A New Formulation of Multi-category Decision-Theoretic Rough Sets

Bing Zhou

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
zhou200b@cs.uregina.ca

Abstract. As a natural extension to the rough set approximations with two decision classes, this paper provides a new formulation of multi-category decision-theoretic rough sets. A three-way decision is added to each class which gives the user the flexibility of making a deferred decision. Different misclassification errors are treated separately with the notion of loss functions from Bayesian decision theory. The losses incurred for making deferred and rejective decisions to each class are also considered. The main objective of this paper is to tackle the limitations of the previous related work, and therefore provide a more complete solution to multi-category decision making.

Keywords: Multi-category classification, probabilistic approximation, three-way decision, Bayesian decision Theory, rough sets.

1 Introduction

In Pawlak rough set theory [5] and its generalized probabilistic approaches [2,6,8,9,12,14,16], all decision classes are treated as the same in the interpretation and applications of approximations and three regions. In other words, the same pair of thresholds are used to defined the positive, negative and boundary regions. As a natural extension to these original studies, rough set approximations for multi-category decision problems using different pairs of thresholds have been discussed in several early literatures.

Yao and Zhao [13] used a method to change an m-category classification problem into m two-category classification problems based on the framework of decision-theoretic rough set model (DTRS) [12]. The results from two-category classification can be immediately applied. The m pairs of thresholds can be systematically calculated based on the well established Bayesian decision theory, and interpreted in terms of more practically operable notions such as cost, risk, benefit etc. Some practice issues when applying Yao and Zhao's idea to classify new objects are further discussed by Liu et al. [4] with the aid from Bayesian decision procedure. However, their work assumes that the losses incurred for misclassifying an object into any substitution classes are the same. This assumption does not always hold in many real world applications [15]. For example, in a

J.T. Yao et al. (Eds): RSKT 2011, LNCS 6954, pp. 514–522, 2011.
© Springer-Verlag Berlin Heidelberg 2011

medical diagnose case, misclassifying an patient who has regular flu to cancer costs more than misclassifying this patient to have pneumonia.

Ślęzak [7] suggested an approach for defining the three probabilistic regions based on pair-wise comparisons of categories. A matrix of thresholds is used, with a pair of thresholds on the likelihood ratio of each pair of categories. Although the approach is mathematically appealing and sound, it suffers from a lack of guidelines and systematic methods on how to determining the required thresholds, one may have difficulties in estimating all thresholds.

In the multi-category solution given in the original Bayesian decision theory, one has the option of assigning the object to one of the m classes. In other words, each class is associated with an action of accepting the object to be a member of that class. Lingras et al. [3] proposed a rough multi-category decision theoretic framework based on DTRS by building a similarity relation between each object and an action corresponding to a subset of categories. The final classification in both of these approaches is made by choosing the action with the minimum expected loss. However, one has to make an immediate decision to either accept nor reject the object to be a member of one of the classes, and the losses incurred for making rejections to different classes are not considered.

In this paper, a probabilistic approximation for multi-category decision problems based on the three-way decision approach of DTRS [10,11] is introduced. Instead of making an immediate accept or reject decision, a third option of making a deferred decision is added to each class. This gives the user the flexibility of refusing to make a decision in close cases. For example, if the doctor cannot diagnose between a few different types of flu based on a patient's symptoms, a series of diagnose tests can be performed to gather more information to help the doctor making the decision. Moreover, the losses incurred for misclassifying an object into different substitution classes are treated differently, and the losses incurred for making deferred and rejective decisions to different classes are considered. The goal of this paper is to tackle the limitations in the previous work on multi-category decision making of rough set, and therefore provide a more complete solution.

The rest of the paper is organized as follows. In Section 2, I briefly review the three-way decisions introduced in DTRS models. Section 3 includes three parts: I first review the existing work on multi-category classification with probabilistic rough set models; a new formulation of this problem is then proposed and its differences with other existing work are analyzed; an example is given to demonstrate the usefulness of the new approach. Section 4 concludes the paper and points out possible future work.

2 Three-Way Decisions with DTRS

Let $R \subseteq U \times U$ be an equivalence relation on the universe U, namely, R is reflexive, symmetric, and transitive. The pair $apr = (U, R)$ is called an approximation space. The equivalence relation R induces a partition of U, denoted by U/R. The basic building blocks in constructing rough set theory are the equivalence classes of R. For an object $x \in U$, the equivalence class containing x is given by $[x]_R = [x] = \{y \in U \mid xRy\}$; we omit the subscript R when the equivalence relation R is understood.

In probabilistic rough set models, the degree of overlap between an equivalence class $[x]$ and a set of objects C is considered. A conditional probability is used to state the degree of overlap and a pair of threshold values α and β with $\alpha > \beta$ on the probability is used to define three probabilistic regions. Let $Pr(C|[x])$ be the probability of an object belonging to C given that the object is in $[x]$. A fundamental result of DTRS models is (α, β)-probabilistic positive, boundary and negative regions defined by [12]:

$$\begin{aligned} \mathrm{POS}_{(\alpha,\beta)}(C) &= \{x \in U \mid Pr(C|[x]) \geq \alpha\}, \\ \mathrm{BND}_{(\alpha,\beta)}(C) &= \{x \in U \mid \beta < Pr(C|[x]) < \alpha\}, \\ \mathrm{NEG}_{(\alpha,\beta)}(C) &= \{x \in U \mid Pr(C|[x]) \leq \beta\}. \end{aligned} \tag{1}$$

We accept an object x to be a member of C if the conditional probability is greater than α. We reject x to be a member of C if the conditional probability is less than β. We neither accept nor reject x to be a member of C if the conditional probability is between of α and β, instead, we make a decision of deferment. The boundary region does not involve acceptance and rejection errors, but it is associated with cost of deferment.

3 Multi-category Classification with DTRS

A probabilistic approximation of multi-category classification is given in this section. The existing multi-category classification solutions are reviewed as comparisons.

3.1 Existing Work

In Bayesian decision theory [1], let $\Omega = \{C_1, C_2, \ldots, C_m\}$ denote a finite set of m classes and $\mathcal{A} = \{a_1, a_2, \ldots, a_m\}$ a finite set of m possible actions. The loss function $\lambda(a_i|C_j)$ is given by an $m \times m$ matrix:

	C_1	C_2	$\cdots$	C_i	$\cdots$	C_m
a_1	$\lambda_{11} = \lambda(a_1\|C_1)$	$\lambda_{12} = \lambda(a_1\|C_2)$	$\cdots$	$\lambda_{1i} = \lambda(a_1\|C_i)$	$\cdots$	$\lambda_{1m} = \lambda(a_1\|C_m)$
a_2	$\lambda_{21} = \lambda(a_2\|C_1)$	$\lambda_{22} = \lambda(a_2\|C_2)$	$\cdots$	$\lambda_{2i} = \lambda(a_2\|C_i)$	$\cdots$	$\lambda_{2m} = \lambda(a_2\|C_m)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
a_m	$\lambda_{m1} = \lambda(a_m\|C_1)$	$\lambda_{m2} = \lambda(a_m\|C_2)$	$\cdots$	$\lambda_{mi} = \lambda(a_m\|C_i)$	$\cdots$	$\lambda_{mm} = \lambda(a_m\|C_m)$

In general, the selection of the values of the loss function can be represented as:

$$\lambda(a_i|C_j) = \begin{cases} 0 & i = j \\ \lambda_{ij} & i \neq j, \end{cases}$$

where λ_{ij} $(i, j = 1, \ldots, m)$ denote the loss incurred for deciding class C_i when the true class is C_j. The expected cost associated with taking action a_i is:

$$R(a_i|[x]) = \sum_{j=1}^{j=m} \lambda(a_i|C_j)Pr(C_j|[x]). \tag{2}$$

The classifier is said to assign an object to class C_i if

$$R(a_i|[x]) < R(a_j|[x]) \qquad \text{for all } j \neq i.$$

Inspired by the above multi-category classification solution, Yao [13] suggested to change an m-category classification problem into m two-category classification problems, and make a three-way decision to each class. For the finite set of m classes $\Omega = \{C_1, C_2, \ldots, C_m\}$, the C_i's form a family of pair-wise disjoint subsets of U, namely, $C_i \cap C_j = \emptyset$ for $i \neq j$, and $\cup C_i = U$. For each C_i, a two-category classification $\{C, C^c\}$ can be defined, where $C = C_i$ and $C^c = C_i^c = \cup_{i \neq j} C_j$. The loss function for each C_i is given by an $3 \times m$ matrix:

	C_1	C_2	$\cdots$	C_i	$\cdots$	C_m
a_{P_i}	$\lambda_{P1} = \lambda(a_{P_i}\|C_i^c)$	$\lambda_{P2} = \lambda(a_{P_i}\|C_i^c)$	$\cdots$	$\lambda_{Pi} = \lambda(a_{P_i}\|C_i)$	$\cdots$	$\lambda_{Pm} = \lambda(a_{P_i}\|C_i^c)$
a_{B_i}	$\lambda_{B1} = \lambda(a_{B_i}\|C_i^c)$	$\lambda_{B2} = \lambda(a_{B_i}\|C_i^c)$	$\cdots$	$\lambda_{Bi} = \lambda(a_{B_i}\|C_i)$	$\cdots$	$\lambda_{Bm} = \lambda(a_{B_i}\|C_i^c)$
a_{N_i}	$\lambda_{N1} = \lambda(a_{N_i}\|C_i^c)$	$\lambda_{N2} = \lambda(a_{N_i}\|C_i^c)$	$\cdots$	$\lambda_{Ni} = \lambda(a_{N_i}\|C_i)$	$\cdots$	$\lambda_{Nm} = \lambda(a_{N_i}\|C_i^c)$

m pairs of thresholds can be systematically calculated based on the given loss functions. The results from Section 2.3 can be immediately applied. However, the losses incurred for making any substitution errors are considered as the same in this approach. That is,

$$\begin{cases} \lambda_{P1} = \lambda_{P2} = \ldots = \lambda_{Pm} = \lambda(a_{Pi}|C_i^c) & \text{for all } i \neq j \\ \lambda_{B1} = \lambda_{B2} = \ldots = \lambda_{Bm} = \lambda(a_{Bi}|C_i^c) & \text{for all } i \neq j \end{cases}$$

This assumption does not always hold in many real world scenario, which makes it not practice in real applications.

Ślęzak [7] introduced a rough Bayesian model (RB) and applied it to multi-category classifications based on pair-wise comparisons of decision classes. For an information table with a set of decision classes $\{0, \ldots, m-1\}$, a matrix of thresholds is given as:

$$\varepsilon = \begin{bmatrix} * & \varepsilon_1^0 & \cdots & \varepsilon_{m-1}^0 \\ \varepsilon_0^1 & * & & \vdots \\ \vdots & & * & \varepsilon_{m-1}^{m-2} \\ \varepsilon_0^{m-1} & \cdots & \varepsilon_{m-2}^{m-1} & * \end{bmatrix}.$$

where $\varepsilon_i^j \in [0, 1)$ for $i \neq j$ is called a significance threshold that expresses whether the degree of belief is strong enough for class C_i with respect to any other class

C_j. The three probabilistic regions are defined as follows:

$$\begin{aligned} \text{POS}_{(\varepsilon)}(C_i) &= \{x \in U \mid \forall_{j:j\neq i} Pr([x]|C_j) \leq \varepsilon_i^j Pr([x]|C_i)\}, \\ \text{BND}_{(\varepsilon)}(C_i) &= \{x \in U \mid \exists_{j:j\neq i} Pr([x]|C_j) > \varepsilon_i^j Pr([x]|C_i) \wedge \\ &\qquad \forall_{j:j\neq i} Pr([x]|C_i) > \varepsilon_j^i Pr([x]|C_j)\}, \\ \text{NEG}_{(\varepsilon)}(C_i) &= \{x \in U \mid \exists_{j:j\neq i} Pr([x]|C_i) \leq \varepsilon_j^i Pr([x]|C_j)\}. \end{aligned} \tag{3}$$

An object x belongs to $\text{POS}_{(\varepsilon)}(C_i)$ if and only if $I_a(x)$ is significantly more likely to occur under C_i than under any other class C_j, $j \neq i$. Object x belongs to $\text{BND}_{(\varepsilon)}(C_i)$ if and only if $I_a(x)$ is not significantly more likely under C_i than under all other C_j but there is also no alternative class, which makes $I_a(x)$ significantly more likely than C_i does. object x belongs to $\text{NEG}_{(\varepsilon)}(C_i)$ if and only if there is an alternative class C_j, which makes $I_a(x)$ significantly more likely than C_i does. However, the selection of significance thresholds can be a subjective and difficult task without systematic guidelines.

3.2 A New Formulation

Similar to Yao's idea, for the finite set of m classes $\Omega = \{C_1, C_2, \ldots, C_m\}$, we make a three-way decision to each class C_i, that is, each C_i is associated with a set of three actions $\mathcal{A} = \{a_{P_i}, a_{B_i}, a_{N_i}\}$, where a_{P_i}, a_{B_i}, and a_{N_i} represent the three actions in deciding $x \in \text{POS}(C_i)$, $x \in \text{BND}(C_i)$, and $x \in \text{NEG}(C_i)$, respectively. The loss function is given by a $3 \times m$ matrix for each C_i:

	C_1	C_2	$\cdots$	C_i	$\cdots$	C_m
a_{P_i}	$\lambda_{P1} = \lambda(a_{P_i}\|C_1)$	$\lambda_{P2} = \lambda(a_{P_i}\|C_2)$	$\cdots$	$\lambda_{Pi} = \lambda(a_{P_i}\|C_i)$	$\cdots$	$\lambda_{Pm} = \lambda(a_{P_i}\|C_m)$
a_{B_i}	$\lambda_{B1} = \lambda(a_{B_i}\|C_1)$	$\lambda_{B2} = \lambda(a_{B_i}\|C_2)$	$\cdots$	$\lambda_{Bi} = \lambda(a_{B_i}\|C_i)$	$\cdots$	$\lambda_{Bm} = \lambda(a_{B_i}\|C_m)$
a_{N_i}	$\lambda_{N1} = \lambda(a_{N_i}\|C_1)$	$\lambda_{N2} = \lambda(a_{N_i}\|C_2)$	$\cdots$	$\lambda_{Ni} = \lambda(a_{N_i}\|C_i)$	$\cdots$	$\lambda_{Nm} = \lambda(a_{N_i}\|C_m)$

Different to Yao's approach, the losses incurred for making substitution errors are considered differently. That is,

$$\begin{cases} \lambda_{Pi} = \lambda(a_{Pi}|C_i^c) & \text{for all } i \neq j \\ \lambda_{Bi} = \lambda(a_{Bi}|C_i^c) & \text{for all } i \neq j \end{cases}$$

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

$$\begin{aligned} R(a_{P_i}|[x]) &= \sum_{j=1}^{j=m} Pr(C_j|[x])\lambda(a_{P_i}|C_j), \\ R(a_{B_i}|[x]) &= \sum_{j=1}^{j=m} Pr(C_j|[x])\lambda(a_{B_i}|C_j), \\ R(a_{N_i}|[x]) &= \sum_{j=1}^{j=m} Pr(C_j|[x])\lambda(a_{N_i}|C_j). \end{aligned} \tag{4}$$

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

(P) If $R(a_{P_i}|[x]) \leq R(a_{B_i}|[x])$ and $R(a_{P_i}|[x]) \leq R(a_{N_i}|[x])$, decide $x \in \mathrm{POS}(C_i)$;

(B) If $R(a_{B_i}|[x]) \leq R(a_{P_i}|[x])$ and $R(a_{B_i}|[x]) \leq R(a_{N_i}|[x])$, decide $x \in \mathrm{BND}(C_i)$;

(N) If $R(a_{N_i}|[x]) \leq R(a_{P_i}|[x])$ and $R(a_{N_i}|[x]) \leq R(a_{B_i}|[x])$, decide $x \in \mathrm{NEG}(C_i)$.

Tie-breaking criteria should be added so that each object is put into only one region.

Ordinarily, the loss incurred for making an error is greater than the loss incurred for being correct, and the loss incurred for making a deferment decision is in between. Consider a special kind of loss functions with:

(c0). $\lambda(a_{P_i}|C_i) \leq \lambda(a_{B_i}|C_i) < \lambda(a_{N_i}|C_i),\ \lambda(a_{N_i}|C_j) \leq \lambda(a_{B_i}|C_j) < \lambda(a_{P_i}|C_j)$, for all $j, j \neq i$.

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

$$
\begin{aligned}
& R(a_{P_i}|[x]) \leq R(a_{B_i}|[x]) \\
\iff\quad & \sum_{j=1}^{j=s} Pr(C_j|[x])\lambda(a_{P_i}|C_j) \leq \sum_{j=1}^{j=s} Pr(C_j|[x])\lambda(a_{B_i}|C_j) \\
\iff\quad & Pr(C_i|[x])\lambda(a_{P_i}|C_i) + \sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])\lambda(a_{P_i}|C_j) \\
& \leq Pr(C_i|[x])\lambda(a_{B_i}|C_i) + \sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])\lambda(a_{B_i}|C_j) \\
\iff\quad & Pr(C_i|[x]) \geq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{B_i}|C_j))}{\lambda(a_{B_i}|C_i) - \lambda(a_{P_i}|C_i)}.
\end{aligned}
$$

Similarly, other conditions of the three rules can be expressed as:

$$R(a_{P_i}|[x]) \leq R(a_{N_i}|[x]) \iff Pr(C_i|[x]) \geq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{P_i}|C_i)},$$

$$R(a_{B_i}|[x]) \leq R(a_{P_i}|[x]) \iff Pr(C_i|[x]) \leq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{B_i}|C_j))}{\lambda(a_{B_i}|C_i) - \lambda(a_{P_i}|C_i)},$$

$$R(a_{B_i}|[x]) \leq R(a_{N_i}|[x]) \iff Pr(C_i|[x]) \geq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{B_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{B_i}|C_i)},$$

$$R(a_{N_i}|[x]) \leq R(a_{P_i}|[x]) \iff Pr(C_i|[x]) \leq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{P_i}|C_i)},$$

$$R(a_{N_i}|[x]) \leq R(a_{B_i}|[x]) \iff Pr(C_i|[x]) \leq \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{B_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{B_i}|C_i)}.$$

By introducing three parameters:

$$\alpha_i = \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{B_i}|C_j))}{\lambda(a_{B_i}|C_i) - \lambda(a_{P_i}|C_i)},$$
$$\beta_i = \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{B_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{B_i}|C_i)},$$
$$\gamma_i = \frac{\sum_{j=1,j\neq i}^{j=s} Pr(C_j|[x])(\lambda(a_{P_i}|C_j) - \lambda(a_{N_i}|C_j))}{\lambda(a_{N_i}|C_i) - \lambda(a_{P_i}|C_i)}. \tag{5}$$

We can express concisely the decision rules (P)-(N) as:

(P) If $Pr(C_i|[x]) \geq \alpha_i$ and $Pr(C_i|[x]) \geq \gamma_i,$ decide $x \in \text{POS}(C_i)$;
(B) If $Pr(C_i|[x]) \leq \alpha_i$ and $Pr(C_i|[x]) \geq \beta_i,$ decide $x \in \text{BND}(C_i)$;
(N) If $Pr(C_i|[x]) \leq \beta_i$ and $Pr(C_i|[x]) \leq \gamma_i,$ decide $x \in \text{NEG}(C_i)$.

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

The conditions of rule (B) suggest that it maybe reasonable to impose the constraint $\alpha_i > \beta_i$ so that the boundary region may be non-empty. We can add a sufficient condition on the loss function to ensure $\alpha_i > \beta_i$ as follow:

$$\text{(c1).} \qquad \frac{\lambda(a_{N_i}|C_i) - \lambda(a_{B_i}|C_i)}{\lambda(a_{B_i}|C_j) - \lambda(a_{N_i}|C_j)} > \frac{\lambda(a_{B_i}|C_i) - \lambda(a_{P_i}|C_i)}{\lambda(a_{P_i}|C_j) - \lambda(a_{B_i}|C_j)}. \tag{6}$$

The condition (c0) and (c1) imply that $\alpha_i > \gamma_i > \beta_i \geq 0$. After tie-breaking, the following simplified rules are obtained:

(P) If $Pr(C_i|[x]) \geq \alpha_i,$ decide $x \in \text{POS}(C_i)$;
(B) If $\beta_i < Pr(C_i|[x]) < \alpha_i,$ decide $x \in \text{BND}(C_i)$;
(N) If $Pr(C_i|[x]) \leq \beta_i,$ decide $x \in \text{NEG}(C_i)$.

From the rules (P), (B), and (N), the (α_i, β_i)-probabilistic positive, negative and boundary regions are given, respectively, by:

$$\begin{aligned} \text{POS}_{(\alpha_i,\beta_i)}(C_i) &= \{x \in U \mid Pr(C_i|[x]) \geq \alpha_i\}, \\ \text{BND}_{(\alpha_i,\beta_i)}(C_i) &= \{x \in U \mid \beta_i < Pr(C_i|[x]) < \alpha_i\}, \\ \text{NEG}_{(\alpha_i,\beta_i)}(C_i) &= \{x \in U \mid Pr(C_i|[x]) \leq \beta_i\}. \end{aligned} \tag{7}$$

We can extend the probabilistic approximations and regions of a single class to a partition. Let π_Ω be a partition of the universe U, defined by the decision attribute Ω. The three regions of the partition π_Ω can be defined as:

$$\begin{aligned} \text{POS}_{(\alpha,\beta)}(\pi_\Omega) &= \bigcup_{1\leq i\leq m} \text{POS}_{(\alpha_i,\beta_i)}(C_i), \\ \text{BND}_{(\alpha,\beta)}(\pi_\Omega) &= \bigcup_{1\leq i\leq m} \text{BND}_{(\alpha_i,\beta_i)}(C_i), \\ \text{NEG}_{(\alpha,\beta)}(\pi_\Omega) &= U - \text{POS}_{(\alpha,\beta)}(\pi_\Omega) \cup \text{BND}_{(\alpha,\beta)}(\pi_\Omega). \end{aligned} \tag{8}$$

It is necessary to have a further study on the probabilistic three regions of a classification, as well as the associated rules. In general, one has to consider

Table 1. A Loss Function Table

	C_1	C_2	C_3	C_4
a_{P_1}	0	2	3	8
a_{B_1}	5	1	1	3
a_{N_1}	10	0	0	0
a_{P_2}	10	0	20	30
a_{B_2}	5	7	10	15
a_{N_2}	0	15	0	0
a_{P_3}	28	25	0	30
a_{B_3}	12	11	10	13
a_{N_3}	0	0	20	0
a_{P_4}	9	5	3	0
a_{B_4}	4	2	1	15
a_{N_4}	0	0	0	30

the problem of rule conflict resolution in order to make effective acceptance, rejection, and abstaining decisions.

3.3 An Example

Consider a medical diagnose example, there is a set of four types of diseases $\Omega = \{C_1, C_2, C_3, C_4\}$. Each disease C_i associated with three actions $\mathcal{A} = \{a_{P_i}, a_{B_i}, a_{N_i}\}$, where a_{P_i} indicates that we accept a patient to have disease C_i, a_{N_i} indicates that we reject a patient to have disease C_i, and a_{B_i} indicates that we neither accept nor reject a patient to have disease C_i, a diagnose test needs to be performed in order to decide whether or not the patient having disease C_i. Suppose the symptoms of a new patient are described by $[x]$, the conditional probabilities of the four diseases can be derived from historical data of the hospital as follows: $Pr(C_1|[x]) = 0.4$, $Pr(C_2|[x]) = 0.2$, $Pr(C_3|[x]) = 0.15$, and $Pr(C_4|[x]) = 0.25$. The loss functions for the four diseases are represented in Table 1. Suppose the four diseases are listed in the increasing order of their severe levels, we can see that the loss incurred for misdiagnosing a patient having disease C_1 to C_4 is higher than misdiagnosing a patient having disease C_1 to C_2 or C_3.

Based on equation (5), we can computer the pairs of thresholds (α_i, β_i) for the four diseases as: $\alpha_1 = 0.35$, $\beta_1 = 0.22$; $\alpha_2 = 1.04$, $\beta_2 = 0.91$; $\alpha_3 = 1.35$, $\beta_3 = 1.03$; $\alpha_4 = 0.19$, $\beta_4 = 0.14$. Now we can compare the conditional probabilities with the corresponding thresholds. Since $Pr(C_1|[x]) = 0.4 > \alpha_1$ and $Pr(C_4|[x]) = 0.25 > \alpha_4$, class C_1 and C_4 are in the positive region. The new patient could have both disease C_1 and C_4, or either one of them. Rule conflict resolution should be added at this point to further distinguish which disease that the patient is more likely to have.

4 Conclusions and Future Work

This paper introduces a probabilistic rough set approximation for an information tables with more than two decision classes. In order to emphasis the semantic interpretation of probabilistic rough sets with three-way decisions, our formulation directly uses three pair-wise disjoint positive, boundary, and negative regions instead of a pair of lower and upper approximations. This approach can be considered as a straightforward generalization of the three-way classification introduced in decision-theoretic rough set models. The differences between our approach and other existing work are analyzed. One may have a further study on three-way decision rules generated from different classes and the associated rule conflict resolutions for real classification applications.

References

1. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. Wiley, New York (1973)
2. Herbert, J.P., Yao, J.T.: Game-theoretic risk analysis in decision-theoretic rough sets. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 132–139. Springer, Heidelberg (2008)
3. Lingras, P., Chen, M., Miao, D.Q.: Rough multi-category decision theoretic framework. In: Wang, G., Li, T., Grzymala-Busse, J.W., Miao, D., Skowron, A., Yao, Y. (eds.) RSKT 2008. LNCS (LNAI), vol. 5009, pp. 676–683. Springer, Heidelberg (2008)
4. Liu, D., Li, T.R., Hu, P., Li, H.X.: Multiple-category classification with decision-theoretic rough sets. In: Yu, J., Greco, S., Lingras, P., Wang, G., Skowron, A. (eds.) RSKT 2010. LNCS(LNAI), vol. 6401, pp. 703–710. Springer, Heidelberg (2010)
5. Pawlak, Z.: Rough Sets, Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers, Dordrecht (1991)
6. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. International Journal of Man-Machine Studies 29, 81–95 (1988)
7. Ślęzak, D.: Rough Sets and Bayes Factor. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets III. LNCS, vol. 3400, pp. 202–229. Springer, Heidelberg (2005)
8. Yao, Y.Y.: Probabilistic approaches to rough sets. Expert Systems 20, 287–297 (2003)
9. Yao, Y.Y.: Probabilistic rough set approximations. International Journal of Approximate Reasoning 49, 255–271 (2008)
10. Yao, Y.Y.: Three-way decisions with probabilistic rough sets. Information Sciences 180, 341–353 (2010)
11. Yao, Y.Y.: The superiority of three-way decisions in probabilistic rough set models. Information Sciences 181, 1080–1096 (2011)
12. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximating concepts. International Journal of Man-Machine Studies 37, 793–809 (1992)
13. Yao, Y.Y., Zhao, Y.: Attribute reduction in decision-theoretic rough set models. Information Sciences 178, 3356–3373 (2008)
14. Yao, Y.Y., Zhou, B.: Naive bayesian rough sets. In: Yu, J., Greco, S., Lingras, P., Wang, G., Skowron, A. (eds.) RSKT 2010. LNCS(LNAI), vol. 6401, pp. 719–726. Springer, Heidelberg (2010)
15. Zhou, B., Yao, Y.Y., Luo, J.G.: A three-way decision approach to email spam filtering. In: Farzindar, A., Kešelj, V. (eds.) Canadian AI 2010. LNCS(LNAI), vol. 6085, pp. 28–39. Springer, Heidelberg (2010)
16. Ziarko, W.: Variable precision rough sets model. Journal of Computer and Systems Sciences 46, 39–59 (1993)