptyset$, or equivalently $[x]_E \subseteq A^c$, namely, $[x]_E$ is totally not a subset of A. For a value between 0 and 1, it may be interpreted as the degree to which $[x]_E$ is a subset of A. By extending the notion of rough membership functions to power set of the universe, one obtains a measure of graded inclusion of two sets [19,20]:

$$v(A, B) = \frac{|A \cap B|}{|A|}. \tag{21}$$

For the case where $A = \emptyset$, we define $v(\emptyset, B) = 1$ and $v(\emptyset, \emptyset) = 1$, namely, the empty set is a subset of any set.

The value $v(A, B)$ can be interpreted as the conditional probability that a randomly selected element from A belongs to B. It may be used to measure the degree to which A is a subset of B. There is a close connection between graded inclusion and fuzzy set inclusion [20].

Measures related to v have been proposed and used by many authors. Ziarko [47] used the measure,

$$c(A, B) = 1 - v(A, B) = 1 - \frac{|A \cap B|}{|A|}, \tag{22}$$

in a variable precision rough set model. One can easily obtain the same results by using v. Skowron and Stepaniuk [20] suggested that graded (vague) inclusion of sets may be measured by a function,

$$v : 2^U \times 2^U \longrightarrow [0, 1], \tag{23}$$

with monotonicity regarding the second argument, namely, for $A, B, C \subseteq U$, $v(A, B) \leq (A, C)$ for any $B \subseteq C$. The function defined by equation (21) is an example of such a measure. In fact, equation (21) considers only the overlap with the first argument, but not the size of the second argument.

Starting from rough membership functions, Skowron and Polkowski [19] introduced the concept of rough inclusion defined by a function $v : 2^U \times 2^U \longrightarrow [0, 1]$ satisfying more properties, in addition to the monotonicity with respect to the second argument. The unit interval $[0, 1]$ can also be generalized to a complete lattice in the definition of rough inclusion [16]. Rough inclusion is only an example for measuring degrees of inclusion in rough mereology. A more detailed discussion on rough mereology and related concepts can be found in [16,19]. Instead of using a measure of graded inclusion, Bryniarski and Wybraniec-Skardowska [1] proposed to use a family of inclusion relations called context relations, indexed by a bounded and partially ordered set called rank set. The unit interval $[0, 1]$ can be treated as a rank set. From a measure

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

of graded inclusion, a context relation with respect to a value $\alpha \in [0,1]$ can be defined by:

$$\subseteq_\alpha = \{(A,B) \mid v(A,B) \geq \alpha\}. \tag{24}$$

In other words, v may be interpreted as a fuzzy relation on 2^U, and $\subseteq_\alpha$ may be interpreted as an α-cut of the fuzzy relation. The use of a complete lattice, or a rank set, corresponds to lattice based fuzzy relations in the theory of fuzzy sets.

3.4 Granule oriented approximations

In an approximation space $apr = (U, E)$, an equivalence class $[x]_E$ is treated as a unit. A granule oriented definition of approximation operators can be used. Approximations of a set are expressible in terms of unions of equivalence granules, namely:

$$\begin{aligned} \underline{apr}(A) &= \bigcup\{[x]_E \mid [x]_E \subseteq A\}, \\ \overline{apr}(A) &= \bigcup\{[x]_E \mid [x]_E \cap A \neq \emptyset\}. \end{aligned} \tag{25}$$

The lower approximation $\underline{apr}(A)$ is the union of those equivalence granules which are subsets of A. The upper approximation $\overline{apr}(A)$ is the union of those equivalence granules which have non-empty intersections with A. This definition is equivalent to the element oriented definition.

In a generalized approximation space $apr = (U, R)$, granule oriented rough set approximations can be defined by generalizing equation (25). The equivalence class $[x]_E$ is replaced by the neighborhood $(x)_R$. One of such generalizations is [30]:

$$\begin{aligned} \underline{apr}'(A) &= \bigcup\{(x)_R \mid x \in U, (x)_R \subseteq A\}, \\ \overline{apr}'(A) &= (\underline{apr}'(A^c))^c. \end{aligned} \tag{26}$$

We generalize the lower approximation as a union of elementary granules and define the upper approximation through duality. While the lower approximation is the union of some granules in U/R, the upper approximation cannot be expressed in this way [30]. The approximation operators $\underline{apr}'$ and $\overline{apr}'$ are different from the element oriented definition. The lower exact sets and upper exact sets are related to $\mathrm{Def}(U/R)$ and $\mathrm{Def}^c(U/R)$.

Theorem 7. *In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, we have:*

$$\begin{aligned} \mathrm{E}(\underline{apr}') &= \{A \subseteq U \mid A = \underline{apr}'(A)\} = \mathrm{Def}(U/R), \\ \mathrm{E}(\overline{apr}') &= \{A \subseteq U \mid A = \overline{apr}'(A)\} = \mathrm{Def}^c(U/R). \end{aligned} \tag{27}$$

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

This theorem generalizes the result of Theorem 1 in the sense that Theorem 1 considers a sub-class of reflexive relations.

Granule oriented approximations can be generalized into graded approximations through a graded inclusion measures. We can replace the relation $\subseteq$ by a relation $\subseteq_\alpha$ for $\alpha \in (0,1]$. For an approximation space $apr = (U, E)$, we define a lower approximation operator as:

$$\begin{aligned} \underline{apr}_\alpha(A) &= \bigcup\{[x]_E \mid [x]_E \subseteq_\alpha A\} \\ &= \bigcup\{[x]_E \mid v([x]_E, A) \geq \alpha\}. \end{aligned} \tag{28}$$

The graded upper approximation operator can be defined by another parameter $\beta \in [0,1)$. If a pair of dual operators is needed, the corresponding graded upper approximation operator can be defined by the dual of $\underline{apr}_\alpha$. The granule oriented graded approximations are the same as those obtained from the element oriented definition.

For a generalized approximation space $apr = (U, R)$, with respect to a value $\alpha \in (0,1]$, we define:

$$\begin{aligned} \underline{apr'}_\alpha(A) &= \bigcup\{(x)_R \mid (x)_R \subseteq_\alpha A\} \\ &= \bigcup\{(x)_R \mid v((x)_R, A) \geq \alpha\}. \end{aligned} \tag{29}$$

The graded upper approximation can be defined by duality. The granule oriented definition produces different approximations from the element oriented definition. By definition, the graded lower approximation of a set can be expressed as the union of some granules. However, not every union of granules can be the lower approximation of a certain set.

Theorem 8. *In a generalized approximation space $apr = (U, R)$ defined by a reflexive relation R, for $\alpha \in (0,1]$ we have:*

$$\begin{aligned} \mathrm{E}(\underline{apr'}_\alpha) &= \{A \subseteq U \mid A = \underline{apr'}_\alpha(A)\} \subseteq \mathrm{E}(\underline{apr'}) = \mathrm{Def}(U/R), \\ \mathrm{E}(\overline{apr}'_{1-\alpha}) &= \{A \subseteq U \mid A = \overline{apr}'_{1-\alpha}(A)\} = \mathrm{E}(\overline{apr}') = \mathrm{Def}^c(U/R). \end{aligned} \tag{30}$$

Both families $\mathrm{E}(\underline{apr'}_\alpha)$ and $\mathrm{E}(\overline{apr}'_{1-\alpha})$ may not necessarily closed under $\cap$ and $\cup$. In general, we have $\mathrm{E}(\underline{apr'}) \neq \mathrm{E}(\underline{apr'}_\alpha)$.

This section not only summarizes the main results from existing studies on rough set approximations in a unified framework, but also presents many new results. From the discussion, we can conclude that parameterized rough set approximations are useful and need further investigation.

4 A Decision-theoretic Model of Rough Sets

In this section, the basic notions of the Bayesian decision procedure for classification is briefly reviewed [3]. Rough set approximation operators are formulated as classifying objects into three disjoint classes, namely, the positive, negative and boundary regions.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

For clarity, we only consider the element oriented definition with respect to the granulated view of the universe induced by an equivalence relation. The same argument can be easily applied to other cases.

4.1 An overview of the Bayesian decision procedure

Let $\Omega = \{w_1, \ldots, w_s\}$ be a finite set of s states, and let $\mathcal{A} = \{a_1, \ldots, a_m\}$ be a finite set of m possible actions. Let $P(w_j|\mathbf{x})$ be the conditional probability of an object x being in state w_j given that the object is described by $\mathbf{x}$. In the following discussions, we assume that these conditional probabilities $P(w_j|\mathbf{x})$ are known.

Let $\lambda(a_i|w_j)$ denote the loss, or cost, for taking action a_i when the state is w_j. For an object with description $\mathbf{x}$, suppose action a_i is taken. Since $P(w_j|\mathbf{x})$ is the probability that the true state is w_j given $\mathbf{x}$, the expected loss associated with taking action a_i is given by:

$$R(a_i|\mathbf{x}) = \sum_{j=1}^{s} \lambda(a_i|w_j)P(w_j|\mathbf{x}). \tag{31}$$

The quantity $R(a_i|\mathbf{x})$ is also called the conditional risk. Given description $\mathbf{x}$, a decision rule is a function $\tau(\mathbf{x})$ that specifies which action to take. That is, for every $\mathbf{x}$, $\tau(\mathbf{x})$ assumes one of the actions, $a_1, \ldots, a_m$. The overall risk $\mathbf{R}$ is the expected loss associated with a given decision rule. Since $R(\tau(\mathbf{x})|\mathbf{x})$ is the conditional risk associated with action $\tau(\mathbf{x})$, the overall risk is defined by:

$$\mathbf{R} = \sum_{\mathbf{x}} R(\tau(\mathbf{x})|\mathbf{x})P(\mathbf{x}), \tag{32}$$

where the summation is over the set of all possible descriptions of objects, i.e., the knowledge representation space. If $\tau(\mathbf{x})$ is chosen so that $R(\tau(\mathbf{x})|\mathbf{x})$ is as small as possible for every $\mathbf{x}$, the overall risk $\mathbf{R}$ is minimized.

The Bayesian decision procedure can be formally stated as follows. For every $\mathbf{x}$, compute the conditional risk $R(a_i|\mathbf{x})$ for $i = 1, \ldots, m$ defined by equation (31), and then select the action for which the conditional risk is minimum. If more than one action minimizes $R(a_i|\mathbf{x})$, any tie-breaking rule can be used.

4.2 Rough set approximation operators

In an approximation space $apr = (U, E)$, with respect to a subset $A \subseteq U$, one can divide the universe U into three disjoint regions, the positive region $\mathrm{POS}(A)$, the negative region $\mathrm{NEG}(A)$, and the boundary region $\mathrm{BND}(A)$:

$$\begin{aligned} \mathrm{POS}(A) &= \underline{apr}(A), \\ \mathrm{NEG}(A) &= U - \overline{apr}(A), \\ \mathrm{BND}(A) &= \overline{apr}(A) - \underline{apr}(A). \end{aligned} \tag{33}$$

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

The lower approximation of a set is the same as the positive region. The upper approximation is the union of the positive and boundary regions, $\overline{apr}(A) = \mathrm{POS}(A) \cup \mathrm{BND}(A)$. One can say with certainty that any element $x \in \mathrm{POS}(A)$ belongs to A, and that any element $x \in \mathrm{NEG}(A)$ does not belong to A. One cannot decide with certainty whether or not an element $x \in \mathrm{BND}(A)$ belongs to A.

In an approximation space $apr = (U, E)$, an element x is viewed as $[x]_E$. That is, the equivalence class containing x is considered to be a description of x. The classification of objects according to approximation operators can be easily fitted into the Bayesian decision-theoretic framework. The set of states is given by $\Omega = \{A, \neg A\}$ indicating that an element is in A and not in A, respectively. We use the same symbol to denote both a subset A and the corresponding state. With respect to three regions, the set of actions is given by $\mathcal{A} = \{a_1, a_2, a_3\}$, where a_1, a_2, and a_3 represent the three actions in classifying an object, deciding $\mathrm{POS}(A)$, deciding $\mathrm{NEG}(A)$, and deciding $\mathrm{BND}(A)$, respectively.

Let $\lambda(a_i|A)$ denote the loss incurred for taking action a_i when an object in fact belongs to A, and let $\lambda(a_i|\neg A)$ denote the loss incurred for taking the same action when the object does not belong to A. The rough membership values $\mu_A(x) = P(A|[x]_E)$ and $\mu_{A^c}(x) = P(\neg A|[x]_E) = 1 - P(A|[x]_E)$ are in fact the probabilities that an object in the equivalence class $[x]_E$ belongs to A and $\neg A$, respectively. The expected loss $R(a_i|[x]_E)$ associated with taking the individual actions can be expressed as:

$$
\begin{aligned}
R(a_1|[x]_E) &= \lambda_{11} P(A|[x]_E) + \lambda_{12} P(\neg A|[x]_E), \\
R(a_2|[x]_E) &= \lambda_{21} P(A|[x]_E) + \lambda_{22} P(\neg A|[x]_E), \\
R(a_3|[x]_E) &= \lambda_{31} P(A|[x]_E) + \lambda_{32} P(\neg A|[x]_E), \qquad (34)
\end{aligned}
$$

where $\lambda_{i1} = \lambda(a_i|A)$, $\lambda_{i2} = \lambda(a_i|\neg A)$, and $i = 1, 2, 3$. The Bayesian decision procedure leads to the following minimum-risk decision rules:

(P) If $R(a_1|[x]_E) \leq R(a_2|[x]_E)$ and $R(a_1|[x]_E) \leq R(a_3|[x]_E)$, decide $\mathrm{POS}(A)$;

(N) If $R(a_2|[x]_E) \leq R(a_1|[x]_E)$ and $R(a_2|[x]_E) \leq R(a_3|[x]_E)$, decide $\mathrm{NEG}(A)$;

(B) If $R(a_3|[x]_E) \leq R(a_1|[x]_E)$ and $R(a_3|[x]_E) \leq R(a_2|[x]_E)$, decide $\mathrm{BND}(A)$.

Tie-breaking rules should be added so that each element is classified into only one region. Since $P(A|[x]_E) + P(\neg A|[x]_E) = 1$, the above decision rules can be simplified such that only the probabilities $P(A|[x]_E)$ are involved. We can classify any object in the equivalence class $[x]_E$ based only on the probabilities $P(A|[x]_E)$, i.e., the rough membership values, and the given loss function λ_{ij} $(i = 1, 2, 3;\ j = 1, 2)$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

Consider a special kind of loss functions with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$. That is, the loss of classifying an object x belonging to A into the positive region POS(A) is less than or equal to the loss of classifying x into the boundary region BND(A), and both of these losses are strictly less than the loss of classifying x into the negative region NEG(A). The reverse order of losses is used for classifying an object that does not belong to A. For this type of loss functions, the minimum-risk decision rules (P)-(B) can be written as:

(P) If $P(A|[x]_E) \geq \gamma$ and $P(A|[x]_E) \geq \alpha$, decide POS(A);
(N) If $P(A|[x]_E) \leq \beta$ and $P(A|[x]_E) \leq \gamma$, decide NEG(A);
(B) If $\beta \leq P(A|[x]_E) \leq \alpha$, decide BND($A$);

where

$$\begin{aligned}\alpha &= \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})},\\ \gamma &= \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})},\\ \beta &= \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}. \end{aligned} \tag{35}$$

By the assumptions, $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, it follows that $\alpha \in (0, 1]$, $\gamma \in (0, 1)$, and $\beta \in [0, 1)$.

A loss function should be chosen in such a way to satisfy the condition $\alpha \geq \beta$. This ensures that the results are consistent with rough set approximations. More specifically, the lower approximation is a subset of the upper approximation, and the boundary region may be non-empty.

Theorem 9. *If a loss function with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ satisfies the condition:*

$$(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \tag{36}$$

then $\alpha \geq \gamma \geq \beta$.

Let $l = (\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31})$ and $r = (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22})$. While l is the product of the differences between the cost of making an incorrect classification and cost of classifying an element into the boundary region, r is the product of the differences between the cost of classifying an element into the boundary region and the cost of a correct classification. A larger value of l, or equivalently a smaller value of r, can be obtained if we move λ_{32} away from λ_{12}, or move λ_{31} away from λ_{21}. In fact, the condition can be intuitively interpreted as saying that cost of classifying an element into the boundary region is closer to the cost of a correct classification than to the cost of an incorrect classification. Such a condition seems to be reasonable.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

When $\alpha > \beta$, we have $\alpha > \gamma > \beta$. After tie-breaking, we obtain the decision rules:

(P1) If $P(A|[x]_E) \geq \alpha$, decide POS(A);
(N1) If $P(A|[x]_E) \leq \beta$, decide NEG(A);
(B1) If $\beta < P(A|[x]_E) < \alpha$, decide BND($A$).

When $\alpha = \beta$, we have $\alpha = \gamma = \beta$. In this case, we use the decision rules:

(P2) If $P(A|[x]_E) > \alpha$, decide POS(A);
(N2) If $P(A|[x]_E) < \alpha$, decide NEG(A);
(B2) If $P(A|[x]_E) = \alpha$, decide BND(A).

For the second set of decision rules, we use a tie-breaking criterion so that the boundary region may be non-empty.

The value of α should be in the range $[0.5, 1]$, in addition to constraint $\alpha \geq \beta$, as suggested by many authors [20,38,40,47]. The following theorem gives the condition for $\alpha \geq 0.5$.

Theorem 10. *If a loss function with* $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ *and* $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ *satisfies the condition:*

$$\lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \tag{37}$$

then $\alpha \geq 0.5$.

Condition (37) says that the difference between the cost of classifying an element not in A into positive region and the cost of classifying the element into the boundary region is more than the difference between the cost of classifying an element in A into the boundary region and a correct classification. It forms part of condition (36). However, they do not imply each other. By combining results from Theorems 9 and 10, we have the condition for $\alpha \geq 0.5$ and $\alpha \geq \beta$.

Corollary 1. *If a loss function with* $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ *and* $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ *satisfies the conditions,*

$$\begin{aligned} &\lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \\ &(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \end{aligned} \tag{38}$$

then $\alpha \geq 0.5$ *and* $\alpha \geq \beta$.

If dual approximation operators are required, one needs to impose additional conditions on a loss function [39].

Theorem 11. *If a loss function with* $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ *and* $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ *satisfies the condition:*

$$(\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}), \tag{39}$$

then $\beta = 1 - \alpha$.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

Condition (39) does not guarantee that $\alpha \geq \beta = 1 - \alpha$, or equivalently $\alpha \geq 0.5$. The condition for $\alpha = 1 - \beta \geq 0.5$ can be obtained by combining conditions (36) and (39), or combining conditions (37) and (39).

Corollary 2. *If a loss function with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ satisfies the two sets of equivalent conditions,*

$$\text{(i).} \quad (\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \quad (\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}); \tag{40}$$

$$\text{(ii).} \quad \lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \quad (\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}); \tag{41}$$

then $\alpha = 1 - \beta \geq 0.5$.

Based on the results obtained so far, we can now investigate loss functions producing existing rough set approximation operators.

Consider the loss function:

$$\lambda_{12} = \lambda_{21} = 1, \quad \lambda_{11} = \lambda_{22} = \lambda_{31} = \lambda_{32} = 0. \tag{42}$$

There is a unit cost if an object belonging to A is classified into the negative region or if an object not belonging to A is classified into the positive region; otherwise there is no cost. This loss function satisfies the conditions given in Corollary 2. A pair of dual approximation operators can be obtained. From equation (35), we have $\alpha = 1 > \beta = 0$, $\alpha = 1 - \beta$, and $\gamma = 0.5$. According to decision rules (P1)-(B1), we obtain the standard rough set approximations [11,12].

Consider another loss function:

$$\lambda_{12} = \lambda_{21} = 1, \quad \lambda_{31} = \lambda_{32} = 0.5, \quad \lambda_{11} = \lambda_{22} = 0. \tag{43}$$

That is, a unit cost is incurred if the system classifies an object belonging to A into the negative region or an object not belonging to A is classified into the positive region; half of a unit cost is incurred if any object is classified into the boundary region. For other cases, there is no cost. The loss function satisfies the conditions given in Corollary 2. In fact, the loss function makes all $\geq$ relations in these conditions become $=$. By substituting these λ_{ij}'s into equation (35), we obtain $\alpha = \beta = \gamma = 0.5$. By using decision rules (P2)-(B2), we obtained the probabilistic rough set approximation proposed by Pawlak *et al.* [14].

The loss function,

$$\lambda_{12} = \lambda_{21} = 4, \quad \lambda_{31} = \lambda_{32} = 1, \quad \lambda_{11} = \lambda_{22} = 0, \tag{44}$$

states that there is no cost for a correct classification, 4 units of cost for an incorrect classification, and 1 unit cost for classifying an object into boundary region. It also satisfies the conditions in Corollary 2. From equation (35), we

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

have $\alpha = 0.75$, $\beta = 0.25$ and $\gamma = 0.5$. By decision rules (P1)-(B1), we have a pair of dual approximation operators $\underline{apr}_{0.75}$ and $\overline{apr}_{0.25}$.

In general, the relationships between the loss function λ and the pair of parameters (α, β) are summarized as follows.

Theorem 12. *For a loss function with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, we have,*

(a1). *α is monotonic non-decreasing with respect to λ_{12} and monotonic non-increasing with respect to λ_{32}.*
(a2). *If $\lambda_{11} < \lambda_{31}$, α is strictly monotonic increasing with respect to λ_{12} and strictly monotonic decreasing with respect to λ_{32}.*
(a3). *α is strictly monotonic decreasing with respect to λ_{31} and strictly monotonic increasing with respect to λ_{11}.*
(b1). *β is monotonic non-increasing with respect to λ_{21} and monotonic non-decreasing with respect to λ_{31}.*
(b2). *If $\lambda_{22} < \lambda_{32}$, β is strictly monotonic decreasing respect to λ_{21} and strictly monotonic increasing with respect to λ_{31}.*
(b3). *β is strictly monotonic increasing with respect to λ_{32} and strictly monotonic decreasing with respect to λ_{22}.*

The connection between threshold values of parameterized rough set approximations and the loss function has significant implications in applying the decision-theoretic model of rough sets. For example, if we increase the cost of an incorrect classification λ_{12} and keep other costs unchanged, the value α would not be decreased. Unlike the variable precision rough set model, the decision-theoretic model requires a loss function. Parameters α and β are determined from the loss function. One may argue that the loss function may be considered as a set of parameters. However, in contrast to standard threshold values, they have an intuitive interpretation. The connections given in the theorem show the consequences of a loss function and provide an interpretation of the required parameters in terms of a more realistic concept of loss or cost. One can easily interpret and measure loss or cost in a real application.

5 Conclusion

Successful applications of the theory of rough sets depend on a clear understanding of the various concepts involved. For this purpose, a decision-theoretic model of rough sets is studied in this paper by focusing on two related fundamental issues, namely, granulation of a universe and approximation in a granulated universe. The decision-theoretic model not only provides a sound basis for rough set theory, but also provides a unified framework in which many existing models of rough set can be derived. The decision model can also be interpreted in terms of a more familiar and interpretable concept known as loss or cost.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

Two granulation structures are examined. An equivalence relation induces a partition of the universe, and a reflexive relation induces a covering of the universe. Under a granulated view of the universe, a subset of universe can be fuzzified and approximated. Fuzzification of a set leads to a rough membership function, which is a special type of fuzzy membership functions. Approximations of a set can be defined in two ways. The element oriented formulation is based on rough membership function, and is related to the notion of α-cut in fuzzy sets. The granule oriented formulation is based on the set inclusion relation, and in general based on a graded set inclusion relation related to rough membership function. The two formulations produce the same results when the universe is granulated by an equivalence relation, and produce different results when the universe is granulated by a reflexive relation.

The families of fixed points of lower and upper approximation operators are studied, which provides insights into our understanding of granulation structures and the induced approximation structures. With a partition defined by an equivalence relation, the families of fixed points are related to a sub-Boolean algebra of the power set of the universe. With a covering defined by a reflexive relation, the families of fixed points are related to closure systems.

The conditions on a loss function are investigated. In particular, we explicitly state the connections between the parameters required for defining graded approximation operators and losses for various classification decisions. This provide an interpretation for parameters used in other models of rough sets. We also identify conditions on a loss function so that other rough set approximation operators, such as the standard approximation operators, probabilistic approximation operators, and variable precision approximation operators can be obtained. The decision-theoretic model is therefore more general than other models.

Acknowledgment

The author wish to thank Sai Ying and an anonymous referee for critical comments and suggestions on the paper.

References

1. Bryniarski, E. and Wybraniec-Skardowska, U. Generalized rough sets in contextual space, in: *Rough Sets and Data Mining*, Lin, T.Y. and Cercone, N. (Eds.), Kluwer Academic Publishers, Boston, 339-354, 1997.
2. Dubois, D. and Prade, H. Similarity-based approximate reasoning, in: *Computational Intelligence: Imitating Life*, J.M. Zurada, R.J. Marks II, C.J. Robinson (Eds.), IEEE Press, New York, 69-80, 1994.
3. Duda, R.O. and Hart, P.E. *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

4. Hirano, S., Inuiguchi, M. and Tsumoto, S. (Eds.), *Proceedings of International Workshop on Rough Set Theory and Granular Computing*, a special issue of the *Bulletin of International Rough Set Society*, Vol. 5, No. 1-2, 2001.
5. Komorowski, J., Pawlak, Z., Polkowski, L. and Skowron, A. Rough sets: a tutorial, in: *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, Springer-Verlag, Singapore, 3-98, 1998.
6. Klir, G.J. and Yuan, B. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, New Jersey, 1995.
7. Lin, T.Y., 1998, Granular computing on binary relations I: data mining and neighborhood systems, II: rough set representations and belief functions, in: *Rough Sets in Knowledge Discovery 1*, Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, 107-140, 1998.
8. Lin. T.Y., Yao, Y.Y. and Zadeh, L.A. (Eds.) *Rough Sets, Granular Computing and Data Mining*, Physica-Verlag, Heidelberg, 2001.
9. Orlowska, E. Logic of indiscernibility relations, *Bulletin of the Polish Academy of Sciences, Mathematics*, **33**, 475-485, 1985.
10. Orlowska, E. and Pawlak, Z. Measurement and indiscernibility, *Bulletin of Polish Academy of Sciences. Mathematics*, **32**, 617-624, 1984.
11. Pawlak, Z. Rough sets, *International Journal of Computer and Information Sciences*, **11**, 341-356, 1982.
12. Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
13. Pawlak, Z. and Skowron, A. Rough membership functions, in: *Advances in the Dempster-Shafer Theory of Evidence*, R.R. Yager and M. Fedrizzi and J. Kacprzyk (Eds.), John Wiley and Sons, New York, 251-271, 1994.
14. Pawlak, Z., Wong, S.K.M. and Ziarko, W. Rough sets: probabilistic versus deterministic approach, *International Journal of Man-Machine Studies*, **29**, 81-95, 1988.
15. Pedrycz, W. *Granular Computing: An Emerging Paradigm*, Springer-Verlag, Berlin, 2001.
16. Polkowski, L. and Skowron, A. Rough mereology: a new paradigm for approximate reasoning, *International Journal of Approximate Reasoning*, **15**, 333-365, 1996.
17. Polkowski, L. and Skowron, A. Towards adaptive calculus of granules, *Proceedings of 1998 IEEE International Conference on Fuzzy Systems*, 111-116, 1998.
18. Skowron, A. Toward intelligent systems: calculi of information granules, *Bulletin of International Rough Set Society*, **5**, 9-30, 2001.
19. Skowron, A and Polkowski, L. Rough mereology and analytical morphology, in: *Incomplete Information: Rough Set Analysis*, Orlowska, O. (Ed.), Physica-Verlag, Heidelberg, 399-437, 1998.
20. Skowron, A. and Stepaniuk, J. Tolerance approximation spaces, *Fundamenta Informaticae*, **27**, 245-253, 1996.
21. Skowron, A. and Stepaniuk, J. Information granules and approximation spaces, *Proceedings of Seventh International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 354-361, 1998.
22. Skowron, A. and Stepaniuk, J. Information granules: towards foundations of granular computing, *International Journal of Intelligent Systems*, **16**, 57-85, 2001.

23. Slowinski, R. and Vanderpooten, D. Similarity relation as a basis for rough approximations, in: *Advances in Machine Intelligence & Soft-Computing IV*, Wang, P.P. (Ed.), Department of Electrical Engineering, Duke University, Durham, North Carolina, 17-33, 1997.
24. Slowinski, R. and Vanderpooten, D. A generalized definition of rough approximations based on similarity, *IEEE Transactions on Data and Knowledge Engineering*, —bf 12, 331-336, 2000.
25. Wolpert, D.H. Off-training set error and a priori distinctions between algorithms, Technical Report SFI-TR-95-01-003, The Santa Fe Institute, 1995, http://acoma.santafe.edu/sfi/publications/Working-Papers/95-01-003.ps.
26. Wolpert, D.H. and Macready, W.G. No free lunch theorems for search, Technical Report SFI-TR-95-02-010, The Santa Fe Institute, 1996, http://acoma.santafe.edu/sfi/publications/Working-Papers/95-02-010.ps.
27. Wong, S.K.M. and Ziarko, W. Comparison of the probabilistic approximate classification and the fuzzy set model, *Fuzzy Sets and Systems*, **21**, 357-362, 1987.
28. Yager, R.R. and Filev,D. Operations for granular computing: mixing words with numbers, *Proceedings of 1998 IEEE International Conference on Fuzzy Systems*, 123-128, 1998.
29. Yao, Y.Y. Two views of the theory of rough sets in finite universes, *International Journal of Approximation Reasoning*, **15**, 291-317, 1996.
30. Yao, Y.Y. Relational interpretations of neighborhood operators and rough set approximation operators, *Information Sciences*, **111**, 239-259, 1998.
31. Yao, Y.Y. Generalized rough set models, in: *Rough Sets in Knowledge Discovery*, Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, 286-318, 1998.
32. Yao, Y.Y. A comparative study of fuzzy sets and rough sets, *Information Sciences*, **109**, 227-242, 1998.
33. Yao, Y.Y. On generalizing Pawlak approximation operators, *Rough Sets and Current Trends in Computing, Proceedings of the First International Conference*, Springer, Berlin, 298-307, 1998.
34. Yao, Y.Y. Granular computing: basic issues and possible solutions, *Proceedings of the 5th Joint Conference on Information Sciences*, **I**, 186-189, 2000.
35. Yao, Y.Y. Rough sets and interval fuzzy sets, *Proceedings of 20th International Meeting of the North American Fuzzy Information Processing Society*, to appear, 2001.
36. Yao, Y.Y. Information granulation and rough set approximation, *International Journal of Intelligent Systems*, **16**, 87-104, 2001.
37. Yao, Y.Y. and Lin, T.Y. Generalization of rough sets using modal logic, *Intelligent Automation and Soft Computing, an International Journal*, **2**, 103-120, 1996.
38. Yao, Y.Y. and Wong, S.K.M. A decision theoretic framework for approximating concepts, *International Journal of Man-machine Studies*, **37**, 793-809, 1992.
39. Yao, Y.Y., Wong, S.K.M. and Lin, T.Y. A review of rough set models, in: *Rough Sets and Data Mining: Analysis for Imprecise Data*, Lin, T.Y. and Cercone, N. (Eds.), Kluwer Academic Publishers, Boston, 47-75, 1997.
40. Yao, Y.Y., Wong, S.K.M. and Lingras, P. A decision-theoretic rough set model, in: *Methodologies for Intelligent Systems*, **5**, Z.W. Ras, M. Zemankova and M.L. Emrich (Eds.), New York: North-Holland, 17-24, 1990.

Yao, Y.Y., Information granulation and approximation in a decision-theoretic model of rough sets,
in: Rough-neuro Computing: a Way to Computing with Words,
Polkowski, L., Pal, S.K., and Skowron, A. (Eds), Springer, Berlin, pp. 491-518, 2003.

41. Yao, Y.Y. and Zhang, J.P. Interpreting fuzzy membership functions in the theory of rough sets, in: *Rough Sets and Current Trends in Computing, Second International Conference, RSCTC 2000, Lecture Notes in Computer Science 2005*, W. Ziarko and Y.Y. Yao (Eds.), Springer, Berlin, 82-89, 2000.
42. Yao, Y.Y. and Zhong, N. Granular computing using information tables, in: *Data Mining, Rough Sets and Granular Computing*, T.Y. Lin, Y.Y. Yao and L.A. Zadeh (Eds.), Physica-Verlag, Heidelberg, 2001.
43. Yao, Y.Y. and Zhong, N. Potential applications of granular computing in knowledge discovery and data mining, *Proceedings of World Multiconference on Systemics, Cybernetics and Informatics*, **5**, 573-580, 1999.
44. Zadeh, L.A. Fuzzy sets and information granularity, in: *Advances in Fuzzy Set Theory and Applications*, Gupta, N., Ragade, R. and Yager, R. (Eds.), North-Holland, Amsterdam, 3-18, 1979.
45. Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **19**, 111-127, 1997.
46. Zhong, N., Skowron, A. and Ohsuga S. (Eds.), *New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, Springer-Verlag, Berlin, 1999.
47. Ziarko, W. Variable precision rough set model, *Journal of Computer and System Sciences*, **46**, 39-59, 1993.