# Probabilistic rough set approximations

## Yiyu Yao

*Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2*

## Abstract

Probabilistic approaches have been applied to the theory of rough set in several forms, including decision-theoretic analysis, variable precision analysis, and information-theoretic analysis. Based on rough membership functions and rough inclusion functions, we revisit probabilistic rough set approximation operators and present a critical review of existing studies. Intuitively, they are defined based on a pair of thresholds representing the desired levels of precision. Formally, the Bayesian decision-theoretic analysis is adopted to provide a systematic method for determining the precision parameters by using more familiar notions of costs and risks. Results from existing studies are reviewed, synthesized and critically analyzed, and new results on the decision-theoretic rough set model are reported.
© 2007 Elsevier Inc. All rights reserved.

## 1. Introduction

In the standard rough set model proposed by Pawlak [26,27], the lower and upper approximations are defined based on the two extreme cases regarding the relationships between an equivalence class and a set. The lower approximation requires that the equivalence class is a subset of the set. For the upper approximation, the equivalence class must have a non-empty overlap with the set. A lack of consideration for the degree of their overlap unnecessarily limits the applications of rough sets and has motivated many researchers to investigate probabilistic generalizations of the theory [11,14,25,28–30,32,33,42–46,52,54,56–58,60,66,69].

Probabilistic approaches to rough sets have appeared in many forms, such as the decision-theoretic rough set model [52,54,56–58], the variable precision rough set model [14,66,69], the Bayesian rough set model [11,34,35], information-theoretic analysis [3,41], probabilistic rule induction [6,8,13,19–21,31,37–39,64,65,67,68], and many related studies [44,45]. The extensive results increase our understanding of the theory. At the same time, it seems necessary to provide a unified and comprehensive framework so that those results can be put together into an integrated whole, rather than separated studies [54]. Most of the papers in this special issue aim at such a goal. The current paper focuses specifically on the issues of probabilistic approximations. The existing results are revisited and critically reviewed and new results are provided.

---

*E-mail address:* yyao@cs.uregina.ca

Probabilistic rough set approximations can be formulated based on the notions of rough membership functions [28] and rough inclusion [30]. Both notions can be interpreted in terms of conditional probabilities or *a posteriori* probabilities. Threshold values, known as parameters, are applied to a rough membership function or a rough inclusion to obtain probabilistic or parameterized approximations. Three probabilistic models have been proposed and studied intensively. They are the decision-theoretic rough set model [52,54,56,58], the variable precision rough set model [14,66], and the Bayesian rough set model [11,34,35]. The main differences among those models are their different, but equivalent, formulations of probabilistic approximations and interpretations of the required parameters.

The variable precision rough set model treats the required parameters as a primitive notion. The interpretation and the process of determining the parameters are based on rather intuitive arguments and left to empirical studies. There is a lack of theoretical and systematic studies and justifications on the choices of the threshold parameters. In fact, a solution to this problem was reported earlier in a decision-theoretic framework for probabilistic rough set approximations [52,56,58], based on the well-established Bayesian decision procedure for classification [7]. Within the decision-theoretic framework, the required threshold values can be easily interpreted and calculated based on more concrete notions, such as costs and risks. Unfortunately, many researchers are still unaware of the decision-theoretic model and tend to estimate the parameters based on tedious trial-and-error approaches. By explicitly showing the connections of the two models in this paper, we hope to increase further understanding of the theoretical foundations of probabilistic approximations.

The Bayesian rough set model [11,34,35] attempts to provide an alternative interpretation of the required parameters. The model is based on the Bayes' rule that expresses the change from the *a priori* probability to the *a posteriori* probability, and a connection between classification and hypothesis verification. Under specific interpretations, the required parameters can be expressed in terms of various probabilities. It is not difficult to establish connections between the probabilities used in the Bayesian rough set model and the costs used in the decision-theoretic model. Additional parameters are introduced in the Bayesian rough set model. There remains the question of how to interpret and determine the required parameters systematically.

With the objective of bringing together existing studies on probabilistic rough set approximations in a unified and comprehensive framework, the rest of the paper is organized into four parts. In Section 2, we review the basic concepts of the standard rough set approximations. This establishes a basis and guidelines for various probabilistic generalizations. In Section 3, we examine the two fundamental notions of rough membership functions and rough inclusions. They serve as a foundation on which probabilistic rough set approximations can be developed. In Section 4, we critically review different formulations of probabilistic rough set approximations. In Section 5, we explicitly show the conditions on a loss function so that many specific classes of probabilistic rough set approximations introduced in Section 4 can be derived in the decision-theoretic model.

## 2. Standard rough set approximations

Suppose $U$ is a finite and non-empty set called the universe. Let $E \subseteq U \times U$ be an equivalence relation on $U$, i.e., $E$ is reflexive, symmetric, and transitive. The basic building blocks of rough set theory are the equivalence classes of $E$. For an element $x \in U$, the equivalence class containing $x$ is given by

$$[x]_E = \{y \in U | xEy\}. \tag{1}$$

When no confusion arises, we also simply write $[x]$. The family of all equivalence classes is also known as the quotient set of $U$, and is denoted by $U/E = \{[x] | x \in U\}$. It defines a partition of the universe, namely, a family of pairwise disjoint subsets whose union is the universe.

For an equivalence relation, the pair $apr = (U, E)$ is called an approximation space [26,27]. In the approximation space, we only have a coarsened view of the universe. Each equivalence class is considered as a whole granule instead of many individuals [51]. Equivalence classes are the elementary definable, measurable, or observable sets in the approximation space [26,27,50]. By taking unions of elementary definable sets, one can derive larger definable sets. The family of all definable sets contains the empty set $\emptyset$, the whole set $U$, and is closed with respect to set complement, intersection, and union. It is an $\sigma$-algebra over $U$. Furthermore, $\sigma(U/E)$ defines uniquely a topological space $(U, \sigma(U/E))$, in which $\sigma(U/E)$ is the family of all open and closed sets [26].

In general, the sigma-algebra $\sigma(U/E)$ is only a subset of the power set $2^U$. An interesting issue is therefore the representation of undefinable sets in $2^U - \sigma(U/E)$ in terms of definable sets, in order to infer knowledge about undefinable sets. Similar to the interior and closure operators in topological spaces, one can define rough set approximation operators [26,27]. For a subset $A \subseteq U$, its lower approximation is the greatest definable set contained in $A$, and its upper approximation is the least definable set containing $A$. That is, for $A \subseteq U$,

$$
\begin{aligned}
\underline{apr}(A) &= \bigcup \{X | X \in \sigma(U/E), X \subseteq A\}, \\
\overline{apr}(A) &= \bigcap \{X | X \in \sigma(U/E), A \subseteq X\}.
\end{aligned}
\tag{2}
$$

In the study of rough set theory, one often uses the following equivalent definitions [49]:

$$
\begin{aligned}
\underline{apr}(A) &= \{x \in U | \forall y \in U [xEy \Rightarrow y \in A]\}, \\
\overline{apr}(A) &= \{x \in U | \exists y \in U [xEy, y \in A]\}
\end{aligned}
\tag{3}
$$

and

$$
\begin{aligned}
\underline{apr}(A) &= \{x \in U | [x] \subseteq A\}, \\
\overline{apr}(A) &= \{x \in U | [x] \cap A \neq \emptyset\}.
\end{aligned}
\tag{4}
$$

An element is in the lower approximation of $A$ if all of its equivalent elements are in $A$, and an element is in the upper approximation of $A$ if at least one of its equivalent elements is in $A$.

Definition given by Eq. (2) is referred to as the subsystem based definition. Definitions given by Eqs. (3) and (4) are referred to as the element based definitions [53]. Equivalently, from Eq. (4), one can also have a granule based definition:

$$
\begin{aligned}
\underline{apr}(A) &= \bigcup \{[x] \in U/E | [x] \subseteq A\}, \\
\overline{apr}(A) &= \bigcup \{[x] \in U/E | [x] \cap A \neq \emptyset\}.
\end{aligned}
\tag{5}
$$

It provides a new interpretation of rough set approximations. The lower approximation is the union of equivalence classes that are subsets of $A$ and the upper approximation is the union of equivalence classes that have a non-empty intersection with $A$.

Let $A^c$ denote the complement of the set $A$. Some of the useful properties satisfied by the pair of approximation operators are summarized below [26,27,49]: for $A, B \subseteq U$,

(L0) $\quad \underline{apr}(A) = (\overline{apr}(A^c))^c$,

(U0) $\quad \overline{apr}(A) = (\underline{apr}(A^c))^c$;

(L1) $\quad \underline{apr}(A \cap B) = \underline{apr}(A) \cap \underline{apr}(B)$,

(U1) $\quad \overline{apr}(A \cup B) = \overline{apr}(A) \cup \overline{apr}(B)$;

(L2) $\quad \underline{apr}(A \cup B) \supseteq \underline{apr}(A) \cup \underline{apr}(B)$,

(U2) $\quad \overline{apr}(A \cap B) \subseteq \overline{apr}(A) \cap \overline{apr}(B)$;

(L3) $\quad A \subseteq B \Rightarrow \underline{apr}(A) \subseteq \underline{apr}(B)$,

(U3) $\quad A \subseteq B \Rightarrow \overline{apr}(A) \subseteq \overline{apr}(B)$;

(L4) $\quad \underline{apr}(A) \subseteq A$,

(U4) $\quad A \subseteq \overline{apr}(A)$;

(L5) $\quad \underline{apr}(A) = \overline{apr}(\underline{apr}(A))$,

(U5) $\quad \overline{apr}(A) = \underline{apr}(\overline{apr}(A))$;

(L6) $\quad \underline{apr}(A) = \underline{apr}(\underline{apr}(A))$,

(U6) $\quad \overline{apr}(A) = \overline{apr}(\overline{apr}(A))$;

(L7) $\quad \underline{apr}(A) = A \iff A \in \sigma(U/E)$,

(U7) $\quad \overline{apr}(A) = A \iff A \in \sigma(U/E)$.

Properties (L0) and (U0) state that the lower and upper approximations are a pair of dual operators. Properties (L1) and (U1) show the distributivity of $\underline{apr}$ over set intersection, and $\overline{apr}$ over set union. Properties (L2) and (U2) state that the lower approximation operator is not necessarily distributive over set union, and the upper approximation operator is not necessarily distributive over set intersection. According to properties (L3) and (U3), both operators are monotonic with respect to set inclusion. By properties (L4) and (U4), a set lies within its lower and upper approximations. The next two pairs of properties state that the result of applying a consecutive approximation operators is the same as the result of the operator closest to $A$. Properties (L7) and (U7) state that a set and its approximations are the same if and only if the set is a definable set in $\sigma(U/E)$.

Given a subset $A \subseteq U$, the universe can be divided into three disjoint regions, namely, the positive, the negative, and the boundary regions [26]:

$$
\begin{aligned}
\mathrm{POS}(A) &= \underline{apr}(A), \\
\mathrm{NEG}(A) &= \mathrm{POS}(A^c) = (\overline{apr}(A))^c, \\
\mathrm{BND}(A) &= \overline{apr}(A) - \underline{apr}(A).
\end{aligned}
\tag{6}
$$

An element of the positive region $\mathrm{POS}(A)$ definitely belongs to $A$, an element of the negative region $\mathrm{NEG}(A)$ definitely does not belong to $A$, and an element of the boundary region $\mathrm{BND}(A)$ only possibly belongs to $A$.

The three regions and the approximation operators uniquely determine each other. One may therefore use any of the three pairs to represent a subset $A \subseteq U$:

$$
\begin{aligned}
(\mathrm{POS}(A), \mathrm{POS}(A) \cup \mathrm{BND}(A)) &= (\underline{apr}(A), \overline{apr}(A)), \\
(\mathrm{POS}(A), \mathrm{BND}(A)) &= (\underline{apr}(A), \overline{apr}(A) - \underline{apr}(A)), \\
(\mathrm{POS}(A), \mathrm{NEG}(A)) &= (\underline{apr}(A), (\overline{apr}(A))^c).
\end{aligned}
$$

Each of them makes explicit certain particular aspect of the approximations. The first pair is the most commonly used one, defining the lower and upper bounds within which lies the set $A$. It is related to the notions of the core and support of a fuzzy set. The second pair explicitly gives the boundary elements under the approximations. The third pair focuses on what is definitely in $A$, in contrast to what is definitely not in $A$.

## 3. Rough membership functions and rough inclusion

Since the lower and upper approximations are dual operators, it is sufficient to consider one of them. According to Eqs. (3) and (4), generalized approximation operators can be introduced by relaxing the conditions:

(LC)    $\forall y \in U[xEy \Rightarrow y \in A]$,

(SC)    $[x] \subseteq A$.

For the logic condition (LC), one can use probabilistic versions. The results from graded modal logic [10,24], variable precision logic [22], and probabilistic logic [16] may be adopted. For the set-theoretic condition (SC), we may adopt the notion of the degree of set inclusion from many studies, such as approximate reasoning [46,61] and rough mereology [30].

### 3.1. Rough membership functions

The concept of rough membership functions is based on the generalization of the strict logic condition (LC) into a probabilistic version. More specifically, the rough membership value of an element $x$, with respect to a set $A \subseteq U$, is typically defined in terms of a measure of the degree to which the logic condition (LC) is true.

The notion of a rough membership function was explicitly introduced by Pawlak and Skowron [28], although it had been used and studied earlier by Wong and Ziarko [43], Pawlak et al. [29], Yao et al. [58], Yao and Wong [56], and many authors.

Let $P:2^U \to [0, 1]$ be a probability function defined on the power set $2^U$, and $E$ an equivalence relation on $U$. The triplet $apr = (U, E, P)$ is called a probabilistic approximation space [29,43]. For a subset $A \subseteq U$, its rough membership function is given by the conditional probability as follows:

$$\mu_A(x) = P(A|[x]). \tag{7}$$

Rough membership value of an element belonging to $A$ is the probability of the element in $A$ given that the element is in $[x]$. With the probabilistic interpretation of rough membership function, we will use $\mu_A(x)$ and $P(A|[x])$ interchangeably in subsequent discussions.

For a finite universe, the rough membership function is typically computed by Pawlak and Skowron [28]:

$$\mu_A(x) = \frac{|A \cap [x]|}{|[x]|}, \tag{8}$$

where $|A|$ denotes the cardinality of the set $A$.

Rough membership functions satisfy the following properties [28,55]:

(m1)   $\mu_U(x) = 1$,

(m2)   $\mu_\emptyset(x) = 0$,

(m3)   $xEy \Rightarrow \mu_A(x) = \mu_A(y)$,

(m4)   $x \in A \Rightarrow \mu_A(x) \neq 0$,

(m5)   $x \notin A \Rightarrow \mu_A(x) \neq 1$,

(m6)   $\mu_A(x) = 1 \Longleftrightarrow [x] \subseteq A$,

(m7)   $\mu_A(x) > 0 \Longleftrightarrow [x] \cap A \neq \emptyset$,

(m8)   $A \subseteq B \Rightarrow \mu_A(x) \leqslant \mu_B(x)$,

(m9)   $\mu_{A^c}(x) = 1 - \mu_A(x)$,

(m10)   $\mu_{A \cup B}(x) = \mu_A(x) + \mu_B(x) - \mu_{A \cap B}(x)$,

(m11)   $A \cap B = \emptyset \Rightarrow \mu_{A \cup B}(x) = \mu_A(x) + \mu_B(x)$,

(m12)   $\max(0, \mu_A(x) + \mu_B(x) - 1) \leqslant \mu_{A \cap B}(x) \leqslant \min(\mu_A(x), \mu_B(x))$,

(m13)   $\max(\mu_A(x), \mu_B(x)) \leqslant \mu_{A \cup B}(x) \leqslant \min(1, \mu_A(x) + \mu_B(x))$.

Those properties easily follow from the properties of a probability function. While (m1)–(m8) show the properties of rough membership functions, (m9)–(m13) show the properties of set-theoretic operations with rough membership functions. The property (m3) is particularly interesting, which shows that elements in the same equivalence class must have the same degree of membership. In other words, equivalent elements must have the same membership value.

### 3.2. Rough inclusion

The concept of rough inclusion generalizes the set-theoretic condition (SC) in order to capture graded inclusion. The degree to which $[x]$ is included in $A$ depends on both the overlap and non-overlap parts of $[x]$ and $A$.

In the rough set theory literature, the notion of rough inclusion, introduced explicitly by Polkowski and Skowron [30], has been studied using other names, including relative degree of misclassification [66], majority inclusion relation [66], vague inclusion [33], inclusion degrees [46,60,61], and so on.

Recall that the notion of rough membership functions is a generalization of the logic condition (LC). By the equivalence of the two conditions (LC) and (SC), we can extend the notion of a rough membership function to rough inclusion [30,33]. For the maximum membership value 1, we have $[x] \subseteq A$, namely, $[x]$ is a subset of $A$. For the minimum membership value 0, we have $[x] \cap A = \emptyset$, or equivalently $[x] \subseteq A^c$, namely, $[x]$ is totally not a subset of $A$. For a value between 0 and 1, it may be interpreted as the degree to which $[x]$ is a subset of $A$. Thus, one obtains a measure of graded inclusion of two sets [30,32,33,61]:

$$v(B|A) = \frac{|A \cap B|}{|A|}. \tag{9}$$

For the case where $A = \emptyset$, we define $v(B|\emptyset) = 1$, namely, the empty set is a subset of any set. Accordingly, the degree to which the equivalence class $[x]$ is included in a set $A$ is given by

$$v([A|[x]) = \frac{|[x] \cap A|}{|[x]|}. \tag{10}$$

It follows that $v(A|[x]) = \mu_A(x)$. From the properties of a rough membership function, one can easily obtain the corresponding properties of rough inclusion. Similar to a rough membership function, the value of $v(B|A)$ can be interpreted as the conditional probability $P(B|A)$ that a randomly selected element from $A$ belongs to $B$. There is a close connection between graded inclusion and fuzzy set inclusion [33,59].

In the development of the variable precision rough set model, Ziarko [66] used an inverse measure of $v$ called the relative degree of misclassification:

$$c(B|A) = 1 - v(B|A) = 1 - \frac{|A \cap B|}{|A|}. \tag{11}$$

Bryniarski and Wybraniec-Skardowska [4] 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 of graded inclusion, a context relation with respect to a value $\alpha \in [0, 1]$ can be defined by

$$\subseteq_\alpha = \{(A, B)|v(B|A) \geqslant \alpha\}. \tag{12}$$

If one interprets $v$ as a fuzzy relation on $2^U$, the relation $\subseteq_\alpha$ may be interpreted as an $\alpha$-cut of the fuzzy relation. The use of a complete lattice, or a rank set, corresponds to the study of $L$-fuzzy sets and $L$-fuzzy relations in the theory of fuzzy sets [15].

Many proposals have been made to generalize and characterize the notion of graded inclusion. Skowron and Stepaniuk [33] suggested that graded (vague) inclusion of sets may be measured by a function, $v: 2^U \times 2^U \to [0, 1]$, with monotonicity regarding the first argument, namely, for $A$, $B$, $C \subseteq U$, $v(B|A) \leqslant (C|A)$ for any $B \subseteq C$. In this case, the function defined by Eq. (9) is an example of such a measure. Skowron and Polkowski [32] suggested new properties for rough inclusion, in addition to the monotonicity. The unit interval $[0, 1]$ can also be generalized to a complete lattice in the definition of rough inclusion [30]. 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 [30,32].

Zhang and Leung [61] and Xu et al. [46] proposed a generalized notion of inclusion degree in the context of a partially ordered set. Let $(L, \preceq)$ be a partially ordered set. A function $D: L \times L \to [0, 1]$ is called a measure of inclusion degree if it satisfies the following properties [46]: for $a, b, c \in L$,

(i)   $0 \leqslant D(b|a) \leqslant 1$,
(ii)   $a \preceq b \Rightarrow D(b \mid a) = 1$,
(iii)   $a \preceq b \preceq c \Rightarrow D(a|b) \geqslant D(a|c)$,
(iv)   $a \preceq b \Rightarrow D(a|c) \leqslant D(b|c)$.

Property (i) is the normalization condition. Property (ii) ensures that the degree of inclusion reaches the maximum value for the standard inclusion. Properties (iii) and (iv) state two types of monotonicity. When the concept of inclusion degree is applied to the partially ordered set $(2^U, \subseteq)$, we immediately obtain a rough inclusion [60].

## 4. Probabilistic rough set approximations

The standard approximation operators ignore the detailed statistical information of the overlap of an equivalence class and a set [29,43]. By exploring such information, probabilistic approximation operators can be introduced.

## 4.1. Standard approximations as the core and the support of a fuzzy set

According to properties (m6) and (m7) of rough membership functions, $\mu_A(x) = 1$ if and only if for all $y \in U$, $xEy$ implies $y \in A$, and $\mu_A(x) > 0$ if and only if there exists a $y \in U$ such that $xEy$ and $y \in A$. A rough membership function $\mu_A$ may be interpreted as a special kind of fuzzy membership function. Under this interpretation, it is possible to re-express the standard rough set approximations [28,29,43], and to establish their connection to the core and support of a fuzzy set [55], as follows:

$$
\begin{aligned}
\underline{apr}(A) &= \{x \in U | \mu_A(x) = 1\} = \text{core}(\mu_A), \\
\overline{apr}(A) &= \{x \in U | \mu_A(x) > 0\} = \text{support}(\mu_A).
\end{aligned}
\tag{13}
$$

That is, the lower and upper approximations of a set $A$ are in fact the core and support of the fuzzy set $\mu_A$, respectively.

In the theory of fuzzy sets, fuzzy set intersection and union are commonly defined in terms of a pair of t-norm and t-conorm [15]. Suppose fuzzy set complement is defined by $1 - (\cdot)$. For a pair of t-norm and t-conorm, core and support of fuzzy sets satisfy the same properties of (L0)–(L6), and (U0)–(U6), except in which the lower approximation is replaced by the core, and the upper approximation by the support, respectively [55]. If the core and support are interpreted as a qualitative representation of a fuzzy set, one may conclude that the theories of fuzzy sets and rough sets share the same qualitative properties.

## 4.2. The 0.5 probabilistic approximations

An attempt to use probabilistic information for approximations was suggested by Pawlak et al. [29]. Their model is based essentially on the majority rule. An element $x$ is put into the lower approximation of $A$ if the majority of its equivalent elements $[x]$ are in $A$. That is,

$$
\begin{aligned}
\underline{apr}_{0.5} &= \{x \in U | P(A|[x]) > 0.5\}, \\
\overline{apr}_{0.5} &= \{x \in U | P(A|[x]) \geqslant 0.5\}.
\end{aligned}
\tag{14}
$$

The lower and upper 0.5 probabilistic approximation operators are dual to each other. The boundary region consists of those elements whose conditional probabilities are exactly 0.5, which represents maximal uncertainty.

## 4.3. Probabilistic approximations as the α-cuts of a fuzzy set

The standard approximations and 0.5 probabilistic approximations use special points of the probability, namely, the two extreme points 0 and 1, and the middle point 0.5. By considering other values, Yao and Wong [56] introduced more general probabilistic approximations in the decision-theoretic model.

In the theory of fuzzy sets, α-cut and strong α-cut are important notions [15]. For $\alpha \in [0, 1]$, the α-cut and strong α-cut are defined, respectively, by

$$
\begin{aligned}
(\mu_A)_\alpha &= \{x \in U | \mu_A(x) \geqslant \alpha\}, \\
(\mu_A)_{\alpha^+} &= \{x \in U | \mu_A(x) > \alpha\}.
\end{aligned}
\tag{15}
$$

Using α-cuts, the standard rough set approximations can be expressed $\underline{apr}(A) = (\mu_A)_1$ and $\overline{apr}(A) = (\mu_A)_{0^+}$. The 0.5 probabilistic approximations can be expressed as $\underline{apr}_{0.5}(A) = (\mu_A)_{0.5^+}$ and $\overline{apr}_{0.5}(A) = (\mu_A)_{0.5}$.

For generalized probabilistic approximations, a pair of parameters $\alpha, \beta \in [0, 1]$ with $\alpha \geqslant \beta$ are used. The condition $\alpha \geqslant \beta$ ensures that the lower approximation is smaller than the upper approximation in order to be consistent with existing approximation operators. Yao and Wong [56] considered two separate cases, $0 \leqslant \beta < \alpha \leqslant 1$ and $0 \neq \beta = \alpha$. For $0 \leqslant \beta < \alpha \leqslant 1$, the standard rough approximations are extended by the definition [56]:

$$
\begin{aligned}
\underline{apr}_\alpha &= \{x \in U | P(A|[x]) \geqslant \alpha\}, \\
\overline{apr}_\beta &= \{x \in U | P(A|[x]) > \beta\}.
\end{aligned}
\tag{16}
$$

For $\alpha = \beta \neq 0$, the 0.5 probabilistic approximations are extended by the definition [56]:

$$\begin{aligned}
&\underline{apr}_\alpha = \{x \in U | P(A|[x]) > \alpha), \\
&\overline{apr}_\alpha = \{x \in U | P(A|[x]) \geqslant \alpha).
\end{aligned} \tag{17}$$

For $0 < \beta \leqslant \alpha < 1$, Wei and Zhang [42] suggested another version in which the lower approximation is defined by $>$ and the upper approximation by $\geqslant$ instead. One advantage of their definition is that we do not need to have two separated cases. However, their definition cannot produce the standard rough set approximations. As will be shown in the following section, both versions are derivable from the decision-theoretic model if different tie-breaking criteria are used.

With a pair of arbitrary $\alpha$ and $\beta$, the probabilistic approximation operators are not necessarily dual to each other. In order to obtain a pair of dual operators, we set $\beta = 1 - \alpha$. Then, the lower and upper probabilistic approximation operators are dual operators.

Although the above formulation is motivated by the notion of $\alpha$-cuts in fuzzy set theory, similar notions have in fact been considered in many fields, such as variable precision (probabilistic) logic [22], probabilistic modal logic [10], graded/fuzzy modal logic [24], and many others. The use of thresholds on probability values for making a practical decision is in fact a common method in many fields, such as pattern recognition and classification [7], machine learning [23], data mining [2], and information retrieval [40], to name just a few.

Consider the condition $\alpha > \beta$. Based on the definition of Eq. (16), the probabilistic rough set operators satisfy the following properties [56,66]: for $0 \leqslant \beta < \alpha \leqslant 1$, $\alpha' \in (0, 1]$ and $\beta' \in [0, 1)$,

(PL0)  $\underline{apr}_\alpha(A) = (\overline{apr}_{1-\alpha}(A^c))^c$,

(PU0)  $\overline{apr}_\alpha(A) = (\underline{apr}_{1-\alpha}(A^c))^c$;

(PL1)  $\underline{apr}_\alpha(A \cap B) \subseteq \underline{apr}_\alpha(A) \cap \underline{apr}_\alpha(B)$,

(PU1)  $\overline{apr}_\beta(A \cup B) \supseteq \overline{apr}_\beta(A) \cup \overline{apr}_\beta(B)$;

(PL2)  $\underline{apr}_\alpha(A \cup B) \supseteq \underline{apr}_\alpha(A) \cup \underline{apr}_\alpha(B)$,

(PU2)  $\overline{apr}_\beta(A \cap B) \subseteq \overline{apr}_\beta(A) \cap \overline{apr}_\beta(B)$;

(PL3)  $A \subseteq B \Rightarrow \underline{apr}_\alpha(A) \subseteq \underline{apr}_\alpha(B)$,

(PU3)  $A \subseteq B \Rightarrow \overline{apr}_\beta(A) \subseteq \overline{apr}_\beta(B)$;

(PLU4)  $\underline{apr}_\alpha(A) \subseteq \overline{apr}_\beta(A)$;

(PL5)  $\underline{apr}_\alpha(A) = \overline{apr}_\beta(\underline{apr}_\alpha(A))$,

(PU5)  $\overline{apr}_\beta(A) = \underline{apr}_\alpha(\overline{apr}_\beta(A))$;

(PL6)  $\underline{apr}_\alpha(A) = \underline{apr}_{\alpha'}(\underline{apr}_\alpha(A))$,

(PU6)  $\overline{apr}_\beta(A) = \overline{apr}_{\beta'}(\overline{apr}_\beta(A))$;

(PL7)  $\underline{apr}_\alpha(A) = A \Longleftrightarrow A \in \sigma(U/E)$,

(PU7)  $\overline{apr}_\beta(A) = A \Longleftrightarrow A \in \sigma(U/E)$.

They are counterparts of the properties (L0)–(L7) and (U0)–(U7) of the standard rough set approximation operators. As stated earlier, $\underline{apr}_\alpha$ and $\overline{apr}_\beta$ are defined differently for the case when $\alpha = \beta$, where similar properties can be obtained. For probabilistic approximation operators, one can have additional properties: for $\alpha, \alpha' \in (0, 1]$ and $\beta, \beta' \in [0, 1)$,

(PL8)  $\alpha \geqslant \alpha' \Rightarrow \underline{apr}_\alpha(A) \subseteq \underline{apr}_{\alpha'}(A)$,

(PU8)  $\beta \geqslant \beta' \Rightarrow \overline{apr}_\beta(A) \subseteq \overline{apr}_{\beta'}(A)$;

(PL9)  $\underline{apr}(A) = \underline{apr}_1(A)$,

(PU9)  $\overline{apr}(A) = \overline{apr}_0(A)$.

Properties (PL8) and (PU8) show that both probabilistic approximation operators are monotonic decreasing with respect to the parameters $\alpha$ and $\beta$. Properties (PL9) and (PU9) establish the connection between probabilistic approximation operators and the standard approximation operators.

## 4.4. The variable precision and parameterized rough set models

With the introduction of rough inclusion, the standard approximation space can be generalized to $apr = (U, I, v)$, where $I:U \to 2^U$ is an information function and $v$ is a measure of rough inclusion [25,33]. The mapping $[\cdot]$ that maps an element to its equivalence class is an instance of the information function $I$. By applying threshold values on a rough inclusion $v$, it is possible to derive variable precision or parameterized approximations [11,33,66] by generalizing Eq. (4) of the element based definition or Eq. (5) of the granule based definition. The variable precision rough set model is one of the well known such formulations [66].

In formulating the variable precision rough set model, Ziarko [66] used the relative degree of misclassification function $c$ and the granule based definition of approximations. To be consistent with the previous and subsequent discussions, we present a slightly different, but equivalent, formulation based on the rough inclusion,

$$v(A|[x]) = \frac{|A \cap [x]|}{|[x]|} = P(A|[x]), \tag{18}$$

and the element based definition. As mentioned in the earlier discussion, based on the rough inclusion $v$, we can define different levels of set inclusion [4,66]:

$$[x] \subseteq_\alpha A \iff v(A|[x]) \geqslant \alpha \iff P(A|[x]) \geqslant \alpha, \tag{19}$$

where $\alpha \in (0, 1]$.

When defining the lower approximation, the majority requirement of the variable precision rough set model suggests that more than 50% of elements in an equivalence class $[x]$ must be in $A$ in order for $x$ to be in the lower approximation. In other words, the set-theoretic condition (SC) must hold to a degree greater than 0.5. We need to choose the threshold value $\alpha$ in the range (0.5, 1). By generalizing Eq. (4), the $\alpha$-level lower approximation is given by: for $\alpha \in (0.5, 1]$,

$$\underline{apr}_\alpha(A) = \{x \in U | [x] \subseteq_\alpha A\} = \{x \in U | P(A|[x]) \geqslant \alpha\}. \tag{20}$$

The corresponding upper approximation is defined based on the dual of the lower approximation:

$$\overline{apr}_{1-\alpha}(A) = (\underline{apr}_\alpha(A^c))^c = \{x \in U | P(A|[x]) > 1 - \alpha\}. \tag{21}$$

The condition $0.5 < \alpha \leqslant 1$ implies $0 \leqslant 1 - \alpha < 0.5$. It follows that the lower approximation is a subset of the upper approximation. The pair of parameters $(\alpha, 1 - \alpha)$ is referred to as the symmetric bounds, as it produces a pair of dual approximation operators $(\underline{apr}_\alpha, \overline{apr}_{1-\alpha})$.

Variable precision rough sets with asymmetric bounds were examined by Katzberg and Ziarko [14]. It is only required that $0 \leqslant \beta < \alpha \leqslant 1$, where $\beta$ is used to define the upper approximation and $\alpha$ is used to define the lower approximation, as defined in Eq. (16). Although $\underline{apr}_\alpha$ and $\overline{apr}_\beta$ are not necessarily dual to each other, Yao and Wong [56] showed that the two pairs of operators, $(\underline{apr}_\alpha, \overline{apr}_\beta)$ and $(\underline{apr}_{1-\beta}, \overline{apr}_{1-\alpha})$ are complement to each other.

For the special rough inclusion $v(A|[x]) = P(A|[x])$, the probabilistic approximations from the decision-theoretic model and the variable precision model are equivalent. The main differences are their formulations. The decision-theoretic model systematically derives many types of approximation operators and provides theoretical guidelines for the estimation of required parameters, while the variable precision model relies much on intuitive arguments.

Ślęzak and Ziarko [35] and Ślęzak [34] introduced the Bayesian rough set model in an attempt to provide an alternative interpretation of the required parameters in the variable precision rough set model. By setting the parameters as the *a priori* probabilities, a pair of probabilistic approximations is defined by [35]: for $A \subseteq U$,

$$\underline{apr}_{P(A)}(A) = \{x \in U | P(A|[x]) > P(A)\},$$
$$\overline{apr}_{P(A)}(A) = \{x \in U | P(A|[x]) \geqslant P(A)\}. \tag{22}$$

They correspond to the case where $\alpha = \beta = P(A)$. Ślęzak [34] examined a more complicated version of Bayesian rough set approximations by comparing probabilities $P([x]|A)$ and $P([x]|A^c)$:

$$
\begin{aligned}
\underline{bapr}_{\delta_1}(A) &= \{x \in U | P([x]|A) \geqslant \delta_1 P([x]|A^c)\}, \\
\overline{bapr}_{\delta_2}(A) &= \{x \in U | P([x]|A) \geqslant \delta_2 P([x]|A^c)\},
\end{aligned}
\tag{23}
$$

where $\delta_1$ and $\delta_2$ are parameters. Based on the Bayes' rule, one can easily find their corresponding variable precision approximations [34]. The corresponding parameters of the variable precision approximations are expressed in terms the probability $P(A)$, $\delta_1$ and $\delta_2$. In comparison with the variable precision model, the new parameters of the Bayesian rough set model are less intuitive and their estimation becomes a challenge.

Greco et al. [11,12] observed that rough membership functions and rough inclusions, as defined by the conditional probabilities $P(A|[x])$, consider the overlap of $A$ and $[x]$ and do not explicitly consider the overlap of $A$ and $[x]^c$. By considering both overlaps, they introduced a relative rough membership function:

$$
\hat{\mu}_A(x) = \frac{|A \cap [x]|}{|[x]|} - \frac{|A \cap [x]^c|}{|[x]^c|} = P(A|[x]) - P(A|[x]^c).
\tag{24}
$$

The relative rough membership function is an instance of a class of measures known as the Bayesian confirmation measures [9]. By incorporating a confirmation measure to the existing parameterized models, they proposed two-parameterized approximations [11,12]:

$$
\begin{aligned}
\underline{apr}_{\alpha,a} &= \{x \in U | P(A|[x]) \geqslant \alpha \text{ and } bc([x], A) \geqslant a\}, \\
\overline{apr}_{\beta,b} &= \{x \in U | P(A|[x]) > \beta \text{ or } bc([x], A) > b\},
\end{aligned}
\tag{25}
$$

where $bc(\cdot)$ is a Bayesian confirmation measure, and $a$ and $b$ are parameters in the range of $bc(\cdot)$. They have shown that the variable precision and Bayesian rough set models are special cases. More details of the two-parameterized approximation models can be found in their paper in this issue [12]. The extra parameters may make the model more effective, which at the same time leads to more difficulties in estimating those parameters.

## 5. The decision-theoretic rough set model

A fundamental difficulty with the probabilistic, variable precision, and parameterized approximations introduced in the last section is the physical interpretation of the required threshold parameters, as well as systematic methods for setting the parameters. This difficulty has in fact been resolved in the decision-theoretic model of rough sets proposed earlier [52,56,58]. This section reviews and summarizes the main results of the decision-theoretic framework and its connections to other studies. It draws extensive results from two previous papers [52,56] on the one hand and re-interprets these results on the other. For clarity, we only consider the element based definition of probabilistic approximation operators. The same argument can be easily applied to other cases.

### 5.1. An overview of the Bayesian decision procedure

Bayesian decision procedure deals mainly with making decision with minimum risk or cost under probabilistic uncertainty. We present an overview by following the discussion in the textbook by Duda and Hart [7], in which more detailed information can be found.

Let $\Omega = \{w_1, \ldots, w_s\}$ be a finite set of $s$ states, and let $A = \{a_1, \ldots, a_m\}$ 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 $\mathbf{x}$. Without loss of generality, we simply 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

$$
R(a_i|\mathbf{x}) = \sum_{j=1}^{s} \lambda(a_i|w_j) P(w_j|\mathbf{x}).
\tag{26}
$$

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

$$\mathbf{R} = \sum_{\mathbf{x}} R(\tau(\mathbf{x})|\mathbf{x})P(\mathbf{x}), \tag{27}$$

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 Eq. (26), 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.

## 5.2. Probabilistic rough set approximations

In an approximation space $apr = (U, E)$, all elements in the equivalence class $[x]$ share the same description [27,43]. For a given subset $A \subseteq U$, the approximation operators partition the universe into three disjoint classes POS($A$), NEG($A$), and BND($A$). Furthermore, one decides how to assign $x$ into the three regions based on the conditional probability $P(A|[x])$. It follows that the Bayesian decision procedure can be immediately applied to solve this problem [52,56,58].

For deriving the probabilistic approximation operators, we have the following problem. The set of states is given by $\Omega = \{A, A^c\}$ 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 $A = \{a_1, a_2, a_3\}$, where $a_1$, $a_2$, and $a_3$ represent the three actions in classifying an object, namely, deciding POS($A$), deciding NEG($A$), and deciding 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 $\lambda(a_i|A^c)$ denote the loss incurred for taking the same action when the object does not belong to $A$. The rough membership values $\mu_A(x) = P(A|[x])$ and $\mu_{A^c}(x) = P(A^c|[x]) = 1 - P(A \mid [x])$ are in fact the probabilities that an object in the equivalence class $[x]$ belongs to $A$ and $A^c$, respectively. The expected loss $R(a_i|[x])$ associated with taking the individual actions can be expressed as

$$\begin{aligned}
R(a_1|[x]) &= \lambda_{11}P(A|[x]) + \lambda_{12}P(A^c|[x]), \\
R(a_2|[x]) &= \lambda_{21}P(A|[x]) + \lambda_{22}P(A^c|[x]), \\
R(a_3|[x]) &= \lambda_{31}P(A|[x]) + \lambda_{32}P(A^c|[x]),
\end{aligned} \tag{28}$$

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

(P)    If $R(a_1|[x]) \leqslant R(a_2|[x])$ and $R(a_1|[x]) \leqslant R(a_3|[x])$, decide POS($A$);

(N)    If $R(a_2|[x]) \leqslant R(a_1|[x])$ and $R(a_2|[x]) \leqslant R(a_3|[x])$, decide NEG($A$);

(B)    If $R(a_3|[x]) \leqslant R(a_1|[x])$ and $R(a_3|[x]) \leqslant R(a_2|[x])$, decide BND($A$).

Tie-breaking rules should be added so that each element is classified into only one region.

Since $P(A|[x]) + P(A^c|[x]) = 1$, the above decision rules can be simplified so that only the probabilities $P(A|[x])$ are involved. We can classify any object in the equivalence class $[x]$ based only on the probabilities $P(A|[x])$, i.e., the rough membership values, and the given loss function $\lambda_{ij}$, $i = 1, 2, 3$ and $j = 1, 2$.

Consider a special kind of loss functions with $\lambda_{11} \leqslant \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leqslant \lambda_{32} < \lambda_{12}$. 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 $P(A|[x]) \geqslant \gamma$ and $P(A|[x]) \geqslant \alpha$, decide POS($A$);

(N)   If $P(A|[x]) \leqslant \beta$ and $P(A|[x]) \leqslant \gamma$, decide NEG($A$);

(B)   If $\beta \leqslant P(A|[x]) \leqslant \alpha$, decide BND($A$);

where

$$
\alpha = \frac{\lambda_{12} - \lambda_{32}}{(\lambda_{31} - \lambda_{32}) - (\lambda_{11} - \lambda_{12})},
$$
$$
\gamma = \frac{\lambda_{12} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{11} - \lambda_{12})}, \tag{29}
$$
$$
\beta = \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}.
$$

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

If a loss function with $\lambda_{11} \leqslant \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leqslant \lambda_{32} < \lambda_{12}$ further satisfies the condition:

$$
(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geqslant (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \tag{30}
$$

then $\alpha \geqslant \gamma \geqslant \beta$. The condition ensures that probabilistic rough set approximations are consistent with the standard rough set approximations. In other words, the lower approximation is a subset of the upper approximation, and the boundary region may be non-empty.

The physical meaning of condition (30) may be interpreted as follows. 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 $\lambda_{32}$ away from $\lambda_{12}$, or move $\lambda_{31}$ away from $\lambda_{21}$. 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.

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

(P1)   If $P(A|[x]) \geqslant \alpha$, decide POS($A$);

(N1)   If $P(A|[x]) \leqslant \beta$, decide NEG($A$);

(B1)   If $\beta < P(A|[x]) < \alpha$, decide BND($A$).

Based on the relationship between approximations and the three regions, we obtain the probabilistic approximations:

$$
\underline{apr}_\alpha(A) = \{x \in U | P(A|[x]) \geqslant \alpha\},
$$
$$
\overline{apr}_\beta(A) = \{x \in U | P(A|[x]) > \beta\}.
$$

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

(P2)   If $P(A|[x]) > \alpha$, decide POS($A$);

(N2)   If $P(A|[x]) < \alpha$, decide NEG($A$);

(B2)   If $P(A|[x]) = \alpha$, 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. Probabilistic approximations can be obtained, which is similar to the 0.5 probabilistic approximations introduced by Pawlak et al. [29].

As an example to illustrate the probabilistic approximations, consider a loss function:

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

It 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 the boundary region. From Eq. (29), we 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 a loss function $\lambda$ and the pair of parameters $(\alpha, \beta)$ can be established. For a loss function with $\lambda_{11} \leqslant \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leqslant \lambda_{32} < \lambda_{12}$, we have [52]:

- $\alpha$ is monotonic non-decreasing with respect to $\lambda_{12}$ and monotonic non-increasing with respect to $\lambda_{32}$.
- If $\lambda_{11} < \lambda_{31}$, $\alpha$ is strictly monotonic increasing with respect to $\lambda_{12}$ and strictly monotonic decreasing with respect to $\lambda_{32}$.
- $\alpha$ is strictly monotonic decreasing with respect to $\lambda_{31}$ and strictly monotonic increasing with respect to $\lambda_{11}$.
- $\beta$ is monotonic non-increasing with respect to $\lambda_{21}$ and monotonic non-decreasing with respect to $\lambda_{31}$.
- If $\lambda_{22} < \lambda_{32}$, $\beta$ is strictly monotonic decreasing respect to $\lambda_{21}$ and strictly monotonic increasing with respect to $\lambda_{31}$.
- $\beta$ is strictly monotonic increasing with respect to $\lambda_{32}$ and strictly monotonic decreasing with respect to $\lambda_{22}$.

Such connections between the required parameters of probabilistic rough set approximations and loss functions have significant implications in applying the decision-theoretic model of rough sets. For example, if we increase the cost of an incorrect classification $\lambda_{12}$ and keep other costs unchanged, the value $\alpha$ would not be decreased. Parameters $\alpha$ and $\beta$ are determined from a loss function. One may argue that a loss function may be considered as a set of parameters. However, in contrast to the standard threshold values, they are not abstract notions, but have an intuitive interpretation. One can easily interpret and measure loss or cost in a real application. In fact, the results and ideas of the decision-theoretic model have been successfully applied to many fields, including data analysis and data mining [5,42,44,62], information retrieval [17,36], feature selection [48], web-based support systems [47], intelligent agents [18], and email classifications [63]. Some authors have generalized the decision-theoretic model to multiple regions [1].

### 5.3. Derivations of existing probabilistic approximations

By imposing various conditions on a loss function, we can easily derive other more specific probabilistic rough set approximations introduced by many researchers.

### 5.3.1. Probabilistic rough set approximations

In the development of decision-theoretic model, we have considered the basic condition:

(B)   $\lambda_{11} \leqslant \lambda_{31} < \lambda_{21}, \quad \lambda_{22} \leqslant \lambda_{32} < \lambda_{12}.$

From the condition (B), we cannot infer the desired relationship between $\alpha$ and $\beta$. In order to make sure that $\alpha \geqslant \beta$, we have further imposed a condition:

(G)   $(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geqslant (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}).$

A loss function satisfying both (B) and (G) guarantees that $\alpha \geqslant \beta$. We therefore obtain the variable precision rough set approximation with asymmetric bounds [14]. The lower and upper approximations, $\underline{apr}_\alpha$ and $\overline{apr}_\beta$, are not necessarily dual operators.

### 5.3.2. Majority based probabilistic rough set approximations

In addition to the constraint $\alpha \geqslant \beta$, many authors suggested that the value of $\alpha$ should be in the range [0.5, 1]. With this condition, an element is in the lower approximation if the majority of its equivalent elements are in the set [33,56,58,66]. If a loss function satisfies condition (B) and the following condition (M):

(M)   $\lambda_{12} - \lambda_{32} \geqslant \lambda_{31} - \lambda_{11},$

then $\alpha \geqslant 0.5$. Condition (M) 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.

Condition (M) is only a part of condition (G). They do not imply each other. The conditions for $\alpha \geqslant 0.5$ and $\alpha \geqslant \beta$ are (B), (G), and (M).

### 5.3.3. Dual probabilistic rough set approximations

Properties (PL0) and (PU0) suggest one can define dual probabilistic approximation operators based on a pair of parameters $(\alpha, \beta)$ with $\alpha + \beta = 1$. The dual probabilistic approximation operators were examined in the decision-theoretic rough set model [52,57] and the variable precision rough set model [66].

If dual approximation operators are required, one needs to impose additional conditions on a loss function [52,57]. If a loss function satisfies condition (B) and the following condition (D):

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

then $\beta = 1 - \alpha$.

Conditions (B) and (D) do not guarantee that $\alpha \geqslant \beta = 1 - \alpha$, or equivalently $\alpha \geqslant 0.5$ and $\alpha + \beta = 1$. The condition for $\alpha = 1 - \beta \geqslant 0.5$ can be obtained by combining conditions (G) and (D), or combining conditions (M) and (D). In other words, for $\alpha = 1 - \beta \geqslant 0.5$, we have two sets that are equivalent. One set consists of (B), (G), and (D), the other set consists of (B), (M), and (D).

### 5.3.4. Standard rough set approximations

Consider the loss function:

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

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 (B), (G), (M), and (D). A pair of dual approximation operators can be obtained. From Eq. (29), 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 [26,27].

The loss function for deriving the standard rough set approximations is intuitively appealing. There may exist more than one loss function to produce the standard rough set approximations. If a loss function satisfies (B) and the condition:

(S)   $\lambda_{11} = \lambda_{31}, \quad \lambda_{32} = \lambda_{22},$

we have $\alpha = 1$ and $\beta = 0$. The condition (S) requires that the cost of classifying an element not in $A$ into the negative region (i.e., a correct classification) is the same as classifying the element into the boundary region, and the cost of classifying an element in $A$ into the positive region (i.e., a correct classification) is the same as classifying the element into the boundary region. By condition (B), those costs should be strictly less than that of incorrect classification. That is, if a loss function satisfies conditions $\lambda_{11} = \lambda_{31} < \lambda_{21}$ and $\lambda_{22} = \lambda_{32} < \lambda_{12}$, we derive the standard rough set approximations.

### 5.3.5. The 0.5 probabilistic rough set approximations

For the derivation of 0.5 probabilistic rough set approximations [29], we need $\alpha = \beta = 0.5$. It suggests that we can consider conditions (M) and (D) together. Thus, we examine the special case where the $\geqslant$ relation in (M) becomes the equality $=$. Suppose a loss function satisfies (B) and the condition:

(P)   $\lambda_{12} - \lambda_{32} = \lambda_{31} - \lambda_{11}, \quad \lambda_{32} - \lambda_{22} = \lambda_{21} - \lambda_{31}.$

By substituting these $\lambda_{ij}$'s into Eq. (29), we obtain $\alpha = \beta = \gamma = 0.5$. From decision rules (P2)–(B2), we obtain the 0.5 probabilistic rough set approximations proposed by Pawlak et al. [29].

Consider the loss function:

$$\lambda_{12} = \lambda_{21} = 1, \quad \lambda_{31} = \lambda_{32} = 0.5, \quad \lambda_{11} = \lambda_{22} = 0. \tag{33}$$

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. This loss function has a very clear and concrete physical interpretation. It satisfies the conditions (B) and (P), which produces the required parameters $\alpha = \beta = 0.5$.

### 5.3.6. The Bayesian and two-parameterized rough set models

It has been shown that one can find the corresponding variable precision approximations for the Bayesian rough set approximations [34]. It has also been shown that both variable precision and Bayesian rough set models may be viewed as special cases of the two-parameterized model [34]. As illustrated by the previous discussion, variable precision approximations can be derived naturally in the decision-theoretic rough set model. Consequently, it is a relatively easy, although may be tedious, task to interpret the results of the Bayesian and the two-parameterized rough set models in the decision-theoretic framework.

The parameters of the Bayesian and the two-parameterized models may be mathematically expressed in terms various probabilities and loss functions. However, the mixture of probabilities and loss functions may decrease the simplicity and understandability of the decision-theoretic model. In solving many practical problems, it is extremely important to strive for the right balance between the simplicity and the power of a model. Although the introduction of extra parameters may increase the power and flexibility of a model, such a power cannot be materialized unless a simple and systematic procedure exists for estimating those parameters. Future research efforts may be put on the study of this problem.

## 6. Conclusion

Several forms of probabilistic approaches to rough sets have appeared in the last decade and new proposals were made recently. It is evident that a general framework is needed for comparing and synthesizing existing results. A revisit to probabilistic rough set approximations suggest that the Bayesian decision-theoretic framework can help us to achieve this goal.

In this paper, we critically reviewed existing studies on the probabilistic rough set approximations. Results from the decision-theoretic model, the variable precision model, the Bayesian rough set model, and the two-parameterized model are pooled together and studied based on the notions of rough membership functions and rough inclusion. Since both notions are defined by the same conditional probabilities, one can formulate probabilistic rough set approximations by using any one of them. The decision-theoretic model uses rough membership functions and the variable precision model uses rough inclusions. Although the same results are produced, the variable precision model suffers from a fundamental difficulty in the interpretation and determination of the required parameters. In contrast, the decision-theoretic model adopts loss functions as a primitive notion and derives systematically all required parameters. By providing a concrete physical interpretation of loss functions, the decision-theoretic model provides theoretical guidelines on the application of approximations. More specifically, approximations lead to loss or risk, and the decision-theoretic model ensures that such loss is minimal.

The Bayesian rough set model aims at interpreting the parameters of the variable precision model based on the Bayes factor. The two-parameterized model extends one-parameterized approximations by introducing also threshold values on a Bayesian confirmation measure. Both models bring new insights into probabilistic rough set approximations. A problem of the two models is a lack of a systematic procedure for setting the required parameters. Although it is possible to link them mathematically to loss functions, their physical meanings need to be further explored.

## References

[1] M.M.E. Abd El-Monsef, N.M. Kilany, Decision analysis via granulation based on general binary relation, International Journal of Mathematics and Mathematical Sciences 2007 (2007), doi:10.1155/2007/12714. Available at <http://www.hindawi.com/GetArticle.aspx?>.

[2] R. Agrawal, T. Imielinski, A. Swami, Mining association rules between sets of items in large databases, in: Proceedings of ACM Special Interest Group on Management of Data, 1993, pp. 207–216.

[3] T. Beauboef, F.E. Petry, G. Arora, Information-theoretic measures of uncertainty for rough sets and rough relational databases, Information Sciences 109 (1998) 185–195.

[4] E. Bryniarski, U. Wybraniec-Skardowska, Generalized rough sets in contextual space, in: T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining, Kluwer Academic Publishers, Boston, 1997, pp. 339–354.

[5] J.S. Deogun, V.V. Raghavan, A. Sarkar, H. Sever, Data mining: trends in research and development, in: T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining, Kluwer Academic Publishers, Boston, 1997, pp. 9–45.

[6] J.S. Deogun, V.V. Raghavan, H. Sever, Rough set based classification methods and extended decision tables, in: Proceedings of the International Workshop on Rough Sets and Soft Computing, RSSC'04, 1994, pp. 302–309.

[7] R.O. Duda, P.E. Hart, Pattern Classification and Scene Analysis, Wiley, New York, 1973.

[8] I. Düntsch, G. Gediga, Roughian: rough information analysis, International Journal of Intelligent Systems 16 (2001) 121–147.

[9] E. Eells, B. Fitelson, Symmetries and asymmetries in evidential support, Philosophical Studies 107 (2002) 129–142.

[10] M. Fattorosi-Barnaba, G. Amati, Modal operators with probabilistic interpretations I, Studia Logica XLVI (1987) 383–393.

[11] S. Greco, B. Matarazzo, R. Sowiński, Rough membership and Bayesian confirmation measures for parameterized rough sets, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of RSFDGrC'05, LNAI, vol. 3641, 2005, pp. 314–324.

[12] S. Greco, B. Matarazzo, R. Słowiński, Parameterized rough set model using rough membership and Bayesian confirmation measures, International Journal of Approximate Reasoning 49 (2008) 285–300.

[13] J.W. Grzymala-Busse, LERS – a system for learning from example based on rough sets, in: R. Słowiński (Ed.), Intelligent Support: Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht, 1992, pp. 3–18.

[14] J.D. Katzberg, W. Ziarko, Variable precision rough sets with asymmetric bounds, in: W. Ziarko (Ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery, Springer, London, 1994, pp. 167–177.

[15] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, New Jersey, 1995.

[16] W.D. Lee, S.R. Ray, Probabilistic inference for variable certainty decisions, in: Proceedings of the 14th ACM Annual Conference on Computer Science, 1986, pp. 482.

[17] Y. Li, C. Zhang, J.R. Swanb, Rough set based model in information retrieval and filtering, in: Proceedings of the 5th International Conference on Information Systems Analysis and Synthesis, 1999, pp. 398–403.

[18] Y. Li, C. Zhang, J.R. Swanb, An information filtering model on the Web and its application in JobAgent, Knowledge-Based Systems 13 (2000) 285–296.

[19] T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston, 1997.

[20] J.S. Mi, W.Z. Wu, W.X. Zhang, Approaches to knowledge reduction based on variable precision rough set model, Information Sciences 159 (2004) 255–272.

[21] D. Miao, L. Hou, A comparison of rough set methods and representative inductive learning algorithms, Fundamenta Informaticae 59 (2004) 203–219.

[22] R.S. Michalski, P.H. Winston, Variable precision logic, Artificial Intelligence 29 (1986) 121–146.

[23] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.

[24] A. Nakamura, J.M. Gao, On a KTB-modal fuzzy logic, Fuzzy Sets and Systems 45 (1992) 327–334.

[25] S.H. Nguyen, A. Skowron, J. Stepaniuk, Granular computing: a rough set approach, Computational Intelligence 17 (2001) 514–544.

[26] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341–356.

[27] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.

[28] Z. Pawlak, A. Skowron, Rough membership functions, in: R.R. Yager, M. Fedrizzi, J. Kacprzyk (Eds.), Advances in the Dempster–Shafer Theory of Evidence, John Wiley and Sons, New York, 1994, pp. 251–271.

[29] Z. Pawlak, S.K.M. Wong, W. Ziarko, Rough sets: probabilistic versus deterministic approach, International Journal of Man–Machine Studies 29 (1988) 81–95.

[30] L. Polkowski, A. Skowron, Rough mereology: a new paradigm for approximate reasoning, International Journal of Approximate Reasoning 15 (1996) 333–365.

[31] L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery 1, vols. 1–2, Physica-Verlag, Heidelberg, 1998.

[32] A. Skowron, L. Polkowski, Rough mereology and analytical morphology, in: E. Orlowska (Ed.), Incomplete Information: Rough Set Analysis, Physica-Verlag, Heidelberg, 1998, pp. 399–437.

[33] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae 27 (1996) 245–253.

[34] D. Śle¸zak, Rough sets, Bayes factor, LNCS Transactions on Rough Sets III, LNCS 3400 (2005) 202–229.

[35] D. Śle¸zak, W. Ziarko, The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning 40 (2005) 81–91.

[36] P. Srinivasan, M.E. Ruiz, D.H. Kraft, J. Chen, Vocabulary mining for information retrieval: rough sets and fuzzy sets, Information Processing and Management 37 (2001) 15–38.

[37] J. Stefanowski, On rough set based approaches to induction of decision rules, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery 1, Physica-Verlag, Heidelberg, 1998.

[38] S. Tsumoto, Automated extraction of medical expert system rules from clinical databases on rough set theory, Information Sciences 112 (1998) 67–84.

[39] S. Tsumoto, Accuracy and coverage in rough set rule induction, in: Rough Sets and Current Trends in Computing, Proceedings of RSCTC'02, LNAI, vol. 2475, 2002, pp. 373–380.

[40] C.J. van Rijsbergen, Information Retrieval, Butterworths, London, 1979.

[41] G.Y. Wang, Rough reduction in algebra view and information view, International Journal of Intelligent Systems 18 (2003) 679–688.

[42] L.L. Wei, W.X. Zhang, Probabilistic rough sets characterized by fuzzy sets, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 12 (2004) 47–60.

[43] S.K.M. Wong, W. Ziarko, Comparison of the probabilistic approximate classification and the fuzzy set model, Fuzzy Sets and Systems 21 (1987) 357–362.

[44] W.Z. Wu, Upper and lower probabilities of fuzzy events induced by a fuzzy set-valued mapping, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of RSFDGrC'05, LNAI, vol. 3641, 2005, pp. 345–353.

[45] W.Z. Wu, Y. Leung, W.X. Zhang, On generalized rough fuzzy approximation operators, LNCS Transactions on Rough Sets V, LNCS 4100 (2006) 263–284.

[46] Z.B. Xu, J.Y. Liang, C.Y. Dang, K.S. Chin, Inclusion degree: a perspective on measures for rough set data analysis, Information Sciences 141 (2002) 227–236.

[47] J.T. Yao, J.P. Herbert, Web-based support systems based on rough set analysis, in: Rough Sets and Emerging Intelligent Systems Paradigms, Proceedings of RSEISP'07, LNAI, vol. 4585, 2007, pp. 360–370.

[48] J.T. Yao, M. Zhang, Feature selection with adjustable criteria, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of RSFDGrC'05, LNAI, vol. 3641, 2005, pp. 204–213.

[49] Y.Y. Yao, Two views of the theory of rough sets in finite universes, International Journal of Approximation Reasoning 15 (1996) 291–317.

[50] Y.Y. Yao, On generalizing Pawlak approximation operators, in: Rough Sets and Current Trends in Computing, Proceedings of RSCTC'98, LNAI, vol. 1424, 1998, pp. 298–307.

[51] Y.Y. Yao, Information granulation and rough set approximation, International Journal of Intelligent Systems 16 (2001) 87–104.

[52] Y.Y. Yao, Information granulation and approximation in a decision-theoretical model of rough sets, in: S.K. Pal, L. Polkowski, A. Skowron (Eds.), Rough-neural Computing: Techniques for Computing with Words, Springer, Berlin, 2003, pp. 491–518.

[53] Y.Y. Yao, On generalizing rough set theory, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of RSFDGrC'03, LNAI, vol. 2639, 2003, pp. 44–51.

[54] Y.Y. Yao, Probabilistic approaches to rough sets, Expert Systems 20 (2003) 287–297.

[55] Y.Y. Yao, Semantics of fuzzy sets in rough set theory, LNCS Transactions on Rough Sets II, LNCS 3135 (2004) 297–318.

[56] Y.Y. Yao, S.K.M. Wong, A decision theoretic framework for approximating concepts, International Journal of Man–machine Studies 37 (1992) 793–809.

[57] Y.Y. Yao, S.K.M. Wong, T.Y. Lin, A review of rough set models, in: T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining: Analysis for Imprecise Data, Kluwer Academic Publishers, Boston, 1997, pp. 47–75.

[58] Y.Y. Yao, S.K.M. Wong, P. Lingras, A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova, M.L. Emrich (Eds.), Methodologies for Intelligent Systems 5, North-Holland, New York, 1990, pp. 17–24.

[59] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1976) 338–353.

[60] M. Zhang, L.D. Xu, W.X. Zhang, H.Z. Li, A rough set approach to knowledge reduction based on inclusion degree and evidence reasoning theory, Expert Systems 20 (2003) 298–304.

[61] W.X. Zhang, Y. Leung, Theory of including degrees and its applications to uncertainty inference, in: Soft Computing in Intelligent Systems and Information Processing, Proceedings of 1996 Asian Fuzzy System Symposium, 1996, pp. 496–501.

[62] W.X. Zhang, W.Z. Wu, J.Y. Liang, D.Y. Li, Rough Set Theory and Methodology, Xi'an Jiaotong University Press, Xi'an, China, 2001 (in Chinese).

[63] W.Q. Zhao, Y.L. Zhu, An email classification scheme based on decision-theoretic rough set theory and analysis of email security, in: Proceedings of 2005 IEEE Region 10 TENCON, doi: 10.1109/TENCON.2005.301121.

[64] N. Zhong, J.Z. Dong, S. Ohsuga, Data mining: a probabilistic rough set approach, in: L. Polkowski, A. Skowron (Eds.), Rough Sets in Knowledge Discovery 2, Physica-Verlag, Heidelberg, 1998, pp. 127–146.

[65] N. Zhong, A. Skowron, A rough set-based knowledge discovery process, International Journal of Mathematics and Computer Science 11 (2001) 603–619.

[66] W. Ziarko, Variable precision rough set model, Journal of Computer and System Science 46 (1993) 39–59.

[67] W. Ziarko, Acquisition of hierarchy-structured probabilistic decision tables and rule from data, Expert Systems 20 (2003) 305–310.

[68] W. Ziarko, Probabilistic rough sets, in: Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of RSFDGrC'05, LNAI, vol. 3641, 2005, pp. 283–293.

[69] W. Ziarko, Probabilistic approach to rough sets, International Journal of Approximate Reasoning 49 (2008) 272–284.