Yao, Y.Y. Decision-theoretic rough set models,
Rough Sets and Knowledge Technology, Second International Conference, RSKT 2007, Proceedings, LNAI 4481, pp. 1-12, 2007.

Decision-Theoretic Rough Set Models

Yiyu Yao

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

Abstract. Decision-theoretic rough set models are a probabilistic extension of the algebraic rough set model. The required parameters for defining probabilistic lower and upper approximations are calculated based on more familiar notions of costs (risks) through the well-known Bayesian decision procedure. We review and revisit the decision-theoretic models and present new results. It is shown that we need to consider additional issues in probabilistic rough set models.

Keywords. Bayesian decision theory, decision-theoretic rough sets, probabilistic rough sets, variable precision rough sets

1 Introduction

Ever since the introduction of rough set theory by Pawlak in 1982 [10, 11, 13], many proposals have been made to incorporate probabilistic approaches into the theory [14, 29–31]. They include, for example, rough set based probabilistic classification [24], 0.5 probabilistic rough set model [15], decision-theoretic rough set models [32, 33], variable precision rough set models [6, 35], rough membership functions [12], parameterized rough set models [14, 17], and Bayesian rough set models [2, 18, 19]. The results of these studies increase our understanding of the rough set theory and its domain of applications.

The decision-theoretic rough set models [29, 32, 33] and the variable precision rough set models [6, 35, 36] were proposed in the early 1990's. The two models are formulated differently in order to generalize the 0.5 probabilistic rough set model [15]. In fact, they produce the same rough set approximations [29, 30]. Their main differences lie in their respective treatment of the required parameters used in defining the lower and upper probabilistic approximations. The decision-theoretic models systematically calculate the parameters based on a loss function through the Bayesian decision procedure. The physical meaning of the loss function can be interpreted based on more practical notions of costs and risks. In contrast, the variable precision models regard the parameters as primitive notions and a user must supply those parameters. A lack of a systematic method for parameter estimation has led researchers to use many ad hoc methods based on trial and error.

The results and ideas of the decision-theoretic model, based on the well established and semantically sound Bayesian decision procedure, have been successfully applied to many fields, such as data analysis and data mining [3, 7, 16, 21–23, 25, 34], information retrieval [8, 20], feature selection [27], web-based support systems [26], and intelligent agents [9]. Some authors have generalized the decision-theoretic model to multiple regions [1].

The main objective of this paper is to revisit the decision-theoretic rough set model and to present new results. In order to appreciate the generality and flexibility of the model, we show explicitly the conditions on the loss function under which other models can be derived, including the standard rough set model, 0.5 probabilistic rough set model, and both symmetric and asymmetric variable precision rough set models. Furthermore, the decision-theoretic model is extended from a two-class classification problem into a many-class classification problem. This enables us to observe that some of the straightforward generalizations of notions and measures of the algebraic rough set model may not necessarily be meaningful in the probabilistic models.

2 Algebraic Rough Set Approximations

Let U be a finite and nonempty set and E an equivalence relation on U. The pair $apr = (U, E)$ is called an approximation space [10, 11]. The equivalence relation E induces a partition of U, denoted by U/E. The equivalence class containing x is given by $[x] = \{y \mid xEy\}$.

The equivalence classes of E are the basic building blocks to construct algebraic rough set approximations. For a subset $A \subseteq U$, its lower and upper approximations are defined by [10, 11]:

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

The lower and upper approximations, $\underline{apr}, \overline{apr} : 2^U \longrightarrow 2^U$, can be interpreted as a pair of unary set-theoretic operators [28]. 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 set complement of A. Other equivalent definitions and additional properties of approximation operators can be found in [10, 11, 28].

Based on the rough set approximations of A, one can divide the universe U into three disjoint regions, 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 or not an element $x \in \mathrm{BND}(A)$ belongs to A.

We can easily extend the concepts of rough set approximations and regions of a single set to a partition of the universe. Consider first a simple case. A set

$\emptyset \neq A \neq U$ induces a partition $\pi_A = \{A, A^c\}$ of U. The approximations of the partition π_A are defined by:

$$\begin{aligned}\underline{apr}(\pi_A) &= (\underline{apr}(A), \underline{apr}(A^c)) = (\underline{apr}(A), (\overline{apr}(A))^c);\\ \overline{apr}(\pi_A) &= (\overline{apr}(A), \overline{apr}(A^c)) = (\overline{apr}(A), (\underline{apr}(A))^c).\end{aligned} \tag{3}$$

Based on the positive, boundary and negative regions of A and A^c, we have the corresponding three disjoint regions of π_A:

$$\begin{aligned}\mathrm{POS}(\pi_A) &= \mathrm{POS}(A) \cup \mathrm{POS}(A^c),\\ \mathrm{BND}(\pi_A) &= \mathrm{BND}(A) \cup \mathrm{BND}(A^c) = \mathrm{BND}(A) = \mathrm{BND}(A^c),\\ \mathrm{NEG}(\pi_A) &= U - \mathrm{POS}(\pi_A) \cup \mathrm{BND}(\pi_A) = \emptyset.\end{aligned} \tag{4}$$

In general, let $\pi = \{A_1, A_2, \ldots, A_m\}$ be a partition of the universe U. We have:

$$\begin{aligned}\underline{apr}(\pi) &= (\underline{apr}(A_1), \underline{apr}(A_2), \ldots, \underline{apr}(A_m));\\ \overline{apr}(\pi) &= (\overline{apr}(A_1), \overline{apr}(A_2), \ldots, \overline{apr}(A_m)).\end{aligned} \tag{5}$$

We can extend the notions of three regions to the case of a partition:

$$\begin{aligned}\mathrm{POS}(\pi) &= \bigcup_{1 \leq i \leq m} \mathrm{POS}(A_i),\\ \mathrm{BND}(\pi) &= \bigcup_{1 \leq i \leq m} \mathrm{BND}(A_i),\\ \mathrm{NEG}(\pi) &= U - \mathrm{POS}(\pi) \cup \mathrm{BND}(\pi) = \emptyset.\end{aligned} \tag{6}$$

It can be verified that $\mathrm{POS}(\pi) \cap \mathrm{BND}(\pi) = \emptyset$ and $\mathrm{POS}(\pi) \cup \mathrm{BND}(\pi) = U$.

The positive and boundary regions can be used to derive two kinds of rules, namely, certain and probabilistic rules, or deterministic and non-deterministic rules [5, 11]. More specifically, for $[x] \subseteq \mathrm{POS}(A_i)$ and $[x'] \subseteq \mathrm{BND}(A_i)$, we have the two kinds of rules, respectively, as follows:

$$[x] \xrightarrow{c=1} A_i, \qquad [x'] \xrightarrow{0<c<1} A_i, \tag{7}$$

where the confidence measure c of a rule is defined by:

$$c = \frac{|A_i \cap [x]|}{|[x]|}, \tag{8}$$

and $|\cdot|$ is the cardinality of a set. It can be verified that $\underline{apr}(A_i) \cap \underline{apr}(A_j) = \emptyset$ for $i \neq j$. An element of U belongs to at most one positive region of A_i's. On the other hand, an element may be in more than one boundary region. Thus, an element may satisfy at most one certain rule, or satisfy more than one probabilistic rule.

To quantify the degree of dependency of the partition π on the partition U/E, many authors use only the positive region. For example, Pawlak suggests the following measure [11]:

$$r(\pi | U/E) = \frac{|\mathrm{POS}(\pi)|}{|U|}. \tag{9}$$

Yao, Y.Y. Decision-theoretic rough set models,
Rough Sets and Knowledge Technology, Second International Conference, RSKT 2007, Proceedings, LNAI 4481, pp. 1-12, 2007.

This measure only considers the coverage of certain rules. The effects of uncertain rules are not considered. Since $\mathrm{NEG}(\pi) = \emptyset$, $\mathrm{POS}(\pi) \cap \mathrm{BND}(\pi) = \emptyset$, and $\mathrm{POS}(\pi) \cup \mathrm{BND}(\pi) = U$, it may be sufficient to consider only $\mathrm{POS}(\pi)$. As we will show later, when those notions are generalized into a probabilistic version, it is necessary to consider all three regions.

3 Probabilistic Rough Set Approximations

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 [4], and apply the procedure for the construction of probabilistic approximations [32, 33].

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

$$R(a_i|\mathbf{x}) = \sum_{j=1}^{s} \lambda(a_i|w_j) P(w_j|\mathbf{x}). \tag{10}$$

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:

$$\mathbf{R} = \sum_{\mathbf{x}} R(\tau(\mathbf{x})|\mathbf{x}) P(\mathbf{x}), \tag{11}$$

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 (10) 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 Probabilistic rough set approximation operators

In an approximation space $apr = (U, E)$, an equivalence class $[x]$ is considered to be the description of x. The partition U/E is the set of all possible descriptions.

The classification of objects according to approximation operators can be easily fitted into the Bayesian decision-theoretic framework. The set of states is given by $\Omega = \{A, A^c\}$ indicating that an element is in A and not in A, respectively. We use the same symbol to denote both a subset A and the corresponding state. With respect to the three regions, the set of actions is given by $\mathcal{A} = \{a_1, a_2, a_3\}$, where a_1, a_2, and a_3 represent the three actions in classifying an object, deciding POS(A), deciding NEG(A), and deciding BND(A), respectively. Let $\lambda(a_i|A)$ denote the loss incurred for taking action a_i when an object belongs to A, and let $\lambda(a_i|A^c)$ denote the loss incurred for taking the same action when the object does not belong to A.

The probabilities $P(A|[x])$ and $P(A^c|[x])$ are the probabilities that an object in the equivalence class $[x]$ belongs to A and A^c, respectively. The expected loss $R(a_i|[x])$ associated with taking the individual actions can be expressed as:

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

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

(P) If $R(a_1|[x]) \leq R(a_2|[x])$ and $R(a_1|[x]) \leq R(a_3|[x])$, decide POS($A$);
(N) If $R(a_2|[x]) \leq R(a_1|[x])$ and $R(a_2|[x]) \leq R(a_3|[x])$, decide NEG($A$);
(B) If $R(a_3|[x]) \leq R(a_1|[x])$ and $R(a_3|[x]) \leq R(a_2|[x])$, decide BND($A$).

Tie-breaking criteria should be added so that each element is classified into only one region. Since $P(A|[x]) + P(A^c|[x]) = 1$, we can simplify the rules to classify any object in $[x]$ based only on the probabilities $P(A|[x])$ and the loss function λ_{ij} ($i = 1, 2, 3$; $j = 1, 2$).

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

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

where

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

$$\beta = \frac{\lambda_{32} - \lambda_{22}}{(\lambda_{21} - \lambda_{22}) - (\lambda_{31} - \lambda_{32})}. \tag{13}$$

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

For a loss function with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, more results about the required parameters α, β and γ are summarized as follows [29]:

1. If a loss function satisfies the condition:
$$(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \tag{14}$$
then $\alpha \geq \gamma \geq \beta$.
2. If a loss function satisfies the condition:
$$\lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \tag{15}$$
then $\alpha \geq 0.5$.
3. If a loss function satisfies the conditions,
$$\begin{aligned} &\lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \\ &(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \end{aligned} \tag{16}$$
then $\alpha \geq 0.5$ and $\alpha \geq \beta$.
4. If a loss function satisfies the condition:
$$(\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}), \tag{17}$$
then $\beta = 1 - \alpha$.
5. If a loss function satisfies the two sets of equivalent conditions,
$$\text{(i).} \quad \begin{aligned} &(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) \geq (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}), \\ &(\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}); \end{aligned} \tag{18}$$
$$\text{(ii).} \quad \begin{aligned} &\lambda_{12} - \lambda_{32} \geq \lambda_{31} - \lambda_{11}, \\ &(\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}); \end{aligned} \tag{19}$$
then $\alpha = 1 - \beta \geq 0.5$.

The condition of Case 1 guarantees that the probabilistic lower approximation of a set is a subset of its probabilistic upper approximation. The condition of Case 2 ensures that a lower approximation of A consists of those elements whose majority equivalent elements are in A. The condition of Case 4 results in a pair of dual lower and upper approximation operators. Case 3 is a combination of Cases 1 and 2. Case 5 is the combination of Cases 1 and 4 or the combination of Cases 3 and 4.

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

$$\begin{aligned} &\text{(P1)} \quad \text{If } P(A|[x]) \geq \alpha, \text{ decide POS}(A); \\ &\text{(N1)} \quad \text{If } P(A|[x]) \leq \beta, \text{ decide NEG}(A); \\ &\text{(B1)} \quad \text{If } \beta < P(A|[x]) < \alpha, \text{ decide BND}(A). \end{aligned}$$

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

$$\begin{array}{ll} \text{(P2)} & \text{If } P(A|[x]) > \alpha, \text{ decide POS}(A); \\ \text{(N2)} & \text{If } P(A|[x]) < \alpha, \text{ decide NEG}(A); \\ \text{(B2)} & \text{If } P(A|[x]) = \alpha, \text{ decide BND}(A). \end{array}$$

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

3.3 Derivations of other probabilistic models

Based on the general decision-theoretic rough set model, it is possible to construct specific models by considering various classes of loss functions. In fact, many existing models can be explicitly derived.

The standard rough set model [15, 24] Consider a loss function:

$$\lambda_{12} = \lambda_{21} = 1, \quad \lambda_{11} = \lambda_{22} = \lambda_{31} = \lambda_{32} = 0. \tag{20}$$

There is a unit cost if an object in A^c is classified into the positive region or if an object in A is classified into the negative region; otherwise there is no cost. From equation (13), we have $\alpha = 1 > \beta = 0$, $\alpha = 1 - \beta$, and $\gamma = 0.5$. From decision rules (P1)-(B1), we can compute the approximations as $\underline{apr}_{(1,0)}(A) = \text{POS}_{(1,0)}(A)$ and $\overline{apr}_{(1,0)}(A) = \text{POS}_{(1,0)}(A) \cup \text{BND}_{(1,0)}(A)$. For clarity, we use the subscript $(1, 0)$ to indicate the parameters used to define lower and upper approximations. The standard rough set approximations are obtained as [15, 24]:

$$\begin{aligned} \underline{apr}_{(1,0)}(A) &= \{x \in U \mid P(A|[x]) = 1\}, \\ \overline{apr}_{(1,0)}(A) &= \{x \in U \mid P(A|[x]) > 0\}, \end{aligned} \tag{21}$$

where $P(A|[x]) = |A \cap [x]|/|[x]|$.

The 0.5 probabilistic model [15] Consider a loss function:

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

A unit cost is incurred if an object in A^c is classified into the positive region or an object in A is classified into the negative region; half of a unit cost is incurred if any object is classified into the boundary region. For other cases, there is no cost. By substituting these λ_{ij}'s into equation (13), we obtain $\alpha = \beta = \gamma = 0.5$. By using decision rules (P2)-(B2), we obtain the 0.5 probabilistic approximations [15]:

$$\begin{aligned} \underline{apr}_{(0.5,0.5)}(A) &= \{x \in U \mid P(A|[x]) > 0.5\}, \\ \overline{apr}_{(0.5,0.5)}(A) &= \{x \in U \mid P(A|[x]) \geq 0.5\}. \end{aligned} \tag{23}$$

The 0.5 model corresponds to the application of the simple majority rule.

The symmetric variable precision rough set model [35] The symmetric variable precision rough set model corresponds to Case 5 of the decision-theoretic model discussed in the last subsection. As suggested by many authors [17, 35], the value of α should be in the range $(0.5, 1]$. This condition can be satisfied by a loss function with $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$ and the condition:

$$\lambda_{12} - \lambda_{32} > \lambda_{31} - \lambda_{11}. \tag{24}$$

The condition of symmetry, i.e., $\beta = 1 - \alpha$, is guaranteed by the additional condition:

$$(\lambda_{12} - \lambda_{32})(\lambda_{32} - \lambda_{22}) = (\lambda_{31} - \lambda_{11})(\lambda_{21} - \lambda_{31}). \tag{25}$$

By decision rules (P1)-(B1), we obtain the probabilistic approximations of the symmetric variable precision rough set model [35]:

$$\begin{aligned} \underline{apr}_{(\alpha,1-\alpha)}(A) &= \{x \in U \mid P(A|[x]) \geq \alpha\}, \\ \overline{apr}_{(\alpha,1-\alpha)}(A) &= \{x \in U \mid P(A|[x]) > 1 - \alpha\}. \end{aligned} \tag{26}$$

They are defined by a single parameter $\alpha \in (0.5, 1]$. The symmetry of parameters α and $\beta = 1 - \alpha$ implies the duality of approximations, i.e., $\underline{apr}_{(\alpha,1-\alpha)}(A) = (\overline{apr}_{(\alpha,1-\alpha)}(A^c))^c$ and $\overline{apr}_{(\alpha,1-\alpha)}(A) = (\underline{apr}_{(\alpha,1-\alpha)}(A^c))^c$.

As an example, consider a loss function:

$$\lambda_{12} = \lambda_{21} = 4, \quad \lambda_{31} = \lambda_{32} = 1, \quad \lambda_{11} = \lambda_{22} = 0. \tag{27}$$

It can be verified that the function satisfies the conditions given by equations (24) and (25). From equation (13), we have $\alpha = 0.75$, $\beta = 0.25$ and $\gamma = 0.5$. By decision rules (P1)-(B1), we have:

$$\begin{aligned} \underline{apr}_{(0.75,0.25)}(A) &= \{x \in U \mid P(A|[x]) \geq 0.75\}, \\ \overline{apr}_{(0.75,0.25)}(A) &= \{x \in U \mid P(A|[x]) > 0.25\}. \end{aligned} \tag{28}$$

In general, higher costs of mis-classification, namely, λ_{12} and λ_{21}, increase the α value [29].

The asymmetric variable precision rough set model [6] The asymmetric variable precision rough set model corresponds to Case 1 of the decision-theoretic model. The condition on the parameters of the model is given by $0 \leq \beta < \alpha \leq 1$. In addition to $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$ and $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, a loss function must satisfy the following conditions:

$$(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) > (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22}). \tag{29}$$

By decision rules (P1)-(B1), we obtain the probabilistic approximations of the asymmetric variable precision rough set model [6]:

$$\begin{aligned} \underline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid P(A|[x]) \geq \alpha\}, \\ \overline{apr}_{(\alpha,\beta)}(A) &= \{x \in U \mid P(A|[x]) > \beta\}. \end{aligned} \tag{30}$$

They are no longer dual operators.

Consider a loss function:

$$\lambda_{12} = 4, \quad \lambda_{21} = 2, \quad \lambda_{31} = \lambda_{32} = 1, \quad \lambda_{11} = \lambda_{22} = 0. \tag{31}$$

The function satisfies the conditions given in equation (29). From equation (13), we have $\alpha = 0.75$, $\beta = 0.50$ and $\gamma = 2/3$. By decision rules (P1)-(B1), we have:

$$\begin{aligned}\underline{apr}_{(0.75,0.50)}(A) &= \{x \in U \mid P(A|[x]) \geq 0.75\},\\ \overline{apr}_{(0.75,0.50)}(A) &= \{x \in U \mid P(A|[x]) > 0.50\}.\end{aligned} \tag{32}$$

They are not dual operators.

4 Probabilistic Approximations of a Partition

The decision-theoretic rough set model examined in the last section is based on the two-class classification problem, namely, classifying an object into either A or A^c. In this section, we extend the formulation to the case of more than two classes.

Let $\pi = \{A_1, A_2, \ldots, A_m\}$ be a partition of the universe U, representing m classes. For this m-class problem, we can solve it in terms of m two-class problems. For example, for the class A_i, we have $A = A_i$ and $A^c = U - A_i = \bigcup_{i \neq j} A_j$. For simplicity, we assume the same loss function for all classes A_i's. Furthermore, we assume that the loss function satisfies the conditions: (i) $\lambda_{11} \leq \lambda_{31} < \lambda_{21}$, (ii) $\lambda_{22} \leq \lambda_{32} < \lambda_{12}$, and (iii) $(\lambda_{12} - \lambda_{32})(\lambda_{21} - \lambda_{31}) > (\lambda_{31} - \lambda_{11})(\lambda_{32} - \lambda_{22})$. It follows that $\alpha > \gamma > \beta$. By decision rules (P1)-(B1), we have the positive, boundary and negative regions:

$$\begin{aligned}\text{POS}_{(\alpha,\beta)}(A_i) &= \{x \in U \mid P(A|[x]) \geq \alpha\},\\ \text{BND}_{(\alpha,\beta)}(A_i) &= \{x \in U \mid \beta < P(A|[x]) < \alpha\},\\ \text{NEG}_{(\alpha,\beta)}(A_i) &= \{x \in U \mid P(A|[x]) \leq \beta\}.\end{aligned} \tag{33}$$

The lower and upper approximations are given by:

$$\begin{aligned}\underline{apr}_{(\alpha,\beta)}(A_i) &= \text{POS}_{(\alpha,\beta)}(A_i) = \{x \in U \mid P(A|[x]) \geq \alpha\},\\ \overline{apr}_{(\alpha,\beta)}(A_i) &= \text{POS}_{(\alpha,\beta)}(A_i) \cup \text{BND}_{(\alpha,\beta)}(A_i) = \{x \in U \mid P(A|[x]) > \beta\}.\end{aligned} \tag{34}$$

Similar to the algebraic case, we can define the approximations of a partition $\pi = \{A_1, A_2, ..., A_m\}$ based on those approximations of equivalence classes of π. For a partition π, the three regions can be defined by:

$$\begin{aligned}\text{POS}_{(\alpha,\beta)}(\pi) &= \bigcup_{1 \leq i \leq m} \text{POS}_{(\alpha,\beta)}(A_i),\\ \text{BND}_{(\alpha,\beta)}(\pi) &= \bigcup_{1 \leq i \leq m} \text{BND}_{(\alpha,\beta)}(A_i),\\ \text{NEG}_{(\alpha,\beta)}(\pi) &= U - \text{POS}_{(\alpha,\beta)}(\pi) \cup \text{BND}_{(\alpha,\beta)}(\pi).\end{aligned} \tag{35}$$

In contrast to the algebraic case, we have the following different properties of the three regions:

1. The three regions are not necessarily pairwise disjoint. Nevertheless, the family $\{\mathrm{POS}_{(\alpha,\beta)}(\pi), \mathrm{BND}_{(\alpha,\beta)}(\pi), \mathrm{NEG}_{(\alpha,\beta)}(\pi)\}$ is a covering of U.
2. The family of positive regions $\{\mathrm{POS}_{(\alpha,\beta)}(A_i) \mid 1 \leq i \leq m\}$ does not necessarily contain pairwise disjoint sets. That is, it may happen that $\mathrm{POS}_{(\alpha,\beta)}(A_i) \cap \mathrm{POS}_{(\alpha,\beta)}(A_j) \neq \emptyset$ for some $i \neq j$.
3. If $\alpha > 0.5$, the family $\{\mathrm{POS}_{(\alpha,\beta)}(A_i) \mid 1 \leq i \leq m\}$ contains pairwise disjoint sets.
4. If $\beta > 0.5$, the three regions are pairwise disjoint. That is, $\{\mathrm{POS}_{(\alpha,\beta)}(\pi), \mathrm{BND}_{(\alpha,\beta)}(\pi), \mathrm{NEG}_{(\alpha,\beta)}(\pi)\}$ is a partition of U. Furthermore, the family boundary regions $\{\mathrm{BND}_{(\alpha,\beta)}(A_i) \mid 1 \leq i \leq m\}$ contains pairwise disjoint sets.
5. If $\beta = 0$, we have $\mathrm{NEG}_{(\alpha,\beta)}(\pi) = \emptyset$.

When generalizing results from the algebraic rough set model, it is necessary to consider the implications of those properties.

The positive and boundary regions give rise to two kinds of rules. For $[x] \subseteq \mathrm{POS}_{(\alpha,\beta)}(A_i)$ and $[x'] \subseteq \mathrm{BND}_{(\alpha,\beta)}(A_i)$, we have:

$$[x] \xrightarrow{c>\alpha} A_i, \qquad [x'] \xrightarrow{\beta<c<\alpha} A_i. \tag{36}$$

Unlike the algebraic rough set model, the probabilistic positive region may also produce non-deterministic rules. The negative region $\mathrm{NEG}_{(\alpha,\beta)}(\pi)$ consists of all those objects that cannot be classified by the above rules.

In the application of probabilistic rough set models, some authors proposed straightforward generalizations of the notions of the algebraic rough set model. For example, the dependency of partition π on U/E is still quantified by using only the probabilistic positive region. The measure used in the variable precision model is [35]:

$$r_{(\alpha,\beta)}(\pi|U/E) = \frac{|\mathrm{POS}_{(\alpha,\beta)}(\pi)|}{|U|}. \tag{37}$$

This may not necessarily be meaningful for the following two reasons. First, the probabilistic positive region may also produce non-deterministic rules. Second, the negative region is no longer empty. It is therefore necessary to consider the impact of the negative region. Similar comments can also be made regarding the generalizations of other measures.

5 Conclusion

A revisit to the decision-theoretic rough set model brings new insights into the probabilistic approaches to rough sets. Different probabilistic models, proposed either before or after the decision-theoretic models, can be easily derived from the decision-theoretic model. More importantly, instead of introducing ad hoc parameters, the Bayesian decision procedure systematically computes the required parameters based on a loss function.

The two-class decision-theoretic model is extended into a many-class model. The results show that some of the straightforward generalizations of the algebraic rough set model may not necessarily be meaningful. From our analysis, it becomes clear that we need to examine new and different measures for the probabilistic rough set models. The decision-theoretic rough set model opens an avenue for future research. A promising direction may be to study various measures based on the loss function within the decision-theoretic model.

References

1. Abd El-Monsef, M.M.E. and Kilany, N.M. Decision analysis via granulation based on general binary relation, *International Journal of Mathematics and Mathematical Sciences*, **2007**, Article ID 12714, 2007.
2. Greco, S., Matarazzo, B. and Slowinski, R. Rough membership and bayesian confirmation measures for parameterized rough sets, *LNAI 3641*, 314-324, 2005.
3. Deogun, J.S., Raghavan, V.V., Sarkar, A. and Sever, H. Data mining: trends in research and development, in: T.Y. Lin and Cercone, N. (Eds.), *Rough Sets and Data Mining*, Klower Academic Publishers, Boston, 9-45, 1997.
4. Duda, R.O. and Hart, P.E. *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
5. Grzymala-Busse, J.W. LERS – a system for learning from examples based on rough sets, in: Slowinski, R. (Ed.), *Intelligent Decision Support: Handbook of Apllications and Advances of the Rough Sets Theory*, Kluwer Academic Publishers, Dordrecht, 3-18, 1992.
6. Katzberg, J.D. and Ziarko, W. Variable precision rough sets with asymmetric bounds, in: Rough Sets, Fuzzy Sets and Knowledge Discovery, Ziarko, W. (Ed), Springer, London, 167-177, 1994.
7. Kitchener, M., Beynon, M. and Harrington, C. Explaining the diffusion of medicaid home care waiver programs using VPRS decision rules, *Health Care Management Science*, **7**, 237244, 2004.
8. Li, Y., Zhang, C. and Swanb, J.R. Rough set based model in information retrieval and filtering, *Proceeding of the 5th International Conference on Information Systems Analysis and Synthesis*, 398-403, 1999.
9. Li, Y., Zhang, C. and Swanb, J.R. An information fltering model on the Web and its application in JobAgent, *Knowledge-Based Systems*, **13**, 285-296, 2000.
10. Pawlak, Z. Rough sets, *International Journal of Computer and Information Sciences*, **11**, 341-356, 1982.
11. Pawlak, Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*, Kluwer Academatic Publishers, Boston, 1991.
12. Pawlak, Z. and Skowron, A. Rough membership functions, in: R.R. Yager and M. Fedrizzi and J. Kacprzyk (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*, John Wiley and Sons, New York, 251-271, 1994.
13. Pawlak, Z. and Skowron, A. Rudiments of rough sets, *Information Sciences*, **177**, 3-27, 2007.
14. Pawlak, Z. and Skowron, A. Rough sets: some extensions, *Information Sciences*, **177**, 28-40, 2007.

15. 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.
16. Qiu, G.F., Zhang, W.X., and Wu, W.Z. Characterizations of attributes in generalized approximation representation spaces, *LNAI 3641*, 84-93, 2005.
17. Skowron, A. and Stepaniuk, J. Tolerance approximation spaces, *Fundamenta Informaticae*, **27**, 245-253, 1996.
18. Slezak, D. Rough sets and Bayes factor, *LNAI 3400*, 202-229, 2005.
19. Slezak, D. and Ziarko, W. Attribute reduction in the Bayesian version of variable precision rough set model, *Electronic Notes in Theoretical Computer Science*, **82**, 263-273, 2003.
20. Srinivasan, P., Ruiz, M.E., Kraft, D.H. and Chen, J. Vocabulary mining for information retrieval: rough sets and fuzzy sets, *Information Processing and Management*, **37**, 15-38, 2001.
21. Tsumoto, S. Accuracy and coverage in rough set rule induction, *LNAI 2475*, 373-380, 2002.
22. Tsumoto, S. Statistical independence from the viewpoint of linear algebra, *LNAI 3488*, 56-64, 2005.
23. Wei, L.L. and Zhang, W.X. Probabilistic rough sets characterized by fuzzy sets, *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, **12**, 47-60, 2004.
24. 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.
25. Wu, W.Z. Upper and lower probabilities of fuzzy events induced by a fuzzy set-valued mapping, *LNAI 3461*, 345-353, 2005.
26. Yao, J.T. and Herbert, J.P. Web-based Support Systems based on Rough Set Analysis, manuscript, 2007.
27. Yao, J.T. and Zhang, M. Feature selection with adjustable criteria, *LNAI 3641*, 204-213, 2005.
28. Yao, Y.Y. Two views of the theory of rough sets in finite universes, *International Journal of Approximation Reasoning*, **15**, 291-317, 1996.
29. Yao, Y.Y. Information granulation and approximation in a decision-theoretical model of rough sets, in: Polkowski, L., Pal, S.K., and Skowron, A. (Eds), *Rough-neuro Computing: Techniques for Computing with Words*, Springer, Berlin, 491-516, 2003.
30. Yao, Y.Y. Probabilistic approaches to rough sets, *Expert Systems*, **20**, 287-297, 2003.
31. Yao, Y.Y. Probabilistic rough set approximations, manuscript, 2006.
32. 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.
33. Yao, Y.Y., Wong, S.K.M. and Lingras, P. A decision-theoretic rough set model, in: Z.W. Ras, M. Zemankova and M.L. Emrich (Eds.), *Methodologies for Intelligent Systems*, **5**, North-Holland, New York, 17-24, 1990.
34. Zhang, W.X., Wu, W.Z., Liang, J.Y. and Li, D.Y. *Rough Set Theory and Methodology* (in Chinese), Xi'an Jiaotong University Press, Xi'an, China, 2001.
35. Ziarko, W. Variable precision rough set model, *Journal of Computer and System Sciences*, **46**, 39-59, 1993.
36. Ziarko, W. Acquisition of hierarchy-structured probabilistic decision tables and rules from data, *Expert Systems*, **20**, 305-310, 2003.
37. Ziarko, W. Probabilistic rough sets, *LNAI 3641*, 283-293, 2005.