

## Three-way Investment Decisions with Decision-theoretic Rough Sets

Dun Liu<sup>1\*</sup>, Yiyu Yao<sup>2</sup>, Tianrui Li<sup>3</sup>

<sup>1</sup> Department of Economics and Management, Southwest Jiaotong University  
Chengdu, 610031, P.R.China <sup>†</sup>  
E-mail: newton83@163.com

<sup>2</sup> Department of Computer Science, University of Regina  
Regina, Saskatchewan, Canada. S4S 0A2  
E-mail: yyao@cs.uregina.ca

<sup>3</sup> Department of Information Science and Technology, Southwest Jiaotong University  
Chengdu, 610031, P.R.China  
E-mail: trli@swjtu.edu.cn

Received: 01-02-2009

Accepted: 09-09-2010

### Abstract

The decision-theoretic rough set model is adopted to derive a profit-based three-way approach to investment decision-making. A three-way decision is made based on a pair of thresholds on conditional probabilities. A positive rule makes a decision of investment, a negative rule makes a decision of non-investment, and a boundary rule makes a decision of deferment. Both cost functions and revenue functions are used to calculate the required two thresholds by maximizing conditional profit with the Bayesian decision procedure. A case study of oil investment demonstrates the proposed method.

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

### 1. Introduction

As a methodology to deal with uncertain decision problems, rough set theory (RST) uses a pair of sets, the lower approximation and upper approximation, to describe a set representing instances of a concept.<sup>15</sup> 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.<sup>16</sup> The two

approximations divide the universe into three pairwise disjoint regions: the positive region, boundary region and negative region. The positive region is given by the lower approximation, the boundary region by the difference of upper and lower approximations, and the negative region by the complement of the upper approximation.<sup>27,29</sup>

Corresponding to the three regions, Yao introduces and studies the notion of three-way decisions, consisting of the positive, boundary and negative

\*This work is partially supported by the National Science Foundation of China (No. 60873108), the Doctoral Innovation Foundation of Southwest Jiaotong University (No. 200907), and the Scientific Research Foundation of Graduate School of Southwest Jiaotong University (No. 2009LD), China.

<sup>†</sup>Corresponding author.

rules.<sup>27</sup> The positive rules generated by the positive region make decisions of acceptance, the negative rules generated by the negative region make decisions of rejection, and the boundary rules generated by the boundary region make deferred or non-committed decisions.<sup>29</sup> By considering tolerance of errors of the three-way decisions, Yao et al. introduce Bayesian decision procedure into RST and propose a decision-theoretic rough set (DTRS) model.<sup>25,31,32</sup> The acceptable level of errors or accuracy (a pair of thresholds) can be automatically computed from risk/cost functions.<sup>25,31,32</sup> The optimal decision of acceptance, rejection or deferment with the minimum conditional risk can be directly calculated in DTRS.

A main difficulty in applying DTRS is the estimation of loss functions. To overcome this difficulty, Herbert and Yao introduce a game-theoretic approach to DTRS for learning optimal parameter values.<sup>4,6</sup> Measures of classification ability are interpreted as players in a game, each with a goal of optimizing its value; actions performed in this game consist of increasing or decreasing the size of the classification regions. Zhou and Li use the concepts of optimistic decision, pessimistic decision, and equable decision to describe the cost of misclassification, the thresholds of probabilistic inclusion are calculated based on the minimization of risk cost under respective decision bias.<sup>7,37</sup>

In this paper, we focus on the semantic studies on investment problems. A model of investment is presented based on three-way decisions: investment, non-investment or further investigation. By considering the cost and the revenue simultaneously, a profit-based analysis of three-way investment decisions is given. We choose the optimal action with maximum conditional profit by using Bayesian decision procedure. A pair of a cost function and a revenue function is used to calculate the two thresholds automatically. Our approach is a natural extension of DTRS. Some additional semantics for investment decision are considered.

The rest of the paper is organized as follows. Section 2 reviews basic concepts of DTRS. Section 3 proposes a three-way investment decision model. A case study of oil investment is given in Section 4.

## 2. Overview of Decision-theoretic Rough Sets

Basic concepts, notations and results of probabilistic rough sets as well as their extensions are briefly reviewed in this section,<sup>5,9,11,15,19,24,25,26,27,28,29,30,38</sup> and Liu et.al summarize the two decades research on decision-theoretic rough sets.<sup>9</sup>

Let  $U$  be a finite and nonempty set and  $R$  an equivalence relation on  $U$ . The pair  $apr = (U, R)$  is called an approximation space, The equivalence relation  $R$  induces a partition of  $U$ , denoted by  $U/R$ . For a subset  $X \subseteq U$ , its lower and upper approximations are defined by:

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

where  $[x]$  is the equivalence class containing  $x$ . Based on the rough set approximations of  $X$ , one can divide the universe  $U$  into three pair-wise disjoint regions: the positive region  $POS(X)$ , the boundary region  $BND(X)$ , and the negative region  $NEG(X)$ :

$$\begin{aligned} POS(X) &= \underline{apr}(X); \\ BND(X) &= \overline{apr}(X) - \underline{apr}(X); \\ NEG(X) &= U - \overline{apr}(X). \end{aligned} \quad (2)$$

A rule generated by an element  $x \in POS(X)$  implies that one can certainly accept  $x$  as a member of  $X$ , the rule generated by an element  $x \in NEG(X)$  expresses the fact that one certainly reject  $x$  as a member of  $X$ , and a rule generated by  $x \in BND(X)$  leads to a situation of undecidability. The structure of three disjoint regions motivates the three-way decision-making.

Let  $S = (U, A, V, f)$  be an information table.  $\forall x \in U, X \subseteq U$ , let,

$$Pr(X|[x]) = \frac{|[x] \cap X|}{|[x]|}, \quad (3)$$

where  $|\cdot|$  stands for the cardinality of a set, and  $Pr(X|[x])$  denotes the conditional probability of the classification. This simple method for estimating the conditional probability based on the cardinalities of sets is used as an illustration. In general, one may consider other methods for estimating the probability more accurately.<sup>28</sup>

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

|     |                |                |
|-----|----------------|----------------|
|     | $X (P)$        | $X^c (N)$      |
| $P$ | $\lambda_{PP}$ | $\lambda_{PN}$ |
| $B$ | $\lambda_{BP}$ | $\lambda_{BN}$ |
| $N$ | $\lambda_{NP}$ | $\lambda_{NN}$ |

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

Suppose  $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$  and  $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$ , by using of the Bayesian decision procedure,<sup>1</sup> the  $(\alpha, \beta)$ -probabilistic lower and upper approximations are defined as follows:

$$\begin{aligned} \underline{apr}_{(\alpha, \beta)}(X) &= \{x \in U \mid Pr(X|[x]) \geq \alpha\}, \\ \overline{apr}_{(\alpha, \beta)}(X) &= \{x \in U \mid Pr(X|[x]) > \beta\}. \end{aligned} \quad (4)$$

where

$$\begin{aligned} \alpha &= \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \\ \beta &= \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \end{aligned} \quad (5)$$

In the case with  $0 \leq \beta < \alpha \leq 1$  (see references [28,29] for condition on the loss function in  $\lambda$ ), after tie-breaking, the three-way decision rules are given as follows:

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

The three regions of rough set theory lead to three-way decisions.

The three-way perspective on the theory of rough sets makes it more applicable. The three regions may also be interpreted as the positive verification and negative verification of a hypothesis, as well as undecidability, based on a given piece of evidence.<sup>19</sup> The notion of three-way decisions is related to many studies, such as hypothesis testing,<sup>21</sup> medical decision-making,<sup>10,14</sup> products inspecting process,<sup>23</sup> documents classification,<sup>8</sup> model selection criteria,<sup>2,36</sup> environmental management,<sup>3</sup> data packs selection,<sup>20</sup> and email spam filtering.<sup>34,35</sup>

When  $\alpha = 1$ ,  $\beta = 0$ , the DTRS model becomes the standard Pawlak rough set model;<sup>15</sup> when  $\alpha = \beta = 0.5$ , the DTRS model becomes the 0.5 probabilistic model;<sup>17</sup> when  $\alpha = 1 - \beta$  and  $\alpha \in (0.5, 1]$ , the DTRS model reduces to the VPRS model.<sup>38</sup>

### 3. A Three-way Decision Model of Investment

Rough set theory has been successfully used in many domains, with different interpretations of the theory. In model construction, one focuses more on the syntax aspects. In applications, one considers algorithms, especially attribute reduction algorithms and rule generation algorithms. In this paper, we mainly focus on the semantics studies in investment decisions by using RST. In this section, a three-way decision approach is introduced for investment. The goal is to seek for the maximum expected profit. Profit is estimated from two aspects, revenue and cost; a good investment should have a higher revenue and a lower cost.

The well known Bayesian decision procedure can be explained as follows.<sup>1</sup> Let  $\Omega = \{s_1, s_2, \dots, s_m\}$  be a finite set of  $m$  states, and let  $\mathcal{A} = \{a_1, a_2, \dots, a_n\}$  be a finite set of  $n$  possible actions. Let  $Pr(w_i|\mathbf{x})$  be the conditional probability of an object being in the state  $w_i$  given its description  $\mathbf{x}$ . Let  $\theta(a_j|w_i)$  denote the revenue, or gain, and  $\varphi(a_j|w_i)$  denote the cost, or loss for taking action  $a_j$  when the state is  $w_i$ . For an object  $\mathbf{x}$ , suppose action  $a_j$  is taken. Since  $Pr(w_i|\mathbf{x})$  is the probability that the true state is  $w_i$  given  $\mathbf{x}$ , the expected difference between revenue and cost associated with taking action

$a_j$  is given by the following formula:

$$R(a_j|\mathbf{x}) = \sum_{i=1}^m (\theta(a_j|w_i) - \varphi(a_j|w_i))Pr(w_i|\mathbf{x}). \quad (6)$$

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

Given a description  $\mathbf{x}$ , a decision rule is function  $\tau(\mathbf{x})$  specifying which action to take. The overall profit is expressed as the sum of expected profit associated with a given rule. Since  $R(\tau(\mathbf{x})|\mathbf{x})$  is the conditional profit associated with action  $\tau(\mathbf{x})$ . The overall profit is defined by:

$$\mathcal{R} = \sum_{\mathbf{x} \in U} R(\tau(\mathbf{x})|\mathbf{x})Pr(\mathbf{x}), \quad (7)$$

where the summation is over the set of all possible descriptions of objects. If  $\tau(\mathbf{x})$  is chosen so that  $R(a_j|\mathbf{x})$  is as big as possible for every  $\mathbf{x}$ , the overall profit  $\mathcal{R}$  is maximized. Thus, the Bayesian decision procedure can be formally states as follows. For every  $\mathbf{x}$ , compute the conditional profit  $R(a_j|\mathbf{x})$  for  $i = 1, 2, \dots, n$  and select the action for which the conditional profit is maximum.

To apply the Bayesian decision procedure, the decision-theoretic rough set model starts from defining the three probabilistic regions from which the probabilistic approximations are defined. For investment decisions, the DTRS is slightly modified.

With respect to the three regions, the set of actions, the profit regarding the revenue and cost of classification actions with respect to different states are given by the  $3 \times 2$  matrix:

|       | $X (P)$                      | $X^c (N)$                    |
|-------|------------------------------|------------------------------|
| $a_P$ | $\theta_{PP} - \varphi_{PP}$ | $\theta_{PN} - \varphi_{PN}$ |
| $a_B$ | $\theta_{BP} - \varphi_{BP}$ | $\theta_{BN} - \varphi_{BN}$ |
| $a_N$ | $\theta_{NP} - \varphi_{NP}$ | $\theta_{NN} - \varphi_{NN}$ |

In the matrix,  $\theta_{PP} - \varphi_{PP}$ ,  $\theta_{BP} - \varphi_{BP}$  and  $\theta_{NP} - \varphi_{NP}$  denote the profits incurred for taking actions  $a_P$ ,  $a_B$  and  $a_N$  when an object belongs to  $X$ ; and  $\theta_{PN} - \varphi_{PN}$ ,  $\theta_{BN} - \varphi_{BN}$  and  $\theta_{NN} - \varphi_{NN}$  denote the profits incurred for taking actions  $a_P$ ,  $a_B$  and  $a_N$  when the object does not belong to  $X$ . Let  $\lambda_{PP} = \theta_{PP} - \varphi_{PP}$ ,  $\lambda_{PN} = \theta_{PN} - \varphi_{PN}$ ,  $\lambda_{BP} = \theta_{BP} - \varphi_{BP}$ ,  $\lambda_{BN} = \theta_{BN} - \varphi_{BN}$ ,  $\lambda_{NP} = \theta_{NP} - \varphi_{NP}$  and  $\lambda_{NN} = \theta_{NN} - \varphi_{NN}$ , we obtain

the same results of DTRS. The separation of  $\lambda$  into a revenue function  $\theta$  and a cost function  $\varphi$  makes it to be easily interpretable to an investment manager.

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

$$\begin{aligned} \mathcal{R}(a_P|[x]) &= (\theta_{PP} - \varphi_{PP})Pr(X|[x]) + (\theta_{PN} - \varphi_{PN})Pr(X^c|[x]), \\ \mathcal{R}(a_B|[x]) &= (\theta_{BP} - \varphi_{BP})Pr(X|[x]) + (\theta_{BN} - \varphi_{BN})Pr(X^c|[x]), \\ \mathcal{R}(a_N|[x]) &= (\theta_{NP} - \varphi_{NP})Pr(X|[x]) + (\theta_{NN} - \varphi_{NN})Pr(X^c|[x]). \end{aligned} \quad (8)$$

where the equivalence class  $[x]$  of  $x$  is viewed as description of  $x$ . The Bayesian decision procedure suggests the following maximum-profit decision rules:

- (P) If  $\mathcal{R}(a_P|[x]) \geq \mathcal{R}(a_B|[x])$  and  $\mathcal{R}(a_P|[x]) \geq \mathcal{R}(a_N|[x])$ , decide  $x \in \text{POS}(X)$ ;
- (B) If  $\mathcal{R}(a_B|[x]) \geq \mathcal{R}(a_P|[x])$  and  $\mathcal{R}(a_B|[x]) \geq \mathcal{R}(a_N|[x])$ , decide  $x \in \text{BND}(X)$ ;
- (N) If  $\mathcal{R}(a_N|[x]) \geq \mathcal{R}(a_P|[x])$  and  $\mathcal{R}(a_N|[x]) \geq \mathcal{R}(a_B|[x])$ , decide  $x \in \text{NEG}(X)$ .

Since  $Pr(X|[x]) + Pr(X^c|[x]) = 1$ , we can simplify the rules by using only the probabilities  $Pr(X|[x])$ , the revenue function  $\theta$ , and the cost function  $\varphi$ .

Consider a special revenue function with:

$$(c0). \quad \begin{aligned} \theta_{NP} < \theta_{BP} \leq \theta_{PP}, \\ \theta_{PN} < \theta_{BN} \leq \theta_{NN}. \end{aligned} \quad (9)$$

That is, the revenue of classifying an object  $x$  belonging to  $X$  into the positive region  $\text{POS}(X)$  is no less than the revenue of classifying  $x$  into the boundary region  $\text{BND}(X)$ , and both of these revenues are no less than the revenue of classifying  $x$  into the negative region  $\text{NEG}(X)$ . The reverse order of revenues are used for classifying an object not in  $X$ .

Similarly, consider a special cost function with:

$$(c1). \quad \begin{aligned} \varphi_{PP} \leq \varphi_{BP} < \varphi_{NP}, \\ \varphi_{NN} \leq \varphi_{BN} < \varphi_{PN}. \end{aligned} \quad (10)$$

That is, the cost of classifying an object  $x$  belonging to  $X$  into the positive region  $\text{POS}(X)$  is no more than the loss of classifying  $x$  into the boundary region  $\text{BND}(X)$ , and both of these costs are no more than the cost of classifying  $x$  into the negative region  $\text{NEG}(X)$ . The reverse order of costs are used for classifying an object not in  $X$ .

Under conditions (c0) and (c1), we can simplify decision rules (P)-(N) as follows. For the rule (P), the first condition can be expressed as:

$$\begin{aligned} \mathcal{R}(a_P|[x]) \geq \mathcal{R}(a_B|[x]) &\iff \\ (\theta_{PP} - \varphi_{PP})Pr(X|[x]) + (\theta_{PN} - \varphi_{PN})Pr(X^c|[x]) &\geq \\ (\theta_{BP} - \varphi_{BP})Pr(X|[x]) + (\theta_{BN} - \varphi_{BN})Pr(X^c|[x]). \end{aligned}$$

That is,

$$Pr(X|[x]) \geq \frac{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN})}{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN}) + (\theta_{PP} - \theta_{BP}) + (\varphi_{BP} - \varphi_{PP})}.$$

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

$$\begin{aligned} \mathcal{R}(a_P|[x]) \geq \mathcal{R}(a_N|[x]) &\iff \\ (\theta_{PP} - \varphi_{PP})Pr(X|[x]) + (\theta_{PN} - \varphi_{PN})Pr(X^c|[x]) &\geq \\ (\theta_{NP} - \varphi_{NP})Pr(X|[x]) + (\theta_{NN} - \varphi_{NN})Pr(X^c|[x]). \end{aligned}$$

That is,

$$Pr(X|[x]) \geq \frac{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN})}{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN}) + (\theta_{PP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{PP})}.$$

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

$$\begin{aligned} \mathcal{R}(a_B|[x]) \geq \mathcal{R}(a_P|[x]) &\iff \\ (\theta_{BP} - \varphi_{BP})Pr(X|[x]) + (\theta_{BN} - \varphi_{BN})Pr(X^c|[x]) &\geq \\ (\theta_{PP} - \varphi_{PP})Pr(X|[x]) + (\theta_{PN} - \varphi_{PN})Pr(X^c|[x]). \end{aligned}$$

That is,

$$Pr(X|[x]) \leq \frac{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN})}{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN}) + (\theta_{PP} - \theta_{BP}) + (\varphi_{BP} - \varphi_{PP})}.$$

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

$$\begin{aligned} \mathcal{R}(a_B|[x]) \geq \mathcal{R}(a_N|[x]) &\iff \\ (\theta_{BP} - \varphi_{BP})Pr(X|[x]) + (\theta_{BN} - \varphi_{BN})Pr(X^c|[x]) &\geq \\ (\theta_{NP} - \varphi_{NP})Pr(X|[x]) + (\theta_{NN} - \varphi_{NN})Pr(X^c|[x]). \end{aligned}$$

That is,

$$Pr(X|[x]) \geq \frac{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN})}{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN}) + (\theta_{BP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{BP})}.$$

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

$$\begin{aligned} \mathcal{R}(a_N|[x]) \geq \mathcal{R}(a_P|[x]) &\iff \\ (\theta_{NP} - \varphi_{NP})Pr(X|[x]) + (\theta_{NN} - \varphi_{NN})Pr(X^c|[x]) &\geq \\ (\theta_{PP} - \varphi_{PP})Pr(X|[x]) + (\theta_{PN} - \varphi_{PN})Pr(X^c|[x]); \end{aligned}$$

$$\begin{aligned} \mathcal{R}(a_N|[x]) \geq \mathcal{R}(a_B|[x]) &\iff \\ (\theta_{NP} - \varphi_{NP})Pr(X|[x]) + (\theta_{NN} - \varphi_{NN})Pr(X^c|[x]) &\geq \\ (\theta_{BP} - \varphi_{BP})Pr(X|[x]) + (\theta_{BN} - \varphi_{BN})Pr(X^c|[x]). \end{aligned}$$

That is,

$$Pr(X|[x]) \leq \frac{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN})}{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN}) + (\theta_{PP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{PP})},$$

and

$$Pr(X|[x]) \leq \frac{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN})}{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN}) + (\theta_{BP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{BP})}.$$

For convenience, we denote the three expressions in these conditions by the following three parameters:

$$\begin{aligned} \alpha &= \frac{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN})}{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN}) + (\theta_{PP} - \theta_{BP}) + (\varphi_{BP} - \varphi_{PP})}, \\ \beta &= \frac{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN})}{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN}) + (\theta_{BP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{BP})}, \\ \gamma &= \frac{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN})}{(\theta_{NN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{NN}) + (\theta_{PP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{PP})}. \end{aligned} \quad (11)$$

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

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

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

By setting  $\alpha > \beta$ , we can easily found the following condition on the revenue and cost functions:

$$\frac{(\theta_{PP} - \theta_{BP}) + (\varphi_{BP} - \varphi_{PP})}{(\theta_{BN} - \theta_{PN}) + (\varphi_{PN} - \varphi_{BN})} > \frac{(\theta_{BP} - \theta_{NP}) + (\varphi_{NP} - \varphi_{BP})}{(\theta_{NN} - \theta_{BN}) + (\varphi_{BN} - \varphi_{NN})}$$

From the inequality  $\frac{b}{a} > \frac{d}{c} \implies \frac{b}{a} > \frac{b+d}{a+c} > \frac{d}{c}$ , ( $a, b, c, d > 0$ ), we have:

$$\begin{aligned}
 & \frac{(\theta_{PP} - \theta_{BP}) + (\lambda_{BP} - \lambda_{PP})}{(\theta_{BN} - \theta_{PN}) + (\lambda_{PN} - \lambda_{BN})} \\
 > & \frac{(\theta_{PP} - \theta_{BP}) + (\lambda_{BP} - \lambda_{PP}) + (\theta_{BP} - \theta_{NP}) + (\lambda_{NP} - \lambda_{BP})}{(\theta_{BN} - \theta_{PN}) + (\lambda_{PN} - \lambda_{BN}) + (\theta_{NN} - \theta_{BN}) + (\lambda_{BN} - \lambda_{NN})} \\
 = & \frac{(\theta_{PP} - \theta_{NP}) + (\lambda_{NP} - \lambda_{PP})}{(\theta_{NN} - \theta_{PN}) + (\lambda_{PN} - \lambda_{NN})} \\
 > & \frac{(\theta_{BP} - \theta_{NP}) + (\lambda_{NP} - \lambda_{BP})}{(\theta_{NN} - \theta_{BN}) + (\lambda_{BN} - \lambda_{NN})} \tag{12}
 \end{aligned}$$

This implies that  $0 \leq \beta < \gamma < \alpha \leq 1$ , and the parameter  $\gamma$  is no longer needed. In this case, after tie-breaking, the following simplified rules are obtained:

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

The proposed model is a natural application and extension of DTRS.

As two special cases of the model, one may consider either the revenue function or the cost function. For the revenue only model, we have:

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

For the cost only model, we have:

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

Note that in the profit  $3 \times 2$  matrix in Section 3, some of the values may be less than 0. It is reasonable to impose that the expected value of profit must be more than 0 in any realistic investment. By equation (6), we can directly make the following de-

isions without further computing:

- (P2) If  $\theta_{PP} - \varphi_{PP} \leq 0$  and  $\theta_{PN} - \varphi_{PN} \leq 0$ ,  $\mathcal{R}(a_P|[x]) \leq 0$ , we decide  $x \notin POS(X)$ ;
- (B2) If  $\theta_{BP} - \varphi_{BP} \leq 0$  and  $\theta_{BN} - \varphi_{BN} \leq 0$ ,  $\mathcal{R}(a_B|[x]) \leq 0$ , we decide  $x \notin BND(X)$ ;
- (N2) If  $\theta_{NP} - \varphi_{NP} \leq 0$  and  $\theta_{NN} - \varphi_{NN} \leq 0$ ,  $\mathcal{R}(a_N|[x]) \leq 0$ , we decide  $x \notin NEG(X)$ .

They represent additional semantics in investment decision. We can automatically obtain the threshold values  $\alpha$  and  $\beta$  by using of equation (11) and immediately generate three-way decision rules.

#### 4. A Case Study of Oil Investment

In this section, we illustrate the proposed model by a case study of the investment decision-making in oil exploration.

In a practical investment problem, whether executes a project or not is determined by the profit, which is called expected monetary value (EMV) in oil industry.<sup>13</sup> The EMV of an oil exploration project is based on two possible outcomes: no oil discovery and oil discovery. When no oil is discovered, the EMV of the project is negative, and vice versa.<sup>33</sup> The EMV usually depends on two aspects: one is the revenue, including oil reserves, the output at end of project, etc.; the other is the cost, including the exploration costs, the development costs, the prior investment fee, the opportunity cost, etc. One method to deal with this investment problem is to use decision tree to choose a project for which EMV is maximum, and the final decision is composed of drilling and not drilling. Methodologies based on preference theory, portfolio theory and option theory are also introduced to solve investment appraisal problem in the upstream oil industry.<sup>12</sup>

Yusgiantoro and Hsiao point out that there are three classifications for all oilfields: (i) a high potential basin, which is poorly explored and requires exploration and exploitation; (ii) a fair potential basin which contains marginal oil reserves and requires intensive exploration; and (iii) poor potential hydrocarbon, which may or may not contain oil

Table 1. Avenues and costs of eight types of oilfield (Unit: million US \$)

| $O$   | $\theta_{PP}$ | $\theta_{BP}$ | $\theta_{NP}$ | $\theta_{PN}$ | $\theta_{BN}$ | $\theta_{NN}$ | $\varphi_{PP}$ | $\varphi_{BP}$ | $\varphi_{NP}$ | $\varphi_{PN}$ | $\varphi_{BN}$ | $\varphi_{NN}$ |
|-------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $o_1$ | 5             | 5             | 0             | 0             | 0.5           | 1             | 1              | 2              | 3              | 1              | 0.5            | 0              |
| $o_2$ | 10            | 8             | 0             | 0             | 1             | 2             | 2              | 3              | 5              | 2              | 1              | 0              |
| $o_3$ | 12            | 5             | 0             | 0             | 3             | 6             | 3              | 5              | 8              | 4              | 0.5            | 0              |
| $o_4$ | 7             | 2             | 0             | 0             | 2             | 4             | 1              | 1.5            | 3              | 5              | 1              | 0              |
| $o_5$ | 20            | 10            | 0             | 0             | 7             | 12            | 4              | 7              | 10             | 7              | 3              | 0              |
| $o_6$ | 15            | 8             | 0             | 0             | 3             | 5             | 2              | 3              | 4              | 3              | 1              | 0              |
| $o_7$ | 8             | 4             | 0             | 0             | 10            | 15            | 2              | 4              | 5              | 4              | 3              | 0              |
| $o_8$ | 18            | 15            | 0             | 0             | 1             | 2             | 1              | 2              | 3              | 4              | 3              | 0              |

reserves.<sup>33</sup> This classification may be helpful in obtaining the revenue and cost functions. Macmillan argues that the oil investment decides should divide into three parts: drill, don't drill, and to defer a the decision by acquiring seismic data.<sup>12</sup> With the insights gained from these studies, we use the three-way decision model proposed in Section 3 to solve the problem.

We have a set of 2 states and a set of 3 actions for oil investment. The set of states is given by  $\Omega = \{X, X^c\}$  indicating that an oilfield has oil and does not have oil, respectively. With respect to the three-way decision, the set of actions is given by  $\mathcal{A} = \{a_P, a_B, a_N\}$ , where  $a_P$ ,  $a_B$ , and  $a_N$  represent investment, need further analysis and do not investment, respectively. There are 12 parameters in the model.  $\theta_{PP}$ ,  $\theta_{BP}$ ,  $\theta_{NP}$  and  $\varphi_{PP}$ ,  $\varphi_{BP}$ ,  $\varphi_{NP}$  denote the revenues and costs incurred for taking actions of investment, need further analysis and do not investment when an oilfield has oil;  $\theta_{PN}$ ,  $\theta_{BN}$ ,  $\theta_{NN}$  and  $\varphi_{PN}$ ,  $\varphi_{BN}$ ,  $\varphi_{NN}$  denote the revenues and costs incurred for taking actions of investment, need further analysis and do not investment when a field does not have oil. We assume,  $\theta_{NP} < \theta_{BP} \leq \theta_{PP}$ ,  $\theta_{PN} < \theta_{BN} \leq \theta_{NN}$ ,  $\varphi_{PP} \leq \varphi_{BP} < \varphi_{NP}$  and  $\varphi_{NN} \leq \varphi_{BN} < \varphi_{PN}$ .

Table 1 shows the revenues and costs for eight types of oilfield. The values in the table reflect the following semantics considerations:

- The quantify  $\theta_{\bullet P}$  is the revenues of reserves or output of oil;  $\theta_{NP} = 0$  when the project is not executed.
- The quantify  $\theta_{\bullet N}$  is the revenues when the oil-

field does not have oil. Following the idea of opportunity cost in economics,<sup>18</sup> namely, a person can just do one thing at one time, we consider the revenues/costs of a particular project versus other projects. The quantity  $\theta_{PN}$  denotes the choice of investment, whose revenue is zero;  $\theta_{BN}$  and  $\theta_{NN}$  denote the revenue from another project instead of the oil project.

- The quantify  $\varphi_{\bullet P}$  is the cost when the oilfield has oil,  $\varphi_{PP}$  denotes the cost of prior development fee and exploration fee, and  $\varphi_{BP}$  and  $\varphi_{NP}$  denotes the opportunity cost when one hesitates or does not invest the oil project.
- The quantify  $\varphi_{\bullet N}$  is the cost when the oilfield does not have oil;  $\varphi_{NN} = 0$  when the project is not executed.

From equation (11), we can directly compute the thresholds  $\alpha$  and  $\beta$  for each oilfield in Table 1. The results are summarized in Table 2.

Table 2. Values of  $\alpha$  and  $\beta$  for eight types of oilfield

|       | $\alpha$ | $\beta$ |
|-------|----------|---------|
| $o_1$ | 0.3333   | 0.2000  |
| $o_2$ | 0.4000   | 0.1667  |
| $o_3$ | 0.4194   | 0.3043  |
| $o_4$ | 0.5217   | 0.4615  |
| $o_5$ | 0.4583   | 0.3810  |
| $o_6$ | 0.3846   | 0.2500  |
| $o_7$ | 0.6471   | 0.6154  |
| $o_8$ | 0.2000   | 0.1111  |

It can be seen that different pairs of thresholds are obtained for different types of oilfield. Suppose now that  $Pr(X|o_i) = 0.35$  for all  $o_i \in O$ . By

Table 3. The decision rules when the conditional probability changes

| $Pr(X o_i)$ | POS( $X$ )                                   | BND( $X$ )          | NEG( $X$ )                                   |
|-------------|--|---------------------|--|
| 0.00        | $\emptyset$                                  | $\emptyset$         | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ |
| 0.05        | $\emptyset$                                  | $\emptyset$         | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ |
| 0.10        | $\emptyset$                                  | $\emptyset$         | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ |
| 0.15        | $\emptyset$                                  | $\{o_8\}$           | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$      |
| 0.20        | $\{o_8\}$                                    | $\{o_2\}$           | $\{o_1, o_3, o_4, o_5, o_6, o_7\}$           |
| 0.25        | $\{o_8\}$                                    | $\{o_1, o_2\}$      | $\{o_3, o_4, o_5, o_6, o_7\}$                |
| 0.30        | $\{o_8\}$                                    | $\{o_1, o_2, o_6\}$ | $\{o_3, o_4, o_5, o_7\}$                     |
| 0.35        | $\{o_1, o_8\}$                               | $\{o_2, o_3, o_6\}$ | $\{o_4, o_5, o_7\}$                          |
| 0.40        | $\{o_1, o_2, o_6, o_8\}$                     | $\{o_3, o_5\}$      | $\{o_4, o_7\}$                               |
| 0.45        | $\{o_1, o_2, o_3, o_6, o_8\}$                | $\{o_5\}$           | $\{o_4, o_7\}$                               |
| 0.50        | $\{o_1, o_2, o_3, o_5, o_6, o_8\}$           | $\{o_4\}$           | $\{o_7\}$                                    |
| 0.55        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_8\}$      | $\emptyset$         | $\{o_7\}$                                    |
| 0.60        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_8\}$      | $\emptyset$         | $\{o_7\}$                                    |
| 0.65        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.70        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.75        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.80        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.85        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.90        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 0.95        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |
| 1.00        | $\{o_1, o_2, o_3, o_4, o_5, o_6, o_7, o_8\}$ | $\emptyset$         | $\emptyset$                                  |

three-way decision rules (P1), (B1) and (N1), we have:  $POS(X) = \{o_1, o_8\}$ ,  $BND(X) = \{o_2, o_3, o_6\}$ , and  $NEG(X) = \{o_4, o_5, o_7\}$ .

The three regions are determined both by the conditional probabilities and the thresholds. In Table 3, we list the three regions when the probability changes for 0.00 to 1.00 with a step size of 0.05.

### 5. Conclusions

A three-way decision approach is introduced into investment based on the decision-theoretic rough set model. The three-way rules generated by the positive, boundary and negative regions represent the decisions of investment, deferment, and non-investment. Based on a pair of a revenue function and a cost function, the Bayesian decision procedure is used to systematically compute the required parameters. An example of oil investment is used to show the difference between the three different types of rules. Our results enhance an understanding of

the decision-theoretic rough set model and broaden its applications in investment decision-making.

### References

1. Duda, R., Hart, P.: Pattern Classification and Scene Analysis. Wiley Press, New York. (1973)
2. Forster, M.: Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology*, 44, 205-231. (2000)
3. 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)
4. Herbert, J., Yao, J.T.: Learning optimal parameters in decision-theoretic rough sets. In: *Proceedings of Rough Sets and Knowledge Technology (RSKT'09)*, LNAI 5589, 610-617. (2009)
5. Herbert, J., Yao, J.T.: Criteria for choosing a rough set model. *Computer and Mathematics with Application*, 57, 908-918. (2009)
6. Herbert, J., Yao, J.T.: Game-Theoretic rough sets. *Fundamenta Informaticae*, to appear. (2010)
7. Li, H., Zhou, X.: Risk Decision Making Based on Decision-theoretic Rough Set: A Multi-view Decision



- Model. *International Journal of Computational Intelligence Systems*, to appear. (2010)
8. Li, Y., Zhang, C., Swan, J.: An information filtering model on the Web and its application in JobAgent. *Knowledge-Based Systems*, 13, 285-296. (2000)
  9. Liu, D., Li, H., Zhou, X.: Two decades' research on decision-theoretic rough sets. In: *Proceeding of 9th IEEE International Conference on Cognitive Informatics*, 968-973. (2010)
  10. Liu, D., Li, T., Hu, P., Li, H.: Multiple-category classification with decision-theoretic rough sets. In: *Proceeding of 5th International Conference on Rough Set and Knowledge Technology*, 697-704. (2010)
  11. Liu, D., Yao, Y. Y., Li, T.: Three-way decision-theoretic rough sets. *Journal of Computer Science*, to appear, in Chinese. (2010)
  12. Macmillan, F.: *Risk, Uncertainty and Investment Decision-Making in the Upstream Oil and Gas Industry*. The Ph.D thesis at University of Aberdeen, UK. (2000)
  13. Newendorp, P.: *Decision Analysis for Petroleum Economics*. Second edition, Planning Press, Aurora Colorado. (1996)
  14. Pauker, S., Kassirer, J.: The threshold approach to clinical decision making. *The New England Journal of Medicine*, 302, 1109-1117. (1980)
  15. Pawlak, Z.: Rough sets. *International Journal of Computer and Information Science*, 11, 341-356. (1982)
  16. Pawlak, Z.: *Rough Sets, Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers Press, Dordrecht. (1991)
  17. Pawlak, Z., Wong, S.K.M., Ziarko, W.: Rough sets: probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, 29, 81-95. (1988)
  18. Robert, S., Daniel. L.: *Microeconomics*. The 6th Edition, Printice-Hall Press, New Jersey. (2005)
  19. Ślęzak, D.: Rough sets and Bayes factor. *LNCS Transactions on Rough Sets III*, LNCS 3400, 202-229. (2005)
  20. Ślęzak, D., Wróblewski, J., Eastwood, V., Synak, P.: Brighthouse: an analytic data warehouse for ad-hoc queries. *Proceedings of the VLDB Endowment*, 1, 1337-1345. (2008)
  21. Wald, A.: Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16, 117-186. (1945)
  22. Wong, S.K.M., Ziarko, W.: Comparison of the probabilistic approximate classification and the fuzzy set model. *Fuzzy Sets and Systems*, 21, 357-362. (1987)
  23. Woodward, P., Naylor, J.: An application of Bayesian methods in SPC. *The Statistician*, 42, 461-469. (1993)
  24. Yao, Y.Y.: Probabilistic approaches to rough sets. *Expert Systems*, 20, 287-297. (2003)
  25. Yao, Y.Y.: Decision-theoretic rough set models. In: *Proceeding of RSKT 2007*, LNAI 4481, 1-12. (2007)
  26. Yao, Y.Y.: Probabilistic rough set approximations. *International Journal of Approximate Reasoning*, 49, 255-271. (2008)
  27. Yao, Y.Y.: Three-way decision: an interpretation of rules in rough set theory. In: *Proceeding of RSKT 2009*, LNAI 5589, 642-649. (2009)
  28. Yao, Y.Y.: Two semantic issues in a probabilistic rough set model. *Fundamenta Informaticae*, to appear. (2009)
  29. Yao, Y.Y.: Three-way decisions with probabilistic rough sets. *Information Sciences*, 180, 341-353. (2010)
  30. Yao, Y.Y., Zhou, B.: Naive bayesian rough sets. In: *Proceeding of 5th International Conference on Rough Set and Knowledge Technology*, 729-737. (2010)
  31. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximating concepts. *International Journal of Man-machine Studies*. 37(6): 793-809. (1992)
  32. Yao, Y.Y., Wong, S.K.M., Lingras, P.: A decision-theoretic rough set model. In: *ISMIS 1990*, 17-25. (1990)
  33. Yusgiantoro, P., Hsiao, F.: Production-sharing contracts and decision-making in oil production. *Energy Economics*, 10, 245-256. (1993)
  34. Zhao, W., Zhu, Y.: An email classification scheme based on decision-theoretic rough set theory and analysis of email security. In: *Proceeding of 2005 IEEE Region 10 TENCON*, Digital Object Identifier: 10.1109/TENCON.2005.301121. (2005)
  35. Zhou, B., Yao, Y.Y., Luo, J.: A three-way decision approach to email spam filtering. In: *Proceedings of the 23th Canadian AI 2010*, 28-39. (2010)
  36. Zhou, B., Yao, Y.Y.: In search for effective granularity with DTRS. In: *Proceeding of 9th IEEE International Conference on Cognitive Informatics*, 464-470. (2010)
  37. Zhou, X., Li, H.: A multi-view decision model based on decision-theoretic rough set. In: *Proceedings of Rough Sets and Knowledge Technology (RSKT'09)*, LNAI. 5589, 650-657. (2009)
  38. Ziarko, W.: Variable precision rough set model. *Journal of Computer and System Sciences*. 46, 39-59. (1993)