Fundamenta Informaticae 108 (2011) 249–265
IOS Press

Two Semantic Issues in a Probabilistic Rough Set Model

Yiyu Yao *

Department of Computer Science
University of Regina
Regina, Canada
yyao@cs.uregina.ca

Abstract. Probabilistic rough set models are quantitative generalizations of the classical and qualitative Pawlak model by considering degrees of overlap between equivalence classes and a set to be approximated. The extensive studies, however, have not sufficiently addressed some semantic issues in a probabilistic rough set model. This paper examines two fundamental semantics-related questions. One is the interpretation and determination of the required parameters, i.e., thresholds on probabilities, for defining the probabilistic lower and upper approximations. The other is the interpretation of rules derived from the probabilistic positive, boundary and negative regions. We show that the two questions can be answered within the framework of a decision-theoretic rough set model. Parameters for defining probabilistic rough sets are interpreted and determined in terms of loss functions based on the well established Bayesian decision procedure. Rules constructed from the three regions are associated with different actions and decisions, which immediately leads to the notion of three-way decision rules. A positive rule makes a decision of acceptance, a negative rule makes a decision of rejection, and a boundary rules makes a decision of deferment. The three-way decisions are, again, interpreted based on the loss functions.

Keywords: Decision-theoretic rough sets, probabilistic rough sets, semantics, three-way decisions

1. Semantic Studies on Rough Sets

The theory of rough sets was proposed more than a quarter century ago [19]. There is a vast literature on the theory and applications [23, 24, 30]. An examination of the literature suggests a lack of systematic

Address for correspondence: Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2
*This work is partially supported by a Discovery Grant from NSERC Canada. The author thanks the reviewers for their constructive comments.

studies on the semantics of some of the fundamental notions of rough sets. This has led to inconsistent interpretations of the theory, misuses of the theory, and meaningless generalizations of the theory. All these issues demand a careful examination of the semantics of the theory of rough sets. To appreciate the importance of semantic studies, let us first recall the main ideas from the following views on formulating a theory or a logical system.

Hempel [10] suggests that the formulation of a theory will require the specification of two kinds of principles, called internal principles and bridge principles. The internal principles "will characterize the basic entities and processes invoked by the theory and the laws to which they are assumed to conform" and the bridge principles "will indicate how the processes envisaged by the theory are related to empirical phenomena with which we are already acquainted, and which the theory may then explain, predict, or retrodict."

In the study of modal logics, Hughes and Cresswell [15] identify two important tasks when constructing a logical system. One is to "set up a system of symbols, with rules for combining them into formulae and manipulating these formulae in various ways;" the other is to "give an interpretation, or attach a meaning, to these symbols and formulae." The result of the first task is an *uninterpreted* system; the result of both tasks is an *interpreted* system.

Hempel's bridge principles and Hughes and Cresswell's second task are applicable to the formulation of the theory of rough sets. It is important to pay due attention to both abstract theoretical formulations and concrete physical interpretations. We must precisely articulate and define the theoretical concepts, carefully examine their physical meanings in the context of a particular situation, and be fully aware the appropriateness and limitation of the theory for solving a specific problem. This not only increases the likelihood of the success of an application of the theory but also avoids potential misuses of the theory.

In the context fuzzy set theory, Dubois and Prade [4] investigate three semantics for fuzzy membership functions. They point out that "[s]ets may be useful for many purposes and set membership does not mean the same thing at the operational level in each and every context." Their study also motivates, and may shed some light on, semantic studies of rough sets.

With the insights gained from existing studies on rough sets, it is the time for us to address semantic issues of rough set theory. As an example, this paper investigates systematically the following two semantic issues in probabilistic rough set models:

> **Meaningful construction of probabilistic rough set approximations**: the interpretation and estimation of the required threshold values on probabilities for defining probabilistic approximations.
>
> **Meaningful classification with probabilistic rough set approximations**: the interpretation and application of probabilistic rules derived from the probabilistic positive, boundary, and negative regions.

The first issue concerns the construction of probabilistic rough set approximations and the second issue deals with the utilization of the approximations. They have, in fact, been touched upon briefly in some previous papers [41, 43, 45, 46]. However, they were not discussed explicitly in the context of the semantics of rough sets. Their implications in resolving semantic difficulties and inconsistencies in the existing studies of probabilistic rough sets have not been fully explored. For this reason, this paper reviews and re-interprets some existing results regarding the two semantic issues within the framework of the decision-theoretic rough set model (DTRSM) [41, 45, 46].

This paper may be viewed as a plea for semantic studies in rough set theory. With almost thirty years' development of the theory, it may be the time to look closer at some of its fundamental issues. We need to examine the research efforts on various generalizations and a wide range of applications of rough set theory. We must make a distinction between useful generalizations and useless generalizations and between meaningful applications and meaningless applications, in a similar way that has been well addressed in other fields [1, 26].

2. Two Semantic Issues of Probabilistic Rough Sets

Two semantics issues of probabilistic rough sets are identified from existing research.

2.1. Construction of probabilistic rough set approximations

Let U be a finite and nonempty set and E an equivalence relation on U. The equivalence relation E induces a partition of U, denoted by U/E. From the partition U/E, one can construct a Boolean algebra $B(U/E)$ by collecting the empty set, equivalence classes, and unions of equivalence classes. Each set in the family $B(U/E)$ is called a definable set, as it can be precisely defined by a logic formula in a decision language in an information table [40]. The pair $apr = (U, E)$, or $apr = (U, B(U/E))$, is called an approximation space [19, 20]. Any set not in $B(U/E)$ is called a undefinable set and can be approximated by a pair of definable sets.

Suppose that the equivalence class containing x is denoted by $[x] = \{y \mid xEy\}$. For a subset $A \subseteq U$, its lower and upper approximations are defined by [19, 20]:

$$\begin{aligned}\underline{apr}(A) &= \bigcup\{X \mid X \in B(U/E), X \subseteq A\},\\ &= \{x \in U \mid [x] \subseteq A\};\\ \overline{apr}(A) &= \bigcap\{X \mid X \in B(U/E), A \subseteq X\}\\ &= \{x \in U \mid [x] \cap A \neq \emptyset\}.\end{aligned} \tag{1}$$

That is, $\underline{apr}(A)$ is the largest definable set contained in A, and $\overline{apr}(A)$ is the smallest definable set containing A. They can be interpreted as a pair of unary set-theoretic operators $\underline{apr}, \overline{apr} : 2^U \longrightarrow 2^U$, where 2^U denotes the power set of U. They are dual operators in the sense that $\underline{apr}(A) = (\overline{apr}(A^c))^c$ and $\overline{apr}(A) = (\underline{apr}(A^c))^c$, where A^c is the set complement of A. In this way, we construct a rough set algebra $(2^U, ^c, \underline{apr}, \overline{apr}, \cap, \cup)$ as a set algebra with added operators [38].

Based on the rough set approximations of A, one can divide the universe U into three pair-wise disjoint regions [19], the positive region $\mathrm{POS}(A)$, the boundary region $\mathrm{BND}(A)$, and the negative region $\mathrm{NEG}(A)$:

$$\begin{aligned}\mathrm{POS}(A) &= \underline{apr}(A),\\ \mathrm{BND}(A) &= \overline{apr}(A) - \underline{apr}(A),\\ \mathrm{NEG}(A) &= U - \mathrm{POS}(A) \cup \mathrm{BND}(A) = U - \overline{apr}(A) = (\overline{apr}(A))^c.\end{aligned} \tag{2}$$

Some of these regions may be empty. 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 an element $x \in \mathrm{BND}(A)$ belongs to A.

Alternatively, we can first define three regions as follows:

$$\begin{aligned}
\mathrm{POS}(A) &= \{x \in U \mid [x] \subseteq A\}, \\
\mathrm{BND}(A) &= \{x \in U \mid [x] \cap A \neq \emptyset \wedge [x] \not\subseteq A\}, \\
\mathrm{NEG}(A) &= \{x \in U \mid [x] \cap A = \emptyset\}. \qquad (3)
\end{aligned}$$

In terms of the three regions, the pair of lower and upper approximations are given by:

$$\begin{aligned}
\underline{apr}(A) &= \mathrm{POS}(A), \\
\overline{apr}(A) &= \mathrm{POS}(A) \cup \mathrm{BND}(A) \\
&= (\mathrm{NEG}(A))^c. \qquad (4)
\end{aligned}$$

In other words, the pair of lower and upper approximations and the three regions uniquely define each other. One can formulate the theory of rough sets by using any one of them.

The conditions, $[x] \subseteq A$ and $[x] \cap A \neq \emptyset$, in equation (1) represent qualitative relationships between $[x]$ and A, namely, $[x]$ is contained in A and $[x]$ has an overlap with A, respectively. The degree of overlap is not considered. This observation can be made more explicit based on the notion of rough membership functions [22]. For a subset $A \subseteq U$, Pawlak and Skowron [22] suggest that a rough membership function μ_A can be defined as follows:

$$\mu_A(x) = Pr(A|[x]) = \frac{|A \cap [x]|}{|[x]|}, \qquad (5)$$

where $|\cdot|$ denotes the cardinality of a set. In may be interpreted as the conditional probability that element is in A given that the element is in the equivalence class $[x]$. This simple method for estimating the conditional probability, based on the cardinalities of sets, makes an assumption of the uniform distribution, which may be unrealistic. To avoid this difficulty, Pawlak *et al.* [25] simply assume that a probability function is given. Ziarko [50] interprets the quantity as a measure of inclusion of two sets. Several authors consider other methods for estimating the probability more accurately and realistically [3, 48].

From a rough membership function, the Pawlak approximations can be equivalently defined by [34]:

$$\begin{aligned}
\underline{apr}(A) &= \{x \in U \mid Pr(A|[x]) = 1\}; \\
\overline{apr}(A) &= \{x \in U \mid Pr(A|[x]) > 0\}. \qquad (6)
\end{aligned}$$

They are defined by using the two extreme values, 0 and 1, of probabilities. They are of a qualitative nature; the magnitude of the value $Pr(A|[x])$ is not taken into account. This observation has led to many proposals for probabilistic rough sets, including 0.5-probabilistic rough sets [25], decision-theoretic rough sets [41, 45, 46], variable precision rough set [50], Bayesian rough sets [27, 29], parameterized rough sets [8], game-theoretic rough sets [12, 13], and many others [2, 32, 36]. A more detailed review and exploration of probabilistic approaches to rough sets can be found in references [39, 42, 51].

Probabilistic rough set models attempt to generalize the restrictive definition of the lower and upper approximations by allowing certain acceptable levels of errors. A main result is parameterized probabilistic approximations. This can be done by replacing the values 1 and 0 in equation (6) by a pair of parameters α and β. In order to ensure that the lower approximation is a subset of the upper approximation, we require $\alpha \geq \beta$. In this paper, we consider the case with $\alpha > \beta$. By generalizing equation (6), the (α, β)-probabilistic lower and upper approximations are defined by:

$$\begin{aligned} \underline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \geq \alpha\}, \\ \overline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) > \beta\}. \end{aligned} \tag{7}$$

From the (α, β)-probabilistic lower and upper approximations, we can obtain the (α, β)-probabilistic positive, boundary and negative regions:

$$\begin{aligned} \text{POS}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \geq \alpha\}, \\ \text{BND}_{(\alpha,\beta)}(A) &= \{x \in U \mid \beta < Pr(A|[x]) < \alpha\}, \\ \text{NEG}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \leq \beta\}. \end{aligned} \tag{8}$$

Alternatively, one may start from the three probabilistic regions and define the probabilistic approximations.

There are other methods for defining probabilistic rough set approximations. For example, instead of using the threshold values α and β, Ślęzak and Ziarko [27, 29] define probabilistic approximations based on the comparison of the *a posteriori* probability $Pr(A|[x])$ and the *a priori* probability $Pr(A)$:

$$\begin{aligned} \text{POS}_{sz}(A) &= \{x \in U \mid Pr(A|[x]) > Pr(A)\}, \\ \text{BND}_{sz}(A) &= \{x \in U \mid Pr(A|[x]) = Pr(A)\}, \\ \text{NEG}_{sz}(A) &= \{x \in U \mid Pr(A|[x]) < Pr(A)\}. \end{aligned} \tag{9}$$

It corresponds to the case where $\alpha = \beta$, for which a slightly different definition of probabilistic approximations is used [41, 45, 46]. In some sense, their definition may be viewed as a method for setting concept dependent threshold values. For a subset A, representing the instances of a certain concept, the threshold value is given by $\alpha = \beta = Pr(A)$. Since the conditional probability $Pr(A|[x])$ may be difficult to estimate, one may define probabilistic approximations in terms of the easy to estimate conditional probability $Pr([x]|A)$, or the likelihood ratio $Pr([x]|A)/Pr([x]|A^c)$, through Bayes theorem [27, 43]. More specifically, with respect to the likelihood ratio we have [43]:

$$\begin{aligned} \text{POS}_{(\alpha,\beta)}(A) &= \{x \in U \mid \frac{Pr([x]|A)}{Pr([x]|A^c)} \geq \alpha'\}, \\ \text{BND}_{(\alpha,\beta)}(A) &= \{x \in U \mid \beta' < \frac{Pr([x]|A)}{Pr([x]|A^c)} < \alpha'\}, \\ \text{NEG}_{(\alpha,\beta)}(A) &= \{x \in U \mid \frac{Pr([x]|A)}{Pr([x]|A^c)} \leq \beta'\}, \end{aligned} \tag{10}$$

where $\alpha' = (Pr(A^c)\alpha)/(Pr(A)(1-\alpha))$ and $\beta' = (Pr(A^c)\beta)/(Pr(A)(1-\beta))$. In the rest of this paper, we use the simpler form given by equation (8). The same argument can be easily made with respect to other forms.

The construction process of probabilistic rough set approximations depends crucially on a pair of threshold values. A definition of probabilistic approximations is meaningful only if the pair of parameters can be meaningfully interpreted and practically obtainable. This immediately leads to an important semantic issue in a probabilistic rough set model.

Meaningful construction of probabilistic approximations: This concerns the interpretation of the required parameters α and β. There are several related questions. How do we determine the parameters on a solid theoretical and practical basis? Can we design a systematic method for estimating those parameters?

In many applications of a probabilistic rough set model, one implicitly assumes that the parameters α and β can be supplied by a user based on an intuitive understanding of the levels of tolerance for errors [50]. Unfortunately, there is not any guideline offered to a user; a user often needs to search for a good pair of parameters on a trial and error basis. Such an *ad hoc* approach, although mostly practiced when applying probabilistic rough sets, is hardly satisfactory. A more sound solution to this issue has been given within the decision-theoretic rough set model [41, 45, 46]. The parameters are interpreted in terms of loss functions that can be related to more intuitive terms such as costs, benefits, and risks based on well established Bayesian decision theory. They can be systematically calculated based on the loss functions.

2.2. Classification with Probabilistic Rough Set Approximations

An important application of rough set theory is to induce classification or decision rules that indicate the decision class of an object based on its values on some condition attributes [20, 23].

In rough set theory, a decision or classification rule is expressed in the following general form:

$$\mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), \quad (m_1, \ldots, m_k). \tag{11}$$

where $\mathrm{Des}(\cdot)$ represents a description (typically a logic expression) of a subset of objects [20, 25], $[x]$ is an equivalence class induced by a set of condition attributes, $A \subseteq U$ denote a decision class representing objects with the same label on decision attributes or instances of a concept, and $m_1, \ldots, m_k$ are values of a family of measures characterizing the rule [31, 47]. A decision rule predicts the decision class of an object based on its values on condition attributes. Many methods have been proposed to induce decision rules from either the pair of lower and upper approximations or the three regions. In this paper, we consider the latter methods.

For a given a rule, our confidence on the rule, or the accuracy of the rule, is defined by:

$$c(\mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A)) = Pr(A|[x]) = \frac{|A \cap [x]|}{|[x]|}. \tag{12}$$

Based on the confidence measure, we examine the characteristics of rules derived from the three regions of A. With respect to the positive, boundary and negative, we can derive the following three classes of rules:

$$\begin{array}{lll} \text{for } [x] \subseteq \mathrm{POS}(A): & \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), & (c = 1); \\ \text{for } [x] \subseteq \mathrm{BND}(A): & \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), & (0 < c < 1); \\ \text{for } [x] \subseteq \mathrm{NEG}(A): & \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), & (c = 0). \end{array}$$

The values of the confidence measure reflect truthfully the qualitative nature of these rules. Based on such a characterization, we review three major interpretations of rules in the Pawlak rough set model.

The lower approximation consists of those objects that *certainly* belong to the decision class, and the upper approximation consists of those objects that only *possibly* belong to the decision class [20].

Accordingly, Grzymala-Busse [9] suggests that two categories of rules can be induced: "certain rules" from the lower approximation and "possible rules" from the upper approximation. Lingras and Yao [18] referred to rules from the upper approximation as "plausibilistic rules," based on a connection between lower and upper approximations and belief and plausibility functions, respectively. Since the lower approximation is a subset of the upper approximation, there is an overlap between the two sets of rules. This may be considered to be an undesirable feature.

For objects in the positive and negative regions, we have $c = 1$ and $c = 0$, respectively. Thus, we can make *deterministic decisions* about their memberships in the given decision class. We can only make *nondeterministic decisions* for objects in the boundary region. Based on this observation, Wong and Ziarko [33] propose two types of rules: "deterministic decision rules" for the positive region and the negative region, and "undeterministic rules" for the boundary region. Since the three regions are pair-wise disjoint, the derived rule sets no longer have an overlap.

The confidence of a rule from the positive region is 1, namely, totally *certain*, and that of a rule from the boundary region is between 0 and 1 exclusively, namely, *uncertain*. Pawlak [21] refers to them, respectively, as "certain decision rules" and "uncertain decision rules." In his book [20] and earlier works, Pawlak focuses mainly on positive region and certain rules, as they characterize the objects on which we can make consistent and correct decisions.

Although other classifications and interpretations of rules have been considered in rough set theory, they are basically variations of the three. Within the classical rough set theory, these interpretations are meaningful. They truthfully reflect the *qualitative, statistical*, or *syntactical* nature of rules, *certain* versus *possible* [9], *deterministic* versus *nondeterministic* [33], and *certain* versus *uncertain* [21].

One may directly apply the results from the Pawlak rough set model to a probabilistic rough set model. With respect to three regions, the following three types of rules can be immediately obtained:

$$\begin{aligned}
\text{for } [x] \subseteq \mathrm{POS}_{(\alpha,\beta)}(A): &\quad \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), \quad (c \geq \alpha);\\
\text{for } [x] \subseteq \mathrm{BND}_{(\alpha,\beta)}(A): &\quad \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), \quad (\beta < c < \alpha);\\
\text{for } [x] \subseteq \mathrm{NEG}_{(\alpha,\beta)}(A): &\quad \mathrm{Des}([x]) \longrightarrow \mathrm{Des}(A), \quad (c \leq \beta).
\end{aligned}$$

The values of the confidence measure truthfully reflect the quantitative nature of probabilistic rules. On the other hand, the classification of rules into three types represents the qualitative nature of the rule and is obtained based on the pair of parameters (α, β). An object in the probabilistic positive region does *not certainly* belong to the decision class, but does so with a *high probability*. Like a probabilistic rule from the probabilistic boundary region, a rule from a probabilistic positive region may be *uncertain* and *nondeterministic*.

It is evident that qualitative interpretations of rules of the Pawlak rough set model are no longer appropriate for a probabilistic rough set model. On the other hand, many studies have simply used the same notions of the classical rough set model in a probabilistic model, without considering their differences. There remains a semantic difficulty and confusion in interpreting probabilistic rules.

Meaningful classification with probabilistic approximations: This concerns the interpretation of probabilistic decision rules. There are several related questions. What are the semantic differences between rules from three regions? How to apply probabilistic rules?

The solution to the issue relies on making a distinction between statistical properties of a rule and applications of a rule. The differences, expressed in terms of statistical measures of rules, lead to totally different decisions. It can be adequately explained based on the recently introduced notion of three-way decisions [43], within the framework the decision-theoretic rough set model.

3. An Interpretation of Threshold Parameters

The Bayesian decision procedure deals with making decision with minimum risk based on observed evidence. We present a brief description of the procedure from the book by Duda and Hart [5] and apply the procedure for the interpretation and construction of probabilistic approximations [45, 46].

3.1. 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 $Pr(w_j|\mathbf{x})$ be the conditional probability of an object x being in state w_j given that the object is described by $\mathbf{x}$. Typically, $\mathbf{x}$ is a feature vector representing a set of objects defined by the vector, which is similar to the definition of an equivalence class in rough set theory. In a classification or pattern recognition problem, each state represent a class. The probability $Pr(w_j)$ may be viewed as the *a priori* probability that an arbitrary object is in the class w_j; the probability $Pr(w_j|\mathbf{x})$ is the *a posteriori* probability after observing that the object is described by a feature vector $\mathbf{x}$. 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 $Pr(w_j|\mathbf{x})$ 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:

$$R(a_i|\mathbf{x}) = \sum_{j=1}^{s} \lambda(a_i|w_j) Pr(w_j|\mathbf{x}). \tag{13}$$

The quantity $R(a_i|\mathbf{x})$ is also called the conditional risk.

Given a description $\mathbf{x}$, a decision rule is a function $\tau(\mathbf{x})$ that specifies which action to take. That is, for every $\mathbf{x}$, $\tau(\mathbf{x})$ takes 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:

$$\mathbf{R} = \sum_{\mathbf{x}} R(\tau(\mathbf{x})|\mathbf{x}) Pr(\mathbf{x}), \tag{14}$$

where the summation is over the set of all possible descriptions of objects. 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. Thus, 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 (13) and select the action for which the conditional risk is minimum. If more than one action minimizes $R(a_i|\mathbf{x})$, a tie-breaking criterion can be used.

3.2. Decision-theoretic rough set model

To match the Bayesian decision procedure, the decision-theoretic rough set model starts from defining the three probabilistic regions from which the probabilistic approximations are defined.

We have a set of 2 states and a set of 3 actions for each state. The set of states is given by $\Omega = \{A, A^c\}$ indicating that an element is in A and not in A, respectively. For simplicity, we use the same symbol to denote both a subset A and the corresponding state. With respect to the three regions, the set of actions with respect to a state is given by $\mathcal{A} = \{a_P, a_B, a_N\}$, where a_P, a_B, and a_N represent the three actions in classifying an object x, namely, deciding $x \in \mathrm{POS}(A)$, deciding $x \in \mathrm{BND}(A)$, and deciding $x \in \mathrm{NEG}(A)$, respectively. The losses regarding the risk or cost of those classification actions with respect to different states are given by the 3×2 matrix:

	A (P)	A^c (N)
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

In the matrix, λ_{PP}, λ_{BP} and λ_{NP} denote the losses incurred for taking actions a_P, a_B and a_N, respectively, when an object belongs to A, and λ_{PN}, λ_{BN} and λ_{NN} denote the losses incurred for taking the same actions when the object does not belong to A. Methods for interpreting and estimating the loss functions can be found in the literature of decision theory and some recent papers in the context of DTRSM [14, 49].

The expected losses associated with taking different actions for objects in $[x]$ can be expressed as:

$$\begin{aligned} R(a_P|[x]) &= \lambda_{PP} Pr(A|[x]) + \lambda_{PN} Pr(A^c|[x]), \\ R(a_B|[x]) &= \lambda_{BP} Pr(A|[x]) + \lambda_{BN} Pr(A^c|[x]), \\ R(a_N|[x]) &= \lambda_{NP} Pr(A|[x]) + \lambda_{NN} Pr(A^c|[x]). \end{aligned} \tag{15}$$

The Bayesian decision procedure suggests the following minimum-risk decision rules:

(P) If $R(a_P|[x]) \leq R(a_B|[x])$ and $R(a_P|[x]) \leq R(a_N|[x])$, decide $x \in \mathrm{POS}(A)$;

(B) If $R(a_B|[x]) \leq R(a_P|[x])$ and $R(a_B|[x]) \leq R(a_N|[x])$, decide $x \in \mathrm{BND}(A)$;

(N) If $R(a_N|[x]) \leq R(a_P|[x])$ and $R(a_N|[x]) \leq R(a_B|[x])$, decide $x \in \mathrm{NEG}(A)$.

In order to make sure that the three regions are mutually disjoint, tie-breaking criteria should be added. In other words, with tie-breaking criteria, each object is put into one and only one region. One such a criterion will be introduced in the subsequent discussion.

Since $Pr(A|[x]) + Pr(A^c|[x]) = 1$, we can simplify the rules based only on the probabilities $Pr(A|[x])$ and the loss function λ. Consider a special kind of loss functions with:

$$\begin{aligned} \text{(c0).} \qquad & \lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}, \\ & \lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}. \end{aligned} \tag{16}$$

That is, the loss of classifying an object x belonging to A into the positive region $\mathrm{POS}(A)$ is less than or equal to the loss of classifying x into the boundary region $\mathrm{BND}(A)$, and both of these losses are strictly less than the loss of classifying x into the negative region $\mathrm{NEG}(A)$. The reverse order of losses is used for classifying an object not in A. Under condition (c0), we can simplify decision rules (P)-(N) as follows. For the rule (P), the first condition can be expressed as:

$$\begin{aligned}
& R(a_P|[x]) \leq R(a_B|[x]) \\
\Longleftrightarrow \quad & \lambda_{PP} Pr(A|[x]) + \lambda_{PN} Pr(A^c|[x]) \leq \lambda_{BP} Pr(A|[x]) + \lambda_{BN} Pr(A^c|[x]) \\
\Longleftrightarrow \quad & \lambda_{PP} Pr(A|[x]) + \lambda_{PN}(1 - Pr(A|[x])) \leq \lambda_{BP} Pr(A|[x]) + \lambda_{BN}(1 - Pr(A|[x])) \\
\Longleftrightarrow \quad & Pr(A|[x]) \geq \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}. \qquad (17)
\end{aligned}$$

Similarly, the second condition of rule (P) can be expressed as:

$$\begin{aligned}
& R(a_P|[x]) \leq R(a_N|[x]) \\
\Longleftrightarrow \quad & \lambda_{PP} Pr(A|[x]) + \lambda_{PN} Pr(A^c|[x]) \leq \lambda_{NP} Pr(A|[x]) + \lambda_{NN} Pr(A^c|[x]) \\
\Longleftrightarrow \quad & \lambda_{PP} Pr(A|[x]) + \lambda_{PN}(1 - Pr(A|[x])) \leq \lambda_{NP} Pr(A|[x]) + \lambda_{NN}(1 - Pr(A|[x])) \\
\Longleftrightarrow \quad & Pr(A|[x]) \geq \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \qquad (18)
\end{aligned}$$

The first condition of rule (B) is the converse of the first condition of rule (P). It follows,

$$R(a_B|[x]) \leq R(a_P|[x]) \quad \Longleftrightarrow \quad Pr(A|[x]) \leq \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}. \qquad (19)$$

For the second condition of rule (B), we have:

$$\begin{aligned}
& R(a_B|[x]) \leq R(a_N|[x]) \\
\Longleftrightarrow \quad & \lambda_{BP} Pr(A|[x]) + \lambda_{BN} Pr(A^c|[x]) \leq \lambda_{NP} Pr(A|[x]) + \lambda_{NN} Pr(A^c|[x]) \\
\Longleftrightarrow \quad & \lambda_{BP} Pr(A|[x]) + \lambda_{BN}(1 - Pr(A|[x])) \leq \lambda_{NP} Pr(A|[x]) + \lambda_{NN}(1 - Pr(A|[x])) \\
\Longleftrightarrow \quad & Pr(A|[x]) \geq \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \qquad (20)
\end{aligned}$$

The first condition of rule (N) is the converse of the second condition of rule (P) and the second condition of rule (N) is the converse of the second condition of rule (B). It follows,

$$\begin{aligned}
R(a_N|[x]) \leq R(a_P|[x]) \quad & \Longleftrightarrow \quad Pr(A|[x]) \leq \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}, \\
R(a_N|[x]) \leq R(a_B|[x]) \quad & \Longleftrightarrow \quad Pr(A|[x]) \leq \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \qquad (21)
\end{aligned}$$

To obtain a compact form of the decision rules, we denote the three expressions in these conditions by the following three parameters:

$$\begin{aligned}
\alpha &= \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \\
\beta &= \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}, \\
\gamma &= \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \qquad (22)
\end{aligned}$$

The decision rules (P)-(N) can be expressed concisely as:

$$
\begin{array}{ll}
\text{(P)} & \text{If } Pr(A|[x]) \geq \alpha \text{ and } Pr(A|[x]) \geq \gamma, \text{ decide } x \in \mathrm{POS}(A);\\
\text{(B)} & \text{If } Pr(A|[x]) \leq \alpha \text{ and } Pr(A|[x]) \geq \beta, \text{ decide } x \in \mathrm{BND}(A);\\
\text{(N)} & \text{If } Pr(A|[x]) \leq \beta \text{ and } Pr(A|[x]) \leq \gamma, \text{ decide } x \in \mathrm{NEG}(A).
\end{array}
$$

Each rule is defined by two out of the three parameters.

Those rules can be further simplified by considering some constraints on the threshold values α and β. To be consistent with the interpretation of rough sets, we may assume that probabilistic lower approximation is a subset of probabilistic upper approximation. In addition, it is possible that the boundary region is not empty. To meet those two constraints, the conditions of rule (B) suggest that $\alpha > \beta$ may be a reasonable condition. To obtain the condition on the loss function, we set $\alpha > \beta$, namely,

$$\frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} > \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \tag{23}$$

This immediately leads to the following condition on the loss function [41]:

$$\text{(c1).} \qquad \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}} > \frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}}. \tag{24}$$

The condition (c1) implies that $1 \geq \alpha > \gamma > \beta \geq 0$. In this case, we introduce tie-breaking criteria that are in favor of positive and negative regions. After tie-breaking, the following simplified rules are obtained [41]:

$$
\begin{array}{ll}
\text{(P1)} & \text{If } Pr(A|[x]) \geq \alpha, \text{ decide } x \in \mathrm{POS}(A);\\
\text{(B1)} & \text{If } \beta < Pr(A|[x]) < \alpha, \text{ decide } x \in \mathrm{BND}(A);\\
\text{(N1)} & \text{If } Pr(A|[x]) \leq \beta, \text{ decide } x \in \mathrm{NEG}(A).
\end{array}
$$

The parameter γ is no longer needed. Each object can be put into one and only one region by using rules (P1), (B1) and (N1).

From the rules (P1), (B1) and (N1), the (α, β)-probabilistic positive, negative and boundary regions are given, respectively, by:

$$
\begin{aligned}
\mathrm{POS}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \geq \alpha\},\\
\mathrm{BND}_{(\alpha,\beta)}(A) &= \{x \in U \mid \beta < Pr(A|[x]) < \alpha\},\\
\mathrm{NEG}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \leq \beta\}.
\end{aligned} \tag{25}
$$

The (α, β)-probabilistic lower and upper approximations are defined by:

$$
\begin{aligned}
\underline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) \geq \alpha\},\\
\overline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid Pr(A|[x]) > \beta\}.
\end{aligned} \tag{26}
$$

They are exactly the three probabilistic regions and the pair of probabilistic approximations as introduced in the last section. Therefore, we have given a solid theoretical basis and a practical interpretation of the probabilistic rough sets. The threshold parameters can be systematically calculated from a loss function based on the Bayesian decision procedure. The results therefore provide a solution to the first semantic issue in a probabilistic rough set model.

For the case of $\alpha = \gamma = \beta$, a tie-breaking criterion that is in favor of the boundary region can be used and a different definition of probabilistic approximations can be obtained [45], which is a generalization of 0.5-probabilistic rough set model [25] and is consistent with equation (9). By following a similar derivation, one can find conditions on the loss function for the case of $\alpha = \gamma = \beta$. For the case of $\alpha < \gamma < \beta$, the boundary region is always empty, and one can use rules (P) and (N) to approximate a set by two regions [44]. In this case, the loss function suggests that it is better to use two regions as approximations, rather than three regions. Again, one can derive conditions on the loss function for the case of $\alpha < \gamma < \beta$. The analysis shows that rough set theory is meaningful for the first two cases (i.e., $\alpha > \beta$), although the second case represents a rather peculiar situation. Two regions without a boundary region are sufficient for the last case. A detailed analysis is given in another paper [41]

4. An Interpretation of Probabilistic Regions for Three-Way Decisions

A new interpretation of rules in rough set models is presented by considering both their qualitative and quantitative natures. Qualitatively, rules are grouped into three classes for making three-way decisions; quantitatively, rules are characterized by the values of the associated measures.

4.1. Positive, boundary, and negative rules for three-way decisions

In order to address the problem of indiscernibility of objects due to a lack of information and knowledge, the theory of rough sets introduces three regions for decisions, i.e., $\mathrm{POS}(A)$, $\mathrm{BND}(A)$, and $\mathrm{NEG}(A)$ for objects *certainly* in A, *possibly* in A, and *certainly not* in A. In DTRSM, this three-way decision is modeled as $x \in \mathrm{POS}(A)$, $x \in \mathrm{BND}(A)$, and $x \in \mathrm{NEG}(A)$. This interpretation of three-way decisions seems to capture more truthfully the philosophy of the rough set theory.

Although many studies focus on deriving rules from the individual regions, they do not explicitly record the region information in the induced rules [9, 20, 21, 23, 24, 33]. This leads to an interpretation of rules in terms of certain and uncertain decisions, rather than the different decisions implied by each rule. More specifically, a rule from the positive region implies a decision $x \in \mathrm{POS}(A)$, and a rule from the boundary region implies a decision $x \in \mathrm{BND}(A)$. Semantically, they result in very different decisions. For this reason, the notion of three-way decisions is introduced for modeling decision rules from the three regions [43]. With respect to the three regions, we introduce the following three-way decision rules:

$$
\begin{aligned}
\text{for } [x] \subseteq \mathrm{POS}(A): &\quad \mathrm{Des}([x]) \longrightarrow_P \mathrm{Des}(A), \quad (c = 1);\\
\text{for } [x] \subseteq \mathrm{BND}(A): &\quad \mathrm{Des}([x]) \longrightarrow_B \mathrm{Des}(A), \quad (0 < c < 1);\\
\text{for } [x] \subseteq \mathrm{NEG}(A): &\quad \mathrm{Des}([x]) \longrightarrow_N \mathrm{Des}(A), \quad (c = 0).
\end{aligned}
$$

The subscript of $\longrightarrow$ indicates the type of a rule: a positive rule makes a decision of acceptance, a negative rule makes a decision of rejection, and a boundary rules makes an abstained, a non-committed or a deferred decision [12, 43]. On the other hand, the associated measure reflects our confidence on the decision. The distinction of the *type* of a decision and the *confidence* on the decision is very important; it avoids the semantic inconsistency in explaining rules in the rough set theory.

The three types of rules have very different semantic interpretations, and each of them leads to a different decision. A positive rule allows us to *accept* an object to be a member of A and a negative rule allows us to *reject* an object to be a member of A, with an understanding that such an acceptance or a rejection is made with certain levels of tolerance of errors. On the other hand, a boundary rule does not offer such a definite decision, and due to uncertainty and inconsistency, we are forced to, or happy to, make an indecision or a delayed decision, which warrants a further investigation [37].

The interpretation, based on a three-way decision framework, seems to be more suitable for viewing rules in rough set theory. Pawlak, Wong, and Ziarko [25] considered a similar interpretation based on three-valued decisions, consisting of "yes", "no", and "do not know." Herbert and Yao [11] interpreted the positive and negative rules as providing "immediate decisions" of "yes" and "no", respectively, and boundary rules as providing "delayed decisions" of "wait-and-see." Li, Zhang and Swan [17] considered the three-way decision as providing "three distinct regions of relevance" in the context of information retrieval. Ślęzak *et al*. [28] present another application of three-way decisions for choosing data packs for query optimization. The three-way decisions lead to a ternary classification of data packs into the relevant, irrelevant, and suspect data packs. Lazaridis and Mehrotra [16] use a technique related to three-way rough set approximations in approximating aggregate queries.

According to the three probabilistic regions, one can make three-way decisions based on the following positive, boundary and negative rules:

$$\begin{array}{rll} \text{for } [x] \subseteq \mathrm{POS}_{(\alpha,\beta)}(A): & \mathrm{Des}([x]) \longrightarrow_P \mathrm{Des}(A), & (c \geq \alpha); \\ \text{for } [x] \subseteq \mathrm{BND}_{(\alpha,\beta)}(A): & \mathrm{Des}([x]) \longrightarrow_B \mathrm{Des}(A), & (\beta < c < \alpha); \\ \text{for } [x] \subseteq \mathrm{NEG}_{(\alpha,\beta)}(A): & \mathrm{Des}([x]) \longrightarrow_N \mathrm{Des}(A), & (c \leq \beta). \end{array}$$

Recall that the conditional probability $c = Pr(A|[x])$ is the accuracy and confidence of a rule. Unlike rules in the classical rough set theory, all three types of rules may be *uncertain*. They represent the levels of tolerance in making incorrect decisions. For positive rules, the error rate of accepting a non-member of A as a member of A is defined by $Pr(A^c|[x]) = 1 - Pr(A|[x]) = 1 - c$ and is below $1 - \alpha$. Conversely, the error rate of rejecting a member of A as a non-member of A is given by $Pr(A|[x]) = c$ and is below β. When the conditional probability is too low for acceptance and at the same time too high for rejection, we choose a boundary rule for an abstained decision, an indecision or a delayed or deferred decision. The error rates for putting an object in A and an object not in A into the boundary region are $Pr(A|[x]) = c$ and $Pr(A^c|[x]) = 1 - Pr(A|[x]) = 1 - c$, respectively.

In addition to resolving a semantic inconsistency, the three-way interpretation has additional advantages. We may establish close relationships between rough set analysis, Bayesian decision analysis, and hypothesis testing in statistics [6, 7, 35]. A few related studies are summarized as examples. Woodward and Naylor [35] discuss Bayesian methods in statistical process control. A pair of threshold values on the posterior odds ratio is used to make a three-stage decision about a process: *accept without further inspection*, *adjust (reject) and continue inspecting*, or *continue inspecting*. Forster [6] considers the importance of model selection criteria with a three-way decision: *accept*, *reject* or *suspend judgment*. Goudey [7] discusses three-way statistical inference that supports three possible actions for an environmental manager: *act as if there is no problem*, *act as if there is a problem*, or *act as if there is not yet sufficient information to allow a decision*. With the new interpretation of rules as three-way decisions with tolerance levels of errors, we can greatly enlarge the application domain of rough set theory.

4.2. Cost analysis of three-way decisions

The semantic differences between three types of rules can be further demonstrated by their associated cost or risk within the framework of DTRSM.

For three-way decision rules, their associated different costs can be computed as follows:

$$\begin{aligned} \text{positive rule}: &\quad c * \lambda_{PP} + (1-c) * \lambda_{PN}, \\ \text{boundary rule}: &\quad c * \lambda_{BP} + (1-c) * \lambda_{BN}, \\ \text{negative rule}: &\quad c * \lambda_{NP} + (1-c) * \lambda_{NN}, \end{aligned} \tag{27}$$

where $c = Pr(A|[x])$ for rule $\text{Des}([x]) \longrightarrow_\Lambda \text{Des}(A)$, $\Lambda \in \{P, B, N\}$. In general, these costs are different. Under the conditions (c0) and (c1), the pair of parameters, (α, β), satisfies the property $\alpha > \beta$. By the definition of three regions through (P1), (B1) and (N1), the costs of rules are bounded by:

$$\begin{aligned} \text{positive rule}: &\quad \lambda_{PP} \leq \mathit{cost} \leq \alpha\lambda_{PP} + (1-\alpha) * \lambda_{PN}, \\ \text{boundary rule}: &\quad \text{if } \lambda_{BP} = \lambda_{BN}, \\ &\quad\quad \mathit{cost} = \lambda_{BP} = \lambda_{BN}, \\ &\quad \text{if } \lambda_{BP} < \lambda_{BN}, \\ &\quad\quad \alpha * \lambda_{BP} + (1-\alpha) * \lambda_{BN} < \mathit{cost} < \beta * \lambda_{BP} + (1-\beta) * \lambda_{BN}, \\ &\quad \text{if } \lambda_{BP} > \lambda_{BN}, \\ &\quad\quad \beta * \lambda_{BP} + (1-\beta) * \lambda_{BN} < \mathit{cost} < \alpha * \lambda_{BP} + (1-\alpha) * \lambda_{BN}, \\ \text{negative rule}: &\quad \lambda_{NN} \leq \mathit{cost} \leq \beta * \lambda_{NP} + (1-\beta) * \lambda_{NN}. \end{aligned} \tag{28}$$

The bounds for the positive and negative rules are obtained based on the assumptions (c0) and (c1). The bounds for the boundary rules depend additionallu on the relationships between λ_{BP} and λ_{BN}.

Consider the special case where we assume zero cost for a correct decision, namely, $\lambda_{PP} = \lambda_{NN} = 0$. The costs of rules can be simplified into:

$$\begin{aligned} \text{positive rule}: &\quad (1-c) * \lambda_{PN}, \\ \text{boundary rule}: &\quad c * \lambda_{BP} + (1-c) * \lambda_{BN}, \\ \text{negative rule}: &\quad c * \lambda_{NP}. \end{aligned} \tag{29}$$

In this case, they are only related to the misclassification error rates and therefore are much easier to understand. Assume further that we have a unit cost for misclassification, namely, $\lambda_{PN} = \lambda_{NP} = 1$, the costs of positive and negative rules are reduced to the misclassification error rates:

$$\begin{aligned} \text{positive rule}: &\quad (1-c), \\ \text{boundary rule}: &\quad c * \lambda_{BP} + (1-c) * \lambda_{BN}, \\ \text{negative rule}: &\quad c. \end{aligned} \tag{30}$$

If we assume again that $\lambda_{BP} = \lambda_{BN}$, the cost of a boundary rule becomes $\lambda_{BP} = \lambda_{BN}$. Studies that use only the classification error rates may therefore be considered as a special case.

The notion of three-way decisions enables us to separate the type of decisions implied by each rule and our confidence on the rule. The cost analysis demonstrates that each type of rules is associated with a different cost. This provides a satisfactory explanation of the second semantic issue in a probabilistic rough set model.

5. Conclusion

A lack of studies on semantic issues in rough set theory has led to inconsistent interpretations of the theory, misuses of the theory, and meaningless generalizations of the theory. Semantic studies are a much needed, and perhaps more fruitful, research direction in the next few years.

This paper focuses on two fundamental semantic issues in a probabilistic rough set model. One is the interpretation and determination of the required parameters; the other is the interpretation and application of rules from the probabilistic positive, boundary and negative regions. Within the framework of the decision-theoretic rough set model, we show that the required parameters can be interpreted and determined systematically based on the well-known Bayesian decision procedure with respect to losses associated with various decisions. We introduce the notion of three-way decision rules by constructing positive, negative, and boundary rules from the three probabilistic regions. They represent, respectively, the three-way decisions of acceptance, rejection, and deferment. Similar to the ideas of hypothesis testing in statistics, the decisions of acceptance and rejection are made with certain levels of tolerance for errors. A cost analysis shows the difference between the three different types of rules.

References

[1] Aczél, J. Measuring information beyond communication theory : Some probably useful and some almost certainly useless generalizations, *Information Processing & Management*, **20**, 383-395, 1984.

[2] Ciucci, D. Approximation algebra and framework, *Fundamenta Informaticae*, **94**, 147-161, 2009.

[3] Dembczynski, K., Greco,S., Kotlowski, W. and Slowinski, R. Statistical model for rough set approach to multicriteria classification, *Proceedings of PKDD 2007*, LNAI 4702, 164-175, 2007.

[4] Dubois, D. and Prade, H. The three semantics of fuzzy sets, *Fuzzy Sets and Systems*, **90**, 141-150, 1997.

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

[6] Forster, M.R. Key concepts in model selection: performance and generalizability, *Journal of Mathematical Psychology*, **44**, 205-231, 2000.

[7] Goudey, R. Do statistical inferences allowing three alternative decision give better feedback for environmentally precautionary decision-making, *Journal of Environmental Management*, **85**, 338-344, 2007.

[8] Greco, S., Matarazzo, B. and Słowiński, R. Parameterized rough set model using rough membership and Bayesian confirmation measures, *International Journal of Approximate Reasoning*, **49**, 285-300, 2009.

[9] Grzymala-Busse, J.W. Knowledge acquisition under uncertainty - a rough set approach, *Journal of intelligent and Robotic Systems*, **1**, 3-16, 1988.

[10] Hempel, C.G. *Philosophy of Natural Science*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1966.

[11] Herbert, J.P. and Yao, J.T. Criteria for choosing a rough set model, *Journal of Computers and Mathematics with Applications*, **57**, 908-918, 2009.

[12] Herbert, J.P. and Yao, J.T. Game-theoretic risk analysis in decision-theoretic rough sets, *Proceedings of RSKT'08*, LNAI 5009, 132-139, 2008.

[13] Herbert, J.P. and Yao, J.T. Game-theoretic rough sets, *Fundamenta Informaticae*, this issue, 2010.

[14] Herbert, J.P. and Yao, J.T. Learning optimal parameters in decision-theoretic rough sets, *Proceedings of RSKT'09*, LNAI 5589, 610-617, 2009.

[15] Hughes, G.E. and Cresswell, M.J. *An Introduction to Modal Logic*, Methuen, London, 1968.

[16] Lazaridis, I. and Mehrotra, S. Progressive approximate aggregate queries with a multi-resolution tree structure, *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, 401-412, 2001.

[17] Li, Y., Zhang, C. and Swan, J.R. An information fltering model on the Web and its application in JobAgent, *Knowledge-Based Systems*, **13**, 285-296, 2000.

[18] Lingras, P.J. and Yao, Y.Y. Data mining using extensions of the rough set model, *Journal of the American Society for Information Science*,

[19] Pawlak, Z. Rough sets, *International Journal of Computer and Information Sciences*, **11**, 341-356, 1982.

[20] Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.

[21] Pawlak, Z. Rough sets, decision algorithms and Bayes' theorem, *European Journal of Operational Research*, **136**, 181-189, 2002.

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

[23] Pawlak, Z. and Skowron, A. Rudiments of rough sets, *Information Sciences*, **177**, 3-27, 2007.

[24] Pawlak, Z. and Skowron, A. Rough sets: some extensions, *Information Sciences*, **177**, 28-40, 2007.

[25] 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.

[26] Roberts, F.S. *Measurement Theory*, Addison-Wesley, Reading, Massachusetts, 1979.

[27] Ślęzak, D. Rough sets and Bayes factor, *LNCS Transactions on Rough Sets III*, LNCS 3400, 202229, 2005.

[28] Ślęzak, D., Wróblewski, J., Eastwood, V. and Synak, P. Brighthouse: an analytic data warehouse for ad-hoc queries, *Proceedings of the VLDB Endowment*, **1**, 1337-1345, 2008.

[29] Ślęzak, D. and Ziarko, W. The investigation of the Bayesian rough set model, *International Journal of Approximate Reasoning*, **40**, 81-91, 2005.

[30] Suraj, Z. and Grochowalski, P. The rough set database system: an overview, *LNCS Transactions on Rough Sets III*, LNCS 3400, 190-210, 2005.

[31] Tsumoto, S. Accuracy and coverage in rough set rule induction, *Proceedings of RSCTC'02*, LNAI 2475, 373-380, 2002.

[32] Wei, L.L. and Zhang, W.X. Probabilistic rough sets characterized by fuzzy sets, *Proceedings of RSFDGrC'03*, LNAI 2639, 173-180, 2003.

[33] Wong, S.K.M. and Ziarko, W. Algorithm for inductive learning, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, **34**, 271-276, 1986.

[34] 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.

[35] Woodward, P.W. and Naylor, J.C. An application of Bayesian methods in SPC, *The Statistician*, **42**, 461-469, 1993.

[36] Wu, W.Z. Upper and lower probabilities of fuzzy events induced by a fuzzy set-valued mapping, *Proceedings of RSFDGrC'05*, LNAI 3641, 345-353, 2005.

[37] Yao, J.T. and Herbert, J.P. Web-based support systems with rough set analysis, *Proceedings of RSEISP'07*, LNAI 4585, 360-370, 2007.

[38] Yao, Y.Y. Two views of the theory of rough sets in finite universes, *International Journal of Approximation Reasoning*, **15**, 291-317, 1996.

[39] Yao, Y.Y. Probabilistic approaches to rough sets, *Expert Systems*, **20**, 287-297, 2003.

[40] Yao, Y.Y. A note on definability and approximations, *LNCS Transactions on Rough Sets VII*, LNCS 4400, 274-282, 2007.

[41] Yao, Y.Y. Decision-theoretic rough set models, *Proceedings of RSKT 2007*, LNAI 4481, 1-12, 2007.

[42] Yao, Y.Y. Probabilistic rough set approximations, *International Journal of Approximation Reasoning*, **49**, 255-271, 2008.

[43] Yao, Y.Y. Three-way decisions with probabilistic rough sets, *Information Sciences*, **180**, 341-353, 2010.

[44] Yao, Y.Y. The superiority of three-way decisions in probabilistic rough set models, manuscript, 2010.

[45] 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.

[46] 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.), North-Holland, New York, 17-24, 1990.

[47] Yao, Y.Y. and Zhong, N. An analysis of quantitative measures associated with rules, *Proceedings of PAKDD'99*, LNAI 1974, 479-488, 1999.

[48] Yao, Y.Y. and Zhou, B. Naive Bayesian rough sets, *Proceedings of RSKT 2010*, LNAI 6401, 713-720, 2010.

[49] Zhou, X.Z. and Li, H.X. A multi-view decision model based on decision-theoretic rough set, *Proceedings of RSKT'09*, LNAI 5589, 650-657, 2009.

[50] Ziarko, W. Variable precision rough sets model, *Journal of Computer and Systems Sciences*, **46**, 39-59, 1993.

[51] Ziarko, W. Probabilistic approach to rough sets, *International Journal of Approximate Reasoning*, **49**, 272-284, 2008.