

Y.Y. Yao, Three-way decisions with probabilistic rough sets,
Information Sciences, Vol. 180, No. 3, pp. 341-353, 2010.

Three-way decisions with probabilistic rough sets

Yiyu Yao

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

Abstract

The rough set theory approximates a concept by three regions, namely, the positive, boundary and negative regions. 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 rule makes a decision of abstaining. This paper provides an analysis of three-way decision rules in the classical rough set model and the decision-theoretic rough set model. The results enrich the rough set theory by ideas from Bayesian decision theory and hypothesis testing in statistics. The connections established between the levels of tolerance for errors and costs of incorrect decisions make the rough set theory practical in applications.

Key words: Decision-theoretic rough sets; probabilistic rough sets; three-way decisions; hypothesis testing; Bayesian decision procedure; classification

1 Introduction

When studying and applying any theory, it is important to pay due attentions 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 limitations of the theory for solving a specific problem. This not only increases the likelihood of the

Email address: yyao@cs.uregina.ca (Yiyu Yao).

success of an application of the theory but also avoids potential misuses of the theory. For a healthy development of any theory, it may be inevitable that we periodically revisit, and revise if necessary, its existing formulations and interpretations under the light of new evidence.

Since the introduction of rough sets more than a quarter century ago [14], researchers have accumulated a vast literature on its theory and applications [17,18,24]. With the insights gained from existing studies, in this paper we re-examine the roles of lower and upper approximations in the context of rule induction. From a new semantic interpretation of the positive, boundary and negative regions, we introduce and study the notion of three-way decisions, consisting of positive, boundary and negative rules [39].

The notion of three-way decisions represents a close relationship between rough set analysis, Bayesian decision analysis, and hypothesis testing in statistics [3,4,27,31]. Ślęzak [20] points out a natural correspondence between the fundamental notions of rough sets and statistics. The set to be approximated corresponds to a hypothesis and an equivalence class to a piece of evidence; the three regions correspond to the results of a three-way decision that the hypothesis is verified positively, negatively, or undecidedly based on the evidence. In terms of rules, a positive rule makes a decision of acceptance, a negative rule makes a decision of rejection, and a boundary rule makes an abstained or non-committed decision [39].

The new interpretation of a three-way decision is not so critical in the classical rough set model since both positive and negative rules do not involve any uncertainty. It is essential in the probabilistic rough set models, where acceptance and rejection decisions are made with certain levels of tolerance for errors. Moreover, the Bayesian decision procedure can be used to develop a decision-theoretic rough set model [34,36,37,40,41]. The required parameters of probabilistic rough set approximations can be systematically determined based on costs of various decisions. The incurred costs of decision rules can be analyzed. Ideas from hypothesis testing in statistics may be used in interpreting the required parameters involved in three-way decision rules.

The three-way decision rules come closer to the philosophy of the rough set theory, namely, representing a concept using three regions instead of two [15]. This three-way decision scheme has not been considered explicitly in other theories of machine learning and rule induction, although it has been studied in other fields. By considering three-way decision rules, one may appreciate the true and unique contributions of rough set theory to machine learning and rule induction. With the introduction of the notion of three-way decision, there is a new avenue of research for rough set theory. The cost-based analysis of three-way decisions brings the theory closer to real-world applications where costs play an indispensable role [8,33].

Based on preliminary results in [39], we examine the formulations and interpretations of three-way decision rules as follows. In Section 2, we briefly review a few key studies that lead to three-way decision rules, which provides the motivations and justifications of the current study. Section 3 re-interprets rules in classical rough set model for a single concept and introduces three-way decision rules. Section 4 provides a detailed analysis of three-way decision rules in the decision-theoretic rough set model. Section 5 applies three-way decision rules for two-category classification and briefly touches upon many-category classification.

2 Studies That Lead to Three-way Decision Rules

An important application of rough set theory is rule induction. Fundamental notions, such as the lower and upper approximations of a concept, the positive, boundary and negative regions of a concept, attribute reducts, attribute-value reducts, indiscernibility relation and discernibility matrix, have been proposed and studied for rule induction. With these notions, rule induction can be conceptually modeled as searching for reducts [15,17]. The focus of this paper is on the interpretation of rules, rather than rule induction. A survey of existing studies shows that several interpretations of rules have been commonly used in rough sets research. They have led to a wide range of applications on one hand and to some confusion on the other.

A concept, or more precisely the extension of a concept, is represented by a subset of a universe of objects and is approximated by a pair of definable concepts of a logic language [35,38]. The lower approximation consists of those objects that *certainly* belong to the concept, and upper approximation consists of those objects that only *possibly* belong to the concept [15]. Accordingly, Grzymala-Busse [5] suggested that two categories of rules can be induced: *certain rules* from the lower approximation and *possible rules* from the upper approximation. Lingras and Yao [9] 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 with such an interpretation.

The lower and upper approximations divide the universe of objects into three pair-wise disjoint regions: the lower approximation as the positive region, the complement of the upper approximation as the negative region, and the difference between the upper and lower approximations as the boundary region. For objects in both positive and negative regions, 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 [30] proposed two types of rules: *deterministic decision rules* for positive region and *undeterministic decision rules* for boundary region. Since the three regions are pair-wise disjoint, the derived rule sets no longer have an overlap.

One may associate probabilistic measures, such as accuracy, confidence, and coverage, to rules [25]. The accuracy and confidence of a deterministic rule is 1, namely, totally certain, and that of an undeterministic rule is between 0 and 1 exclusively, namely, uncertain. Thus, Pawlak [16] referred to them, respectively, as *certain decision rules* and *uncertain decision rules*.

Although other classifications and interpretations of rules have been considered in rough set theory, they are basically variations of the above three. Within the classical Pawlak rough set theory, these interpretations make perfect sense. They truthfully reflect the *qualitative, statistical, or syntactical* nature of rules, *certain* versus *possible* [5], *deterministic* versus *nondeterministic* [30], and *certain* versus *uncertain* [16]. In his book and earlier works, Pawlak [15] focused mainly on the positive region and certain rules, as they characterize the objects on which we can make consistent and correct decisions.

When the classical rough set model is generalized into probabilistic rough set models [19,20,28,32,36,37,40,43], we meet difficulties with these qualitative interpretations of rules. An object in the probabilistic positive region does *not certainly* belong to the decision class, but with a *high probability* (i.e., the probability value is above certain threshold). Like a probabilistic rules from the probabilistic boundary region, a rule from the probabilistic positive region may be *uncertain* and *nondeterministic*. There remains a semantic difficulty and confusion in interpreting the two categories of probabilistic rules from, respectively, the probabilistic positive and boundary regions.

In earlier papers [36,39,42], we argued that a solution can be sought from the *semantics* of rules, rather than their *syntactical* characteristics of being *certain* or *uncertain*. Rules are interpreted and classified based on their associated actions and decisions. With respect to the positive, boundary, and negative regions, we introduced the notions of *positive rules*, *boundary rules* and *negative rules*. A positive rule makes a decision of acceptance, a negative rule makes a decision of rejection, and a boundary rule makes an abstained decision that needs further-investigation [33,42]. There are other possibilities for viewing an abstained decision, depending on particular applications. As pointed by one reviewer of this paper, it may happen that a cost-benefit analysis does not justify a decision based on information provided by the boundary region. The new interpretation given by three-way decisions resolves the above mentioned semantic difficulty and offers a new view of rules in rough set theory. This pa-

per makes a further contribution along the same line by giving a more detailed analysis of three-way decision rules.

The concept of three-way decisions plays an important role in many real world decision-making problems. One usually makes a decision based on available information and evidence. When the evidence is insufficient or weak, it might be impossible to make either a positive or a negative decision. One therefore chooses an alternative decision that is neither yes nor no. A few examples are given as illustrations. In the editorial peer review process of many journals, an editor typically makes a three-way decision about a manuscript, namely, acceptance, rejection, or further revision, based on comments from reviewers; the final decision of the third category is subject to comments from reviewers in another round of review [29]. The rationale of this sequential decision-making process may be explained by a framework of a sequential test of a statistical hypothesis proposed by Wald [27]. At any stage of the experiment, a three-way decision is made, namely, to accept the hypothesis being tested, to reject the hypothesis, and to continue the experiment by making an additional observation. The process is terminated when one of the first two decisions is made; another trial is performed if the third decision is made; the process is continued until eventually either the first or the second decision is made. This three-way decision-making strategy is commonly used in medical decision making [10,13]. In the threshold approach to clinical decision making proposed by Pauker and Kassirer [13], a pair of a “testing” threshold and a “test-treatment” threshold is used, with testing threshold being less than the test-treatment threshold. If the probability of disease is smaller than the testing threshold, there should be no treatment and no further testing; if the probability of disease is greater than the test-treatment threshold, treatment should be given without further testing. If the probability is between the two thresholds, a test should be performed so that a decision may be made depending on its results.

To a large extent, the decision-theoretic rough set model [40,41] is based on the same underlying idea of the threshold approach [13]. The three regions of rough set theory lead to three-way decisions; they correspond to the positive verification and negative verification of a hypothesis, as well as undecidability, based on given evidence [20,27]. This three-way decision perspective on the theory of rough sets may have many real-world applications. For example, based on the three regions of the rough set theory, Słezak et al. [21] consider a three-way decision when choosing data packs for query optimization; they classify data packs into the relevant, irrelevant, and suspect data packs. In an information filtering model, Li et al. [8] classify documents into three classes based on a three-way decision, namely, relevant documents, irrelevant documents, and possible relevant documents. The similar ideas can be applied to solve other problems.

3 Rules for a Single Concept

This section provides a brief summary of rules for a single concept in the classical rough set model, which motivates and sets up the stage of the current study.

3.1 The classical rough set model

Consider a simple knowledge representation scheme in which a finite set of objects is described by using a finite set of attributes. Formally, it can be defined by an information table M expressed as the tuple [15]:

$$M = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}), \quad (1)$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values for an attribute $a \in At$, and $I_a : U \rightarrow V_a$ is an information or a description function. It is assumed that the mapping I_a is single-valued. In this case, the value of an object $x \in U$ on an attribute $a \in At$ is denoted by $I_a(x)$.

Given a subset of attributes $A \subseteq At$, we define an indiscernibility relation $ind(A)$ on U as follows:

$$x \text{ } ind(A) \text{ } y \iff \forall a \in A [I_a(x) = I_a(y)]. \quad (2)$$

That is, two objects x and y are indiscernible with respect to A if and only if they have exactly the same value on every attribute in A . The indiscernibility relation $ind(A)$ is an equivalence relation. The equivalence class containing object x is denoted by $[x]_{ind(A)}$, or simply $[x]_A$ and $[x]$ if no confusion arises. The pair $apr = (U, ind(A))$ is called an approximation space.

A concept, in the classical view, is defined by a pair of intension and extension [26]. The extension consists of the instances to which the concept applies; the intension is a set of singly necessary and jointly sufficient conditions that describe the instances of the concept. Concepts are assumed to have well-defined boundaries and their extensions can be precisely defined by sets of objects. In this paper, we sometimes use the terms concepts and sets (i.e., extensions of concepts) interchangeably.

The interpretation of a concept as a pair of intension and extension enables us to study concept formation and learning in a logic setting in terms of intensions and in a set-theoretic setting in terms of extensions [38]. Reasoning

about intensions is based on logic [22]. Concept learning aims at deriving the intension of a concept based on a given extension and deriving relationships between concepts based on the relations between their extensions.

In rough set theory, an approximation space $apr = (U, ind(A))$ serves as a basic tool for inducing rules that describe a concept. The equivalence classes of $ind(A)$ are the most specific concepts that can be defined by using formulas in a decision logic language in an information table [11,15,35]. More specifically, the set $[x]$ can be defined by the formula $\bigwedge_{a \in A} (a = I_a(x))$. The σ -algebra, $\sigma(U/ind(A))$, consists of the empty set and unions of equivalence classes of $ind(A)$ and is the family of all definable sets. Suppose that the extension of a concept is given by a subset $C \subseteq U$. It may not necessarily represent a definable concept. By using equivalence classes, we obtain a pair of definable concepts as its lower and upper approximations [15]:

$$\begin{aligned} \underline{apr}(C) &= \{x \mid x \in U, [x] \subseteq C\}; \\ \overline{apr}(C) &= \{x \mid x \in U, [x] \cap C \neq \emptyset\}. \end{aligned} \quad (3)$$

It follows that $\underline{apr}(C) \subseteq C \subseteq \overline{apr}(C)$. That is, the concept C is approximated from below and above by two definable concepts. When C is a definable concept, we have $\underline{apr}(C) = C = \overline{apr}(C)$.

3.2 Three-way decision rules

The notion of three-way decision rules is introduced based on three disjoint regions produced by the lower and upper approximations. The ideas are briefly mentioned in [34] and further developed in [7,36,39,42].

Based on the approximations of C , one can divide the universe U into three disjoint regions, the positive region $POS(C)$, the boundary region $BND(C)$, and the negative region $NEG(C)$:

$$\begin{aligned} POS(C) &= \underline{apr}(C), \\ BND(C) &= \overline{apr}(C) - \underline{apr}(C), \\ NEG(C) &= U - POS(C) \cup BND(C) = U - \overline{apr}(C) = (\overline{apr}(C))^c. \end{aligned} \quad (4)$$

One can say with *certainty* that any element $x \in POS(C)$ belongs to C , and that any element $x \in NEG(C)$ does not belong to C . One cannot decide with certainty whether or not an element $x \in BND(C)$ belongs to C . With respect to the positive and negative, we can derive positive, boundary, and negative decision rules [7,34,36,39]:

$$\begin{aligned}
\text{Des}([x]) &\longrightarrow_P \text{Des}(C), \quad \text{for } [x] \subseteq \text{POS}(C); \\
\text{Des}([x]) &\longrightarrow_B \text{Des}(C), \quad \text{for } [x] \subseteq \text{BND}(C); \\
\text{Des}([x]) &\longrightarrow_N \text{Des}(C), \quad \text{for } [x] \subseteq \text{NEG}(C);
\end{aligned}$$

where $\text{Des}([x])$ denotes the logic formula defining the equivalence class $[x]$ and $\text{Des}(C)$ is the name of the concept [15,19].

The formulation offers a new interpretation of rules that is more in line with the philosophy of rough set theory. With insufficient information as provided by a set of attributes A , the rough set theory promotes a methodology for three-way decision making [39]. A positive rule is used for accepting, a negative rule for rejecting, and a boundary rule for abstaining. We accept an object as being an instance of a concept C based on a positive rule, reject an object as being an instance of C based on a negative rule, and abstain based on a boundary rule. It is the introduction of such a three-way decision that makes rough set theory unique and different from other theories. In other words, the rough set theory recognizes and models our inability to make a definite acceptance or rejection decision with insufficient information. According to a boundary rule, we can neither accept nor reject an object. On the other hand, when additional information is available, it is possible to narrow down the boundary region and to change from abstained decision to a definite decision [7].

In some sense, three-way decisions may be viewed as intermediate steps in a sequential decision making. In general, one may consider a hierarchical way of decision making. In the literature of rough sets, hierarchical classifiers have been studied by many authors [1,12]. The integration of three-way decision-making and hierarchical classifiers may produce fruitful results.

3.3 Limitations of the classical rough set model

Within the classical rough set model, existing studies on rules are unnecessarily restricted. One can generalize the notion of rules to avoid such limitations.

Pawlak [15] focused mainly on the positive region and the associated positive rules, as these rules produce certain and consistent classification. Based on the fact that the negative region can be expressed as $\text{NEG}(C) = (\text{POS}(C) \cup \text{BND}(C))^c$, the negative rules seem to be redundant. One can reject an object if it does not satisfy a positive or boundary rule. For this reason, only positive and boundary rules are considered in some studies [7,36,42]. In contrast, we explicitly identify three types of rules to emphasize on their semantics and to promote the rough-set philosophy of three-way decision making. Furthermore, if one considers many concepts, instead of a single concept, negative rules will not be redundant [42].

For a decision rule $\text{Des}([x]) \longrightarrow_{\Lambda} \text{Des}(C)$, where $\Lambda \in \{P, B, N\}$, we can associate a probabilistic measure called the accuracy or confidence of the rule as follows:

$$c(\text{Des}([x]) \longrightarrow_{\Lambda} \text{Des}(C)) = Pr(C|[x]) = \frac{|[x] \cap C|}{|[x]|}, \quad (5)$$

where $|\cdot|$ denotes the cardinality of a set, and $Pr(C|[x])$ is the conditional probability of an object in C given that the object is in $[x]$, estimated by using the cardinalities of sets. This method is only one of the possible ways, and perhaps not a very practical way, to estimate probability. In Bayesian inference, one typically transforms $Pr(C|[x])$ into $Pr(C|[x]) = Pr([x]|C)Pr(C)/Pr([x])$ by the Bayes theorem and estimates more practically the probability $Pr([x]|C)$ instead [2,20]. In this paper we use the simple calculation mainly for the purpose of illustration, so that we can concentrate on the basic ideas of tree-way decisions without distracting ourselves from the issue of probability estimation. In real applications, one needs to consider more practical ways to probability estimation.

According to the confidence values, positive, boundary, and negative rules are defined by the conditions $c = 1$, $0 < c < 1$, and $c = 0$, respectively. That is, the decisions of acceptance and rejection are made by using the two extreme values 0 and 1. Rules in classical rough set model are of qualitatively nature. The acceptance and rejection decisions are made without any tolerance of uncertainty.

The qualitative categorization may be too restrictive to be practically useful. In order to resolve this problem, we may recall some results from hypothesis testing in statistics, where the decision of rejecting a null hypothesis is made when the value of the test statistics falls into the rejection region. That is, the rejection of a null hypothesis is normally associated with a risk of erroneous rejection. When such an idea is applied to rough set theory, we need to introduce confidence levels of acceptance, abstaining, and rejection in the three-way decision making. More specifically, we can accept an object into the positive region if there is a very low risk of erroneous acceptance; similarly, we can reject an object if there is a very low risk of erroneous rejection; otherwise, we make an abstained decision. This motivates the decision-theoretic rough set model [34,36,37,40,41], which will be described in the next section.

We can adopt results from many studies on statical inference and decision involving a three-way decision [3,4,31]. Woodward and Naylor[31] discussed 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 [3] considered the importance of model selec-

tion criteria with a three-way decision: *accept*, *reject* or *suspend judgment*. Goudey [4] discussed 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 the rough set theory.

4 Probabilistic Rules for a Single Concept

Probabilistic rough set models allow a tolerance of inaccuracy in lower and upper approximations, or equivalently in the probabilistic positive, boundary, and negative regions. Based on the well established Bayesian decision procedure, the decision-theoretic rough set model provides systematic methods for deriving the required thresholds on probabilities for defining the three regions [7,34,36,37,40,41].

4.1 Decision-theoretic rough set model

The Bayesian decision procedure is widely used in many fields. In developing the decision-theoretic rough set model, we adopt the formulation for classification given in the book by Duda and Hart [2].

With respect to the membership of an object in C and the three-way decision, we have a set of 2 states and a set of 3 actions for each state. The set of states is given by $\Omega = \{C, C^c\}$ indicating that an element is in C and not in C , respectively. For simplicity, we use the same symbol to denote both the set C and the corresponding state. With respect to the three regions, the set of actions with respect to a state is given by $\mathcal{A} = \{P, B, N\}$, where P , B , and N represent the three actions in classifying an object x , namely, deciding $x \in \text{POS}(C)$, deciding $x \in \text{BND}(C)$, and deciding $x \in \text{NEG}(C)$, respectively. The loss function regarding the risk or cost of actions in different states is given by the 3×2 matrix:

	$C (P)$	$C^c (N)$
P	λ_{PP}	λ_{PN}
B	λ_{BP}	λ_{BN}
N	λ_{NP}	λ_{NN}

In the matrix, λ_{PP} , λ_{BP} and λ_{NP} denote the losses incurred for taking actions P , B and N , respectively, when an object belongs to C , and λ_{PN} , λ_{BN} and λ_{NN} denote the losses incurred for taking the same actions when the object does not belong to C .

We assume that the losses induced by different actions are given by experts in a particular domain. For example, in medical decision making, the losses may be interpreted based on the harms of treating patients who do not have the disease and the benefits of treating patients with the disease [10,23]. Losses may be estimated through techniques such as cost-effective analysis and cost-benefit analysis. One may intuitively interpret losses as costs measured in terms of the amount of money, although putting a monetary value is not always easy in practice. The estimation of losses is much domain dependent and needs careful investigation based on domain knowledge when applying the decision-theoretic rough set model.

Given the loss matrix, the expected losses associated with taking different actions for objects in $[x]$ can be expressed as [34,40,41]:

$$\begin{aligned}
R(P|[x]) &= \lambda_{PP}Pr(C|[x]) + \lambda_{PN}Pr(C^c|[x]), \\
R(B|[x]) &= \lambda_{BP}Pr(C|[x]) + \lambda_{BN}Pr(C^c|[x]), \\
R(N|[x]) &= \lambda_{NP}Pr(C|[x]) + \lambda_{NN}Pr(C^c|[x]).
\end{aligned} \tag{6}$$

The Bayesian decision procedure suggests the following minimum-risk decision rules [34,40,41]:

- (P) If $R(P|[x]) \leq R(N|[x])$ and $R(P|[x]) \leq R(B|[x])$, decide $x \in \text{POS}(C)$;
- (B) If $R(B|[x]) \leq R(P|[x])$ and $R(B|[x]) \leq R(N|[x])$, decide $x \in \text{BND}(C)$;
- (N) If $R(N|[x]) \leq R(P|[x])$ and $R(N|[x]) \leq R(B|[x])$, decide $x \in \text{NEG}(C)$.

Tie-breaking criteria should be added so that each object is put into only one region. Since $Pr(C|[x]) + Pr(C^c|[x]) = 1$, we can simplify the rules based only on the probabilities $Pr(C|[x])$ and the loss function λ .

Consider a special kind of loss functions with:

$$\begin{aligned}
\text{(c0)} \quad & \lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}, \\
& \lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}.
\end{aligned} \tag{7}$$

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

- (P) If $Pr(C|[x]) \geq \alpha$ and $Pr(C|[x]) \geq \gamma$, decide $x \in \text{POS}(C)$;
- (B) If $Pr(C|[x]) \leq \alpha$ and $Pr(C|[x]) \geq \beta$, decide $x \in \text{BND}(C)$;
- (N) If $Pr(C|[x]) \leq \beta$ and $Pr(C|[x]) \leq \gamma$, decide $x \in \text{NEG}(C)$;

where the parameters α , β , and γ are defined as:

$$\begin{aligned}
\alpha &= \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \\
\beta &= \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}, \\
\gamma &= \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}.
\end{aligned} \tag{8}$$

In other words, from a loss function one can systematically determine the required threshold parameters.

A closer examination of expressions of α , β and γ suggests that they in fact use the relative values of the loss function rather than the absolute values. That is, as long as the differences of losses inside pairs of parentheses remain the same, one would obtain the same values for α , β and γ , independent of the actual values of the loss function. Figure 1 illustrates these differences and their relationships based on line segments. We use different symbols to denote each line segment. For example, the symbol $\lambda_{(B-P)P}$ denotes the difference of losses for putting an element of C into the boundary and positive regions of C , respectively. It is indeed the increase of losses for moving an object in C from the positive region to boundary region. From the figure, we see that $\lambda_{(N-P)P} = \lambda_{(B-P)P} + \lambda_{(N-B)P}$; the longer line segment is the sum of two shorter line segments. That is, the increase of losses of moving an object in C from the positive region to the negative region in one move is the same as the sum of differences of losses of two moves, first moving the object from the positive region to the boundary region and then moving it from the boundary region to the negative region. Similar results can be observed from the losses of classifying objects not in C .

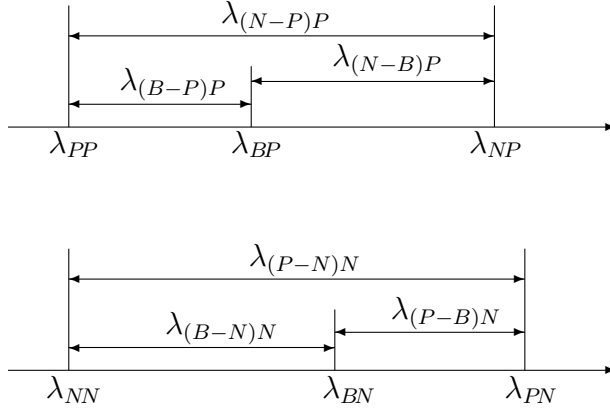


Fig. 1. Line segments interpretation of a loss function

From the analysis, it is sufficient to consider only four basic differences, consisting of $\lambda_{(P-B)N}$, $\lambda_{(B-P)P}$, $\lambda_{(B-N)N}$, and $\lambda_{(N-B)P}$. With the new notations, we have:

$$\begin{aligned}
\alpha &= \frac{\lambda_{(P-B)N}}{\lambda_{(P-B)N} + \lambda_{(B-P)P}}, \\
\beta &= \frac{\lambda_{(B-N)N}}{\lambda_{(B-N)N} + \lambda_{(N-B)P}}, \\
\gamma &= \frac{\lambda_{(P-N)N}}{\lambda_{(P-N)N} + \lambda_{(N-P)P}} \\
&= \frac{\lambda_{(B-N)N} + \lambda_{(P-B)N}}{\lambda_{(B-N)N} + \lambda_{(P-B)N} + \lambda_{(N-B)P} + \lambda_{(B-P)P}}. \tag{9}
\end{aligned}$$

By interpreting differences of losses in terms of line segments, one obtains another intuitive interpretation of the required parameters. As will be discussed later, it is much easier to state and interpret conditions on a loss function based on line segments.

Consider now an additional condition on a loss function [36]:

$$(c1) \quad \lambda_{(P-B)N} \lambda_{(N-B)P} > \lambda_{(B-N)N} \lambda_{(B-P)P}. \tag{10}$$

When $\lambda_{(B-N)N} \neq 0$ and $\lambda_{(N-B)P} \neq 0$, we can equivalently re-express the condition in terms of ratios:

$$(c1) \quad \frac{\lambda_{(P-B)N}}{\lambda_{(B-N)N}} > \frac{\lambda_{(B-P)P}}{\lambda_{(N-B)P}}. \tag{11}$$

In the rest of this paper, we use such a ratio form with the understanding

that all denominators are not zero; if a denominator is indeed zero, the ratio form should read by its corresponding product form. The left hand side of $>$ in (c1) concerns the differences of losses for taking two different actions for the *same* object not in C , and the right hand side considers similar differences of losses for the *same* object in C . Specifically, the left hand side is the ratio of difference of losses of putting an element not in C into the positive region and boundary region of C over the difference of losses of putting the element in the boundary region and negative region of C . From Figure 1, it is easier to interpret this ratio in terms of line segments. In particular, if we fix the values of λ_{NN} and λ_{PN} , the ratio decreases with the increase of λ_{BN} , namely, the ratio is monotonically decreasing with λ_{BN} . One can have a similar interpretation of the right hand side for classifying an element in C . In this case, the ratio is monotonically increasing with λ_{BP} .

Another interpretation of condition (c1) can be obtained by considering the following equivalent condition:

$$(c1') \quad \frac{\lambda_{(N-B)P}}{\lambda_{(B-N)N}} > \frac{\lambda_{(B-P)P}}{\lambda_{(P-B)N}}. \quad (12)$$

The left hand side concerns a pair of the *same* actions, namely, putting an object into the boundary and negative regions, but for two different objects such that one is in C and the other is not in C . On the other hand, the right hand side considers a pair of the *same* actions of putting two different objects, in C and not in C , respectively, into the positive and boundary regions. From Figure 1, one can easily interpret the two ratios involved.

Based on the ratios in the condition (c1'), we now re-express α and β as:

$$\begin{aligned} \alpha &= \left(1 + \frac{\lambda_{(B-P)P}}{\lambda_{(P-B)N}} \right)^{-1}, \\ \beta &= \left(1 + \frac{\lambda_{(N-B)P}}{\lambda_{(B-N)N}} \right)^{-1}. \end{aligned} \quad (13)$$

It follows that $1 \geq \alpha > \gamma > \beta \geq 0$. The following simplified rules are obtained after tie-breaking [36]:

- (P1) If $Pr(C|[x]) \geq \alpha$, decide $x \in \text{POS}(C)$;
- (B1) If $\beta < Pr(C|[x]) < \alpha$, decide $x \in \text{BND}(C)$;
- (N1) If $Pr(C|[x]) \leq \beta$, decide $x \in \text{NEG}(C)$.

The parameter γ is no longer needed.

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

$$\begin{aligned} \text{POS}_{(\alpha, \beta)}(C) &= \{x \in U \mid \text{Pr}(C|[x]) \geq \alpha\}, \\ \text{BND}_{(\alpha, \beta)}(C) &= \{x \in U \mid \beta < \text{Pr}(C|[x]) < \alpha\}, \\ \text{NEG}_{(\alpha, \beta)}(C) &= \{x \in U \mid \text{Pr}(C|[x]) \leq \beta\}. \end{aligned} \tag{14}$$

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

$$\begin{aligned} \underline{\text{apr}}_{(\alpha, \beta)}(C) &= \{x \in U \mid \text{Pr}(C|[x]) \geq \alpha\}, \\ \overline{\text{apr}}_{(\alpha, \beta)}(C) &= \{x \in U \mid \text{Pr}(C|[x]) > \beta\}. \end{aligned} \tag{15}$$

A more detailed derivation of existing probabilistic models can be found in other papers [36,37]. A game-theoretic risk analysis of the required parameters in the decision-theoretic rough set models is given by Herhert and Yao [6].

In real applications of the probabilistic rough set models, one may directly supply the parameters α and β based on an intuitive understanding the levels of tolerance for errors [43]. This means that one indeed uses an intermediate result of the decision-theoretic rough set model without an in-depth understanding of its theoretical foundations [36]. It should be pointed out that the uses of ad hoc parameters α and β may largely due to an unawareness of the well-established Bayesian decision procedure. As a final remark, one may find it much easier to give loss functions that can be related to more intuitive terms such as costs, benefits, and risks, than to give abstract threshold values. This is particular true in situations where the costs can be translated into monetary values.

In other papers [36,37], we have shown explicitly that many earlier probabilistic rough set models can be obtained from the decision-theoretic rough set model when additional conditions are imposed on the losses. Recently, Ślęzak [20] examined two additional formulations of the three probabilistic regions. In one formulation, the threshold values α and β are replaced by the probability $\text{Pr}([x])$. By expressing the losses as functions of $\text{Pr}([x])$, one can derive this formulation. In the other formulation, two probabilities, $\text{Pr}([x]|C)$ and $\text{Pr}([x]|C^c)$, are compared to define the probabilistic regions of C . Instead of using a pair of threshold values on the probability $\text{Pr}(C|[x])$, one uses a pair of threshold values on the likelihood ratio, $\text{Pr}([x]|C)/\text{Pr}([x]|C^c)$. Recall that the odds $\text{Pr}(C|[x])/\text{Pr}(C^c|[x])$ is a monotonic increasing transformation of the probability $\text{Pr}(C|[x])$. A threshold value on the probability can be interpreted as another threshold value on the odds. For the positive region, we

have:

$$P(C|[x]) \geq \alpha \iff \frac{Pr(C|[x])}{Pr(C^c|[x])} \geq \alpha', \quad (16)$$

where $\alpha' = \alpha/(1 - \alpha)$. By Bayes theorem, we can re-express odds in terms of likelihood ratio as follows:

$$\frac{Pr(C|[x])}{Pr(C^c|[x])} = \frac{\frac{Pr([x]|C)Pr(C)}{Pr([x])}}{\frac{Pr([x]|C^c)Pr(C^c)}{Pr([x])}} = \frac{Pr([x]|C)}{Pr([x]|C^c)} \frac{Pr(C)}{Pr(C^c)}. \quad (17)$$

Therefore,

$$\begin{aligned} P(C|[x]) \geq \alpha &\iff \frac{Pr(C|[x])}{Pr(C^c|[x])} \geq \alpha' \\ &\iff \frac{Pr([x]|C)}{Pr([x]|C^c)} \frac{Pr(C)}{Pr(C^c)} \geq \alpha' \\ &\iff \frac{Pr([x]|C)}{Pr([x]|C^c)} \geq \frac{Pr(C^c)}{Pr(C)} \alpha' = \alpha'', \end{aligned} \quad (18)$$

where $\alpha'' = (Pr(C^c)\alpha)/(Pr(C)(1 - \alpha))$. Similar expressions can be obtained for the boundary and negative regions. By expressing losses as functions of $P(C)$ and $P(C^c)$, one can derive the formulation from the decision-theoretic rough set model.

4.2 Probabilistic rules

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

$$\begin{aligned} \text{Des}([x]) &\longrightarrow_P \text{Des}(C), \quad \text{for } [x] \subseteq \text{POS}_{(\alpha,\beta)}(C); \\ \text{Des}([x]) &\longrightarrow_B \text{Des}(C), \quad \text{for } [x] \subseteq \text{BND}_{(\alpha,\beta)}(C); \\ \text{Des}([x]) &\longrightarrow_N \text{Des}(C), \quad \text{for } [x] \subseteq \text{NEG}_{(\alpha,\beta)}(C). \end{aligned}$$

Recall that the conditional probability $c = Pr(C|[x])$ 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 C as a member of C is defined by $Pr(C^c|[x]) = 1 - Pr(C|[x]) = 1 - c$ and is below $1 - \alpha$. For negative rules, the error rate of rejecting a member of

C as a non-member of C is given by $Pr(C|[x]) = c$ and is below β . When the conditional probability is too low (i.e., below threshold α) for acceptance and at the same time too high (i.e., above the threshold β) for rejection, we choose a boundary rule for an abstained decision, an indecision or a delayed decision. The error rates for putting an object in C and an object not in C into the boundary region are $Pr(C|[x]) = c$ and $Pr(C^c|[x]) = 1 - Pr(C|[x]) = 1 - c$, respectively.

The semantics of the three types of rules can be easily explained by their associated different costs:

$$\begin{aligned}
\text{positive rule : } & c * \lambda_{PP} + (1 - c) * \lambda_{PN}, \\
\text{boundary rule : } & c * \lambda_{BP} + (1 - c) * \lambda_{BN}, \\
\text{negative rule : } & c * \lambda_{NP} + (1 - c) * \lambda_{NN},
\end{aligned} \tag{19}$$

where $c = Pr(C|[x])$ for rule $Des([x]) \longrightarrow_{\Lambda} Des(C)$, $\Lambda \in \{P, B, N\}$. Consider the special case where we assume zero cost for a correct classification, namely, $\lambda_{PP} = \lambda_{NN} = 0$. The costs of rules can be simplified into:

$$\begin{aligned}
\text{positive rule : } & (1 - c) * \lambda_{PN}, \\
\text{boundary rule : } & c * \lambda_{BP} + (1 - c) * \lambda_{BN}, \\
\text{negative rule : } & c * \lambda_{NP}.
\end{aligned} \tag{20}$$

In this case, they are only related to the misclassification error rates and therefore 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:

$$\begin{aligned}
\text{positive rule : } & (1 - c), \\
\text{boundary rule : } & c * \lambda_{BP} + (1 - c) * \lambda_{BN}, \\
\text{negative rule : } & c.
\end{aligned} \tag{21}$$

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.

5 Rules for Two-Category Classification

Based on the discussion of the approximations of a single concept, one may study the approximations of many concepts at the same time. Consider first the simple case of a two-category classification problem, where C denotes one

class and the C^c denotes the other class. In the classical rough set theory, we have:

$$\begin{aligned} \text{POS}(C) &= \text{NEG}(C^c), \\ \text{BND}(C) &= \text{BND}(C^c), \\ \text{NEG}(C) &= \text{POS}(C^c). \end{aligned} \tag{22}$$

It follows that the positive, boundary, and negative rules of C correspond, respectively, to the negative, boundary, and positive rules of C^c , and vice versa. Thus, it is sufficient to study only the approximations of C without explicitly considering the approximations of C^c . Unfortunately, these equalities do not necessarily hold for probabilistic approximations. For two-category classification, we must consider the approximations of both classes in three separate cases.

5.1 Probabilistic rules for classification

For a two-category classification problem $\mathcal{C} = \{C, C^c\}$, in general we can use two loss functions, one for C and the other for C^c . For easy analysis, we assume that the same loss function is used for both C and C^c . Then we have two sets of probabilistic rules, the set for C ,

$$\begin{aligned} (\text{P}^+) \quad \text{Des}([x]) &\longrightarrow_P \text{Des}(C), \quad \text{for } [x] \subseteq \text{POS}_{(\alpha,\beta)}(C); \\ (\text{B}^+) \quad \text{Des}([x]) &\longrightarrow_B \text{Des}(C), \quad \text{for } [x] \subseteq \text{BND}_{(\alpha,\beta)}(C); \\ (\text{N}^+) \quad \text{Des}([x]) &\longrightarrow_N \text{Des}(C), \quad \text{for } [x] \subseteq \text{NEG}_{(\alpha,\beta)}(C). \end{aligned}$$

and the set for C^c ,

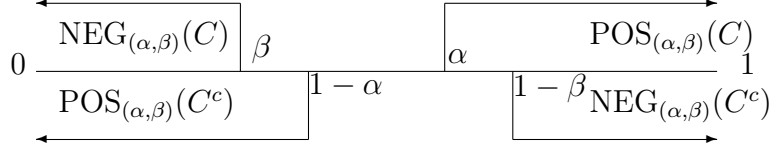
$$\begin{aligned} (\text{P}^-) \quad \text{Des}([x]) &\longrightarrow_P \text{Des}(C^c), \quad \text{for } [x] \subseteq \text{POS}_{(\alpha,\beta)}(C^c); \\ (\text{B}^-) \quad \text{Des}([x]) &\longrightarrow_B \text{Des}(C^c), \quad \text{for } [x] \subseteq \text{BND}_{(\alpha,\beta)}(C^c); \\ (\text{N}^-) \quad \text{Des}([x]) &\longrightarrow_N \text{Des}(C^c), \quad \text{for } [x] \subseteq \text{NEG}_{(\alpha,\beta)}(C^c). \end{aligned}$$

The positive rules (P^+) and (P^-) are used to accept an object to be an instance of C and C^c , respectively; the negative rules (P^+) and (P^-) are used to reject an object to be an instance of C and C^c , respectively; the boundary rules represent an abstained decision.

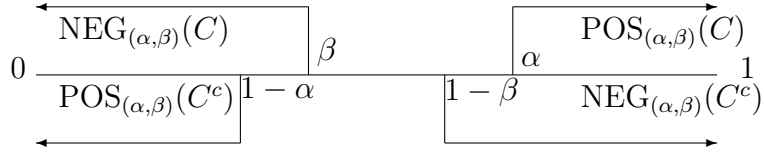
Unlike rules in the classical rough set theory, the acceptance (rejection) of an object to be an instance of C does not imply the rejection (acceptance) of the object to be an instance of C^c , and vice versa. For example, it is possible that we may accept an object to be an instance of C and at the same time do not



(a). Model 1: $\alpha > \beta$ and $\alpha + \beta = 1$



(b). Model 2: $\alpha > \beta$ and $\alpha + \beta < 1$



(c). Model 3: $\alpha > \beta$ and $\alpha + \beta > 1$

Fig. 2. Three models of two-category classification

reject the object to be an instance of C^c . Our three-way decisions for two-category classification depend on the actual values of α and β . Three models are identified and examined next.

5.2 Model 1: $\alpha > \beta$ and $\alpha + \beta = 1$

The model with $\alpha + \beta = 1$ is illustrated by Figure 2(a) based on the probability values $Pr(C|[x])$. If a loss function satisfies (c0) and (c1), we have $\alpha > \beta$. Thus, $\beta < 0.5$ and $\alpha > 0.5$ hold in Model 1.

From the fact that $\alpha = 1 - \beta$ and $\beta = 1 - \alpha$, it follows that

$$\begin{aligned} Pr(C|[x]) \geq \alpha &\iff Pr(C^c|[x]) \leq 1 - \alpha \\ &\iff Pr(C^c|[x]) \leq \beta; \end{aligned}$$

$$\begin{aligned}
Pr(C|[x]) \leq \beta &\iff Pr(C^c|[x]) \geq 1 - \beta \\
&\iff Pr(C^c|[x]) \geq \alpha; \\
\beta < Pr(C|[x]) < \alpha &\iff 1 - \alpha < Pr(C^c|[x]) < 1 - \beta \\
&\iff \beta < Pr(C^c|[x]) < \alpha.
\end{aligned} \tag{23}$$

By the definition of the probabilistic positive, boundary and negative regions, we have:

$$\begin{aligned}
POS_{(\alpha,\beta)}(C) &= NEG_{(\alpha,\beta)}(C^c), \\
BND_{(\alpha,\beta)}(C) &= BND_{(\alpha,\beta)}(C^c), \\
NEG_{(\alpha,\beta)}(C) &= POS_{(\alpha,\beta)}(C^c).
\end{aligned} \tag{24}$$

These relationships are shown in Figure 2(a). They may be viewed as a counterpart of Equation (22) of the classical rough set model.

For the model with $\alpha + \beta = 1$, the following double implications hold between the decisions of acceptance and rejection:

$$\begin{aligned}
&P^+ : \textit{accept} \text{ an object for } C \text{ with confidence } Pr(C|[x]) \geq \alpha \\
&\iff N^- : \textit{reject} \text{ an object for } C^c \text{ with confidence } Pr(C^c|[x]) \leq \beta = 1 - \alpha; \\
&N^+ : \textit{reject} \text{ an object for } C \text{ with confidence } Pr(C|[x]) \leq \beta \\
&\iff P^- : \textit{accept} \text{ an object for } C^c \text{ with confidence } Pr(C^c|[x]) \geq \alpha = 1 - \beta;
\end{aligned}$$

That is, for any object the acceptance of C is equivalent to the rejection of C^c and the rejection of C is equivalent to the acceptance of C^c , and vice versa. Furthermore, the abstained decisions are the same for both C and C^c . Thus, this model may be viewed as a symmetric model.

To derive Model 1, a loss function must satisfy (c0), (c1) and the condition:

$$(c2) \quad \frac{\lambda_{(P-B)N}}{\lambda_{(N-B)P}} = \frac{\lambda_{(B-P)P}}{\lambda_{(B-N)N}}. \tag{25}$$

The condition is again represented in terms of ratios of increases of losses, and can be interpreted based on Figure 1. For a symmetric model, it is expected that the loss function is in some sense symmetric with respect to C and C^c . Consider a loss function satisfying the following additional symmetric conditions:

$$\begin{aligned}
\lambda_{PP} &= \lambda_{NN}, \\
\lambda_{BP} &= \lambda_{BN}, \\
\lambda_{NP} &= \lambda_{PN}.
\end{aligned} \tag{26}$$

Under these conditions, both ratios in (c2) become 1 and hence the loss function satisfies (c2). In other words, the symmetric conditions in Equation (26) characterize a family of loss functions for deriving Model 1. Given conditions (c0) and (c1), condition (c2) is both necessary and sufficient for deriving Model 1, and conditions given by Equation (26) are sufficient but not necessary.

The classical rough set model is defined by $\alpha = 1$ and $\beta = 0$, which satisfies the condition $\alpha + \beta = 1$. Thus, it is a special case of Model 1. From equation (9), it following that given (c0), the following conditions are both necessary and sufficient for the classical rough set model [37]:

$$\lambda_{BP} = \lambda_{PP}, \quad \lambda_{BN} = \lambda_{NN}. \quad (27)$$

A loss function satisfying the above condition satisfies condition (c2) in its product form.

5.3 Model 2: $\alpha > \beta$ and $\alpha + \beta < 1$

Model 2 is defined by the condition $\alpha + \beta < 1$ and is illustrated by Figure 2(b). By the condition $\alpha > \beta$, $\beta < 0.5$ holds for Model 2.

To derive Model 2, a loss function must satisfy the condition:

$$(c3) \quad \frac{\lambda_{(P-B)N}}{\lambda_{(N-B)P}} < \frac{\lambda_{(B-P)P}}{\lambda_{(B-N)N}}. \quad (28)$$

For Model 2, we no longer have the equivalence of accepting C and rejecting C^c , nor the equivalence of rejecting C and accepting C^c . Instead, as shown by Figure 2(b), we only have the following single implications:

$$\begin{aligned} & N^+ : \text{reject an object for } C \text{ with confidence } Pr(C|[x]) \leq \beta \\ \implies & P^- : \text{accept an object for } C^c \text{ with confidence } Pr(C^c|[x]) \geq \alpha; \\ & N^- : \text{reject an object for } C^c \text{ with confidence } Pr(C^c|[x]) \leq \beta \\ \implies & P^+ : \text{accept an object for } C \text{ with confidence } Pr(C|[x]) \geq \alpha. \end{aligned}$$

That is, from rejection of one class we can conclude the acceptance of the other class. However, as shown in Figure 2(b), the reverse of the implications does not hold.

5.4 Model 3: $\alpha > \beta$ and $\alpha + \beta > 1$

Model 3 may be viewed as a mirror model of Model 2. It is defined by the condition $\alpha + \beta > 1$ and is illustrated by Figure 2(c). By the condition $\alpha > \beta$, $\alpha > 0.5$ holds for Model 3.

To derive Model 3, a loss function must satisfy the condition:

$$(c4) \quad \frac{\lambda_{(P-B)N}}{\lambda_{(N-B)P}} > \frac{\lambda_{(B-P)P}}{\lambda_{(B-N)N}}. \quad (29)$$

As shown by Figure 2(c), the following mirror implications of Model 2 hold in Model 3:

$$\begin{aligned} & P^- : \text{accept an object for } C^c \text{ with confidence } Pr(C^c|[x]) \geq \alpha \\ \implies & N^+ : \text{reject an object for } C \text{ with confidence } Pr(C|[x]) \leq \beta; \\ & P^+ : \text{accept an object for } C \text{ with confidence } Pr(C|[x]) \geq \alpha \\ \implies & N^- : \text{reject an object for } C^c \text{ with confidence } Pr(C^c|[x]) \leq \beta. \end{aligned}$$

From acceptance of one class we can conclude the rejection of the other class. The reverse of the implications does not hold.

5.5 Many-category classification

The notion of three-way decision rules can be extended from the two-category classification into many-category classification. There are several ways to achieve this goal. In this section, we only briefly comment on a few issues regarding many-category classification.

Ślęzak [20] suggests an approach for defining the three probabilistic regions based on pair-wise comparisons of categories. A matrix of threshold values is used, with a pair of threshold values on the likelihood ratio of each pair of categories. Although the approach is mathematically appealing and sound, one may have difficulties in estimating all threshold values. One may simplify the model by using the same pair of threshold values for all pairs of categories.

An alternative method is to change an m -category classification problem into m two-category classification problems [42]. Consider an m -category classification problem $\mathcal{C} = \{C_1, C_2, \dots, C_m\}$, where C_i 's form a family of pair-wise disjoint subset of U , namely, $C_i \cap C_j = \emptyset$ for $i \neq j$, and $\cup C_i = U$. For each C_i , we can define a two-category classification by $\{C, C^c\}$, where $C = C_i$

and $C^c = C_i^c = \cup_{i \neq j} C_j$. The results from two-category classification can be immediately applied.

In general, one may use different losses for different categories. The results are a model with different threshold values for different categories, which is indeed used by Ślęzak [20]. Mathematically, the formulation is not much complicated, but the resulting model is more general and flexible. Again, there are practical difficulties in estimating the required parameters. In building a model, one needs to strike the right balance between the simplicity and flexibility, and between mathematical generality and practical applicability, of a model. A simple model is not always an inferior practical model.

Depending on the values of α and β , an equivalence class may produce more than one positive rule. For $\alpha \geq 0.5$, each equivalence class produces at most one positive rule. Similarly, an equivalence class may produce several boundary rules and several negative rules. For $\beta > 0.5$, each equivalence class produces at most one boundary rule. In general, one has to consider the problem of rule conflict resolution in order to make effective acceptance, rejection, and abstaining decisions.

For classification problems with more than two classes, one may not be interested in negative rules. The objects in the negative region of one class may be in the positive and boundary regions of other classes. For this reason, only positive and boundary rules were considered in some earlier studies [36,42]. In general, it is necessary to consider negative rules. The notion of three-way decisions seems to be more accurate and appropriate for interpreting rules in rough set theory.

It is necessary to have a further study on the probabilistic positive, boundary, and negative regions of a classification, as well as the associated rules. In the classical rough set model, these three regions form a partition of the universe. This is not necessarily true in the probabilistic models [36,42]. One needs to impose conditions on the losses in the decision-theoretic set model [40,41], or on the threshold values on the Bayesian rough set model [20], in order to derive pair-wise disjoint regions. Such a subtle difference needs to be carefully examined.

6 Conclusion

By observing a difficulty in interpreting rules in probabilistic rough set models, we introduce the notion of three-way decision rules. Positive, negative, and boundary rules are constructed from the corresponding regions. They represent the results of a three-way decision of acceptance, rejection, or abstaining. Sim-

ilar to the ideas of hypothesis testing in statistics, the decisions of acceptance and rejection are made with certain levels of tolerance for errors. These levels of tolerance are determined systematically based on the well-known Bayesian decision procedure with respect to losses associated with various decisions.

The framework of three-way decision rules is applied to two-category classification. Three models, one symmetric and two non-symmetric, are derived. In the symmetric model, the acceptance of one class is equivalent to the rejection of the other class, and vice versa. It is no longer true in the non-symmetric models. The same ideas of three-way decision rules can be extended to many-category classification.

With probabilistic approximations, we tolerate certain levels of erroneous decisions. This offers a new view for rule induction, which is absent from the classical rough set model. However, some existing studies straightforwardly adopt, or slightly generalize, the notions from the classical rough model without considering the new semantics offered by the probabilistic rough set models. For example, some studies use the sizes of probabilistic regions produced by a subset of attributes and the entire set of attributes, respectively, in the process of attribute reduction [43]. Such a practice is not entirely consistent with the ideas of tolerance of erroneous decisions that motivate probabilistic rough set models [42]. There does not exist a convincing argument to support the claim that an attribute reduct should keep the same probabilistic positive region, except the fact that the classical rough set model requires it. Suppose that our tolerance level of errors is 10% and the probabilistic positive region defined by the entire set of attributes produces 5% errors. It may be unnecessary to require that the probabilistic positive region defined by an attribute reduct also produces 5% errors. Therefore, as future research, we need to re-examine and re-interpret notions from the classical rough model in the new probabilistic setting of three-way decisions.

Acknowledgments

The author would like to thank reviewers of the paper for their constructive comments. This work is partially supported by a discovery grant from NSERC Canada.

References

- [1] Bazan, J.G., Nguyen, S.H., Nguyen, H.S. and Skowron, A. Rough set methods in approximation of hierarchical concepts , *Proceedings of RSCTC'04*, LNAI

3066, 346-355, 2004.

- [2] Duda, R.O. and Hart, P.E. *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
- [3] Forster, M.R. Key concepts in model selection: performance and generalizability, *Journal of Mathematical Psychology*, **44**, 205-231, 2000.
- [4] 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.
- [5] Grzymala-Busse, J.W. Knowledge acquisition under uncertainty - a rough set approach, *Journal of intelligent and Robotic Systems*, **1**, 3-16, 1988.
- [6] 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.
- [7] 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.
- [8] Li, Y., Zhang, C. and Swan, J.R. An information filtering model on the Web and its application in JobAgent, *Knowledge-Based Systems*, **13**, 285-296, 2000.
- [9] Lingras, P.J. and Yao, Y.Y. Data mining using extensions of the rough set model, *Journal of the American Society for Information Science*, **49**, 415-422, 1998.
- [10] Lurie, J.D. and Sox, H.C. Principles of medical decision making, *Spine*, **24**, 493-498, 1999.
- [11] Marek, V.W. and Truszczyński, M. Contributions to the theory of rough sets, *Fundamenta Informaticae*, **39**, 389-409, 1999.
- [12] Nguyen, S.H. and Nguyen, H.S. Hierarchical rough classifiers, *Proceedings of RSEISP'07*, LNAI 4585, 40-50, 2007.
- [13] Pauker, S.G. and Kassirer, J.P. The threshold approach to clinical decision making, *The New England Journal of Medicine*, **302**, 1109-1117, 1980.
- [14] Pawlak, Z. Rough sets, *International Journal of Computer and Information Sciences*, **11**, 341-356, 1982.
- [15] Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
- [16] Pawlak, Z. Rough sets, decision algorithms and Bayes' theorem, *European Journal of Operational Research*, **136**, 181-189, 2002.
- [17] Pawlak, Z. and Skowron, A. Rudiments of rough sets, *Information Sciences*, **177**, 3-27, 2007.
- [18] Pawlak, Z. and Skowron, A. Rough sets: some extensions, *Information Sciences*, **177**, 28-40, 2007.

- [19] 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.
- [20] Ślęzak, D. Rough sets and Bayes factor, *LNCS Transactions on Rough Sets III*, LNCS 3400, 202-229, 2005.
- [21] 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.
- [22] Sowa, J.F. *Conceptual Structures, Information Processing in Mind and Machine*, Addison-Wesley, Reading, Massachusetts, 1984.
- [23] Sox, H.C., Blatt, M.A., Higgins, M.C. and Marton, K.I. *Medical Decision Making*, Butterworth-Heinemann, Stoneham, MA, 1988.
- [24] Suraj, Z. and Grochowalski, P. The rough set database system: an overview, *LNCS Transactions on Rough Sets III*, LNCS 3400, 190-210, 2005.
- [25] Tsumoto, S. Accuracy and coverage in rough set rule induction, *Proceedings of RSTC'02*, LNAI 2475, 373-380, 2002.
- [26] Van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (Eds.) *Categories and Concepts, Theoretical Views and Inductive Data Analysis*, Academic Press, New York, 1993.
- [27] Wald, A. Sequential tests of statistical hypotheses, *The Annals of Mathematical Statistics*, **16**, 117-186, 1945.
- [28] Wei, L.L. and Zhang, W.X. Probabilistic rough sets characterized by fuzzy sets, *Proceedings of RSFDGrC'03*, LNAI 2639, 173-180, 2003.
- [29] Weller, A.C. *Editorial Peer Review: Its Strengths and Weaknesses*, Information Today, Inc., Medford, NJ, 2001.
- [30] Wong, S.K.M. and Ziarko, W. Algorithm for inductive learning, *Bulletin of the Polish Academy of Sciences, Technical Sciences*, **34**, 271-276, 1986.
- [31] Woodward, P.W. and Naylor, J.C. An application of Bayesian methods in SPC, *The Statistician*, **42**, 461-469, 1993.
- [32] 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.
- [33] Yao, J.T. and Herbert, J.P. Web-based support systems based on rough set analysis, *Proceedings of RSEISP'07*, LNAI 4585, 360-370, 2007.
- [34] Yao, Y.Y. Probabilistic approaches to rough sets, *Expert Systems*, **20**, 287-297, 2003.
- [35] Yao, Y.Y. A note on definability and approximations, *LNCS Transactions on Rough Sets VII*, LNCS 4400, 274-282, 2007.

- [36] Yao, Y.Y. Decision-theoretic rough set models, *Proceedings of RSKT'07*, LNAI 4481, 1-12, 2007.
- [37] Yao, Y.Y. Probabilistic rough set approximations, *International Journal of Approximation Reasoning*, **49**, 255-271, 2008.
- [38] Yao, Y.Y. Interpreting concept learning in cognitive informatics and granular computing, *IEEE Transactions on System, Man and Cybernetics, B*, **39**, 855-866, 2009.
- [39] Yao, Y.Y. Three-way decision: an interpretation of rules in rough set theory, *Proceedings of RSKT'09*, LNAI 5589, 642-649, 2009.
- [40] 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.
- [41] Yao, Y.Y., Wong, S.K.M. and Lingras, P. A decision-theoretic rough set model, in: *Methodologies for Intelligent Systems 5*, Z.W. Ras, M. Zemankova and M.L. Emrich (Eds.), New York, North-Holland, 17-24, 1990.
- [42] Yao, Y.Y. and Zhao, Y., Attribute reduction in decision-theoretic rough set models, *Information Sciences*, **178**, 3356-3373, 2008.
- [43] Ziarko, W. Variable precision rough sets model, *Journal of Computer and Systems Sciences*, **46**, 39-59, 1993.