A New Discriminant Analysis Approach under Decision-Theoretic Rough Sets

Dun Liu[1], Tianrui Li[2], and Decui Liang[1]

[1] School of Economics and Management, Southwest Jiaotong University
Chengdu 610031, P.R. China
newton83@163.com, decuiliang@126.com
[2] School of Information Science and Technology, Southwest Jiaotong University
Chengdu 610031, P.R. China
trli@swjtu.edu.cn

Abstract. Discriminant analysis is an effective methodology to deal with the classification problem. However, most common methods including binary logistic regression in discriminant analysis rarely consider the semantics explanations such as losses or costs in decision rules. From the idea of three-way decisions in decision-theoretic rough sets (DTRS), we propose a new discriminant analysis approach by combining DTRS and binary logistic regression. DTRS is utilized to systematically calculate the corresponding thresholds with Bayesian decision procedure. Meanwhile, the binary logistic regression is employed to compute the conditional probability of three-way decisions. An empirical study validates the reasonability and effectiveness of the proposed approach.

Keywords: Decision-theoretic rough sets, discriminant analysis, binary logistic regression, decision making.

1 Introduction

Discriminant analysis is a multivariate statistical method for classifications. As an approach for classifying a set of observations into predefined classes with respect to several variables, it requires the data in an information system satisfies the conditions that the dependent variable is nonmetric and independent variables are metric [4, 12–14]. Usually, the purpose of discriminant analysis is to predict the classification for the observations by using discriminant functions and rules. Nowadays, many common discriminant methodologies, such as binary logistic regression, distance discriminant analysis, fisher discriminant analysis, Bayesian discriminant analysis and stepwise discriminant analysis, have been successfully and widely used in many fields, such as electricity loads [9], face recognition [2], feature extraction [3], etc.

A decision is typically made under some risk and uncertainty. The misclassification may cause losses or costs. However, most of discriminant methodologies rarely consider losses or costs for misclassification, except for Bayesian discriminant analysis. The Bayesian discriminant analysis method only takes account

J.T. Yao et al. (Eds): RSKT 2011, LNCS 6954, pp. 476–485, 2011.

for the misclassification costs in two scenarios, namely, the incorrect acceptance costs and incorrect rejection costs. This method does not consider the deferment scenario and it is actually regarded as a two-way decision [15].

Rough sets describe a set of a concept using the lower approximation and upper approximation [10]. The two approximations divide the universe into three pairwise disjoint regions: the positive region, boundary region and negative region [7]. With respect to the three regions, Yao proposed a three-way decision including the positive, boundary and negative rules. Positive rules make decisions of acceptance, negative rules make decisions of rejection, and boundary rules make deferred or non-committed decisions [15]. The three-way decision depends on a pair of thresholds and conditional probability [7]. By considering the tolerance of errors in the three-way decision, Yao et al. introduced Bayesian decision procedure to propose decision-theoretic rough sets (DTRS) [15], and the pair of thresholds can be directly calculated by minimizing the decision cost with Bayesian theory. He pointed out that the two-way decision is the special case of a three-way decision which considers the deferment scenario [15]. In addition, Yao and Zhou further proposed the naive Bayesian rough sets. The conditional probability is estimated by using the Bayes theorem with naive probabilistic independence assumption for attributes [16], but this approach can only deal with the discrete data. In real applications, the continuous and discrete data may coexist in information systems. Binary logistic regression provides a way to compute the conditional probability in this situation. In the view of semantics, DTRS is utilized to systematically calculate the corresponding thresholds by considering cost or loss. Hence, DTRS can reasonably explain the threshold of binary logistic regression, and is complementary of binary logistic regression. With comparison of Bayesian discriminant analysis, DTRS also adds the deferred decision into the two-way decision. DTRS has been successfully utilized in some learning methods, such as clustering [5] and naive Bayesian classifier [16]. In this paper, we try to introduce the three-way decision in DTRS to discriminant analysis and propose a new discriminant analysis method.

The remainder of this paper is organized as follows: Section 2 provides the basic concepts of discriminant analysis and DTRS. A new discriminant analysis approach under DTRS is proposed in Section 3. Then, a case study of corporate failure prediction is given to illustrate our approach in Section 4. Section 5 concludes the paper and outlines the future work.

2 Preliminaries

Basic concepts, notations and results of the discriminant analysis and DTRS are briefly reviewed in this section [4, 6, 8, 15].

2.1 Discriminant Analysis and Binary Logistic Regression

This subsection introduces the discriminant analysis to calculate the conditional probability for objects in an information table. As a common linear discriminant analysis approach, binary logistic regression method is utilized to deal with the binary classification problem because it can directly compute the probability of occurrence of event.

In binary logistic regression model, the value domain of dependent variable has two categories: occurrence ($d = 1$) and non-occurrence ($d = 0$). Suppose x is an event, the probability of occurrence is denoted by $Pr((d = 1)|x)$, and $Pr((d = 0)|x) = 1 - P((d = 1)|x)$ denotes non-occurrence. The logistic function transformation of $Pr((d = 1)|x)$ is known as the logit transformation as follows.

$$\theta(Pr((d = 1)|x)) = logit(Pr((d = 1)|x)) = ln(\frac{Pr((d = 1)|x)}{1 - Pr((d = 1)|x)}) \quad (1)$$

where the expression of $\theta(Pr((d = 1)|x))$ can be used by a linear function of independent variables denoted by $a_1, a_2, \cdots, a_k$, and (1) can rewrite as:

$$\theta(Pr((d = 1)|x)) = ln(\frac{Pr((d = 1)|x)}{1 - Pr((d = 1)|x)}) = \beta_0 + \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k (2)$$

where β_0 denotes the intercept, and $\beta_1, \beta_2, \cdots, \beta_k$ denote the regression coefficients of $a_1, a_2, \cdots, a_k$ respectively. With the above analysis, the probability of occurrence of event $Pr((d = 1)|x)$ can be expressed as follows:

$$Pr((d = 1)|x) = \frac{e^{\theta(Pr((d=1)|x))}}{1 + e^{\theta(Pr((d=1)|x))}} = \frac{e^{\beta_0 + \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k}}{1 + e^{\beta_0 + \beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k}} \quad (3)$$

The discriminant rules are directly generated as follows.

$$d = \begin{cases} 1 & Pr((d = 1)|x) > 0.5; \\ 0 & Pr((d = 1)|x) \leq 0.5. \end{cases}$$

2.2 Decision-Theoretic Rough Sets

In this subsection, we briefly introduce DTRS model, especially in the aspect of the thresholds of the three-way decision.

For the Bayesian decision procedure, the DTRS model is composed of 2 states and 3 actions [15, 16]. The set of states is given by $\Omega = \{C, \neg C\}$ indicating that an object is in C and not in C, respectively. And the set of actions is given by $\mathcal{A} = \{a_P, a_B, a_N\}$, where a_P, a_B, and a_N represent the three actions in classifying an object x, namely, deciding $x \in \mathrm{POS}(C)$, deciding x should be further investigated $x \in \mathrm{BND}(C)$, and deciding $x \in \mathrm{NEG}(C)$, respectively. The loss function regarding the risk or cost of actions in different states is given by the 3×2 matrix:

	C (P)	$\neg C$ (N)
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

In the matrix, λ_{PP}, λ_{BP} and λ_{NP} denote the losses incurred for taking actions of a_P, a_B and a_N, respectively, when an object belongs to C. Similarly, λ_{PN}, λ_{BN} and λ_{NN} denote the losses incurred for taking the same actions when the object belongs to $\neg C$. $Pr(C|[x])$ is the conditional probability of an object x belonging to C given that the object is described by its equivalence class $[x]$. For an object x, the expected loss $R(a_i|[x])$ associated with taking the individual actions can be expressed as:

$$\begin{aligned} R(a_P|[x]) &= \lambda_{PP}Pr(C|[x]) + \lambda_{PN}Pr(\neg C|[x]), \\ R(a_B|[x]) &= \lambda_{BP}Pr(C|[x]) + \lambda_{BN}Pr(\neg C|[x]), \\ R(a_N|[x]) &= \lambda_{NP}Pr(C|[x]) + \lambda_{NN}Pr(\neg C|[x]). \end{aligned} \quad (4)$$

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

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

Since $Pr(C|[x]) + Pr(\neg C|[x]) = 1$, we simplify the rules based only on the probability $Pr(C|[x])$ and the loss function. By considering a reasonable kind of loss functions with $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$, the decision rules (P)-(N) can be expressed concisely as:

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

The thresholds values α, β, γ are given by: $\alpha = \frac{(\lambda_{PN}-\lambda_{BN})}{(\lambda_{PN}-\lambda_{BN})+(\lambda_{BP}-\lambda_{PP})}$, $\beta = \frac{(\lambda_{BN}-\lambda_{NN})}{(\lambda_{BN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{BP})}$ and $\gamma = \frac{(\lambda_{PN}-\lambda_{NN})}{(\lambda_{PN}-\lambda_{NN})+(\lambda_{NP}-\lambda_{PP})}$.

In addition, as a well-defined boundary region, the conditions of rule (B) suggest that $\alpha > \beta$, that is, $\frac{(\lambda_{BP}-\lambda_{PP})}{(\lambda_{PN}-\lambda_{BN})} < \frac{(\lambda_{NP}-\lambda_{BP})}{(\lambda_{BN}-\lambda_{NN})}$. It implies $0 \leq \beta < \gamma < \alpha \leq 1$. In this case, after tie-breaking, the following simplified rules are obtained:

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

From (P1)-(N1), the threshold parameters α and β can be systematically calculated from the loss functions based on the Bayesian decision procedure, which gives us a solid theoretical basis.

3 The New Discriminant Analysis Approach under Decision-theoretic Rough Sets

As stated in Section 2, DTRS mainly focuses on the classification problem with 2 states $\Omega = \{C, \neg C\}$. Meanwhile, the binary logistic regression can also solve the two classification problem. Following the similarity of the two methodologies, we point out their differences in two aspects. On the one hand, the difference between DTRS and logistic regression is the former lends to a three-way decision and the latter lends to a two-way decision. One the other hand, the logistic regression can compute the conditional probability in an information table, and it is a complementary of DTRS. Therefore, we try to combine the two methodologies together and propose a new discriminant analysis approach under DTRS as follows.

The binary classification corresponds to two states of C and $\neg C$ in DTRS. Combined with logistic regression, C is replaced by $d = 1$ and $\neg C$ is replaced by $d = 0$ or $\neg(d = 1)$ simultaneously. Suppose an information table is composed of n objects $O = \{o_1, o_2, \cdots, o_n)$ and k attributes $A = \{a_1, a_2, \cdots, a_k)$. In discriminant analysis, an object o_i ($i = 1, 2, \cdots, n$) is described by it's corresponding variable's values $a_{i1}, a_{i2}, \cdots, a_{ik}$. For simplicity, the conditional probability of the object o_i belonging to $d = 1$, is straightway expressed by $Pr((d = 1)|o_i)$. In fact, some objects may have different attribute-values in the information table or different prior information so that they may have different loss functions in the same state. In order to meet this fact, we will construct a loss function matrix for every object. For the object o_i, the loss function regarding the risk or cost of actions in different states is given by the 3×2 matrix according to expert experience and priori information.

	$d = 1$ (P)	$\neg(d = 1)$ (N)
a_P	λ^i_{PP}	λ^i_{PN}
a_B	λ^i_{BP}	λ^i_{BN}
a_N	λ^i_{NP}	λ^i_{NN}

where a_P, a_B, and a_N represent the three actions, namely, accepting the result of o_i belonging to $d = 1$, the result of o_i belonging to $d = 1$ should be further investigated, and rejecting the result of o_i belonging to $d = 1$, respectively. λ^i_{PP}, λ^i_{BP} and λ^i_{NP} denote the losses incurred for taking actions of "accepted", "further investigated" and "rejected", respectively, when o_i belongs to $d = 1$. Similarly, λ^i_{PN}, λ^i_{BN} and λ^i_{NN} denote the losses incurred for taking the same actions when o_i belongs to $\neg(d = 1)$. Furthermore, the thresholds are calculated according to the conditions of decision rules (P1)-(N1). With above conditions, the decision rules (P1i)-(N1i) of the object o_i can be expressed as:

(P1i) If $Pr((d = 1)|o_i) \geq \alpha_i$, decide $o_i \in \mathrm{POS}(d = 1)$;

(B1i) If $\beta_i < Pr((d = 1)|o_i) < \alpha_i$, decide $o_i \in \mathrm{BND}(d = 1)$;

(N1i) If $Pr((d = 1)|o_i) \leq \beta_i$, decide $o_i \in \mathrm{NEG}(d = 1)$.

The thresholds values α_i, β_i are given by:

$$\alpha_i = \frac{(\lambda^i_{PN} - \lambda^i_{BN})}{(\lambda^i_{PN} - \lambda^i_{BN}) + (\lambda^i_{BP} - \lambda^i_{PP})},$$

$$\beta_i = \frac{(\lambda^i_{BN} - \lambda^i_{NN})}{(\lambda^i_{BN} - \lambda^i_{NN}) + (\lambda^i_{NP} - \lambda^i_{BP})}.$$

Similarly, the conditional probability $Pr(\neg(d = 1)|o_i)$ has the analogous decision rules. The conditional probability $Pr((d = 1)|o_i)$ can be calculated according to the logistic regression equation as follows.

$$Pr((d = 1)|o_i) = \frac{e^{\beta_0+\beta_1 a_{i1}+\beta_2 a_{i2}+\cdots+\beta_k a_{ik}}}{1 + e^{\beta_0+\beta_1 a_{i1}+\beta_2 a_{i2}+\cdots+\beta_k a_{ik}}}. \tag{5}$$

In the following, the new discriminant analysis approach consists of 5 steps shown in Figure 1.

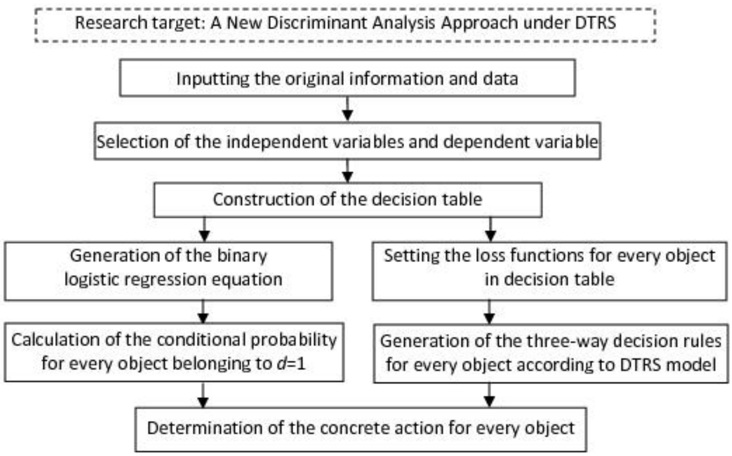

Fig. 1. A new discriminant analysis approach with DTRS

Step 1: Selection of the independent variables and dependent variables, then construct the original information table.

Step 2: Application of binary logistic regression to obtain the logistic regression equation based on the original information table.

Step 3: Calculation of the conditional probability for each object belonging to $d = 1$ according to logistic regression equation, *i.e.*, $Pr((d = 1)|o_i)$ $(i = 1, 2, \cdots, n)$.

Step 4: Generation of decision rules for each object according to DTRS model. For object o_i $(i = 1, 2, \cdots, n)$, we set the loss functions of different actions in two states $(d = 1, \neg(d = 1))$ according to expert experience and prior information, and calculate the thresholds α_i and β_i according to (P1i)-(N1i).

Step 5: Determination of concrete action for each object. For object o_i, we compare the conditional probability $Pr((d = 1)|o_i)$ with the thresholds values α_i and β_i. Then, the concrete action of o_i is determined.

Remark 1. From Fig. 1, DTRS is utilized to generate decision rules and calculate the thresholds for the three-way decision. On the other hand, discriminant analysis brings us a new way on estimating the condition probability of the three-way decision.

4 An Illustration

In this section, a case study of corporate failure prediction [1] is given to explain our approach proposed in Section 3. The data of this case includes 30 failed and 30 non-failed UK companies. Beynon and Peel [1] listed a total of 12 independent variables (or condition attributes) for potential rule generation. The dependent variable (or decision attribute) is denoted as company status (d), which is coded zero for a non-failed firm ($d = 0$) and unity for a failed company ($d = 1$).

In order to validate the significance of logistic regression equation, we utilize SPSS 16.0 to run binary logistic regression for the original information table in [1]. By using backward and forward method for logistic regression, we select 3 independent variables from 12 variables and denote them as a_1 (current liabilities/total assets), a_2 (number of days between account year end and the date the annual report and accounts were filed at company registry) and a_3 (whether the company auditor is a Big6 auditor, coded 1 if yes, coded 0 if no). Table 1 lists the original information for the 3 attributes, and the coefficients in logistic regression equation is shown in Table 2.

Table 1. The original information table of UK companies

O	a_1	a_2	a_3	d	O	a_1	a_2	a_3	d	O	a_1	a_2	a_3	d	O	a_1	a_2	a_3	d
o_1	0.6233	96	0	0	o_{16}	0.3999	271	1	0	o_{31}	0.7034	172	0	1	o_{46}	0.4824	295	0	1
o_2	1.2218	287	1	0	o_{17}	0.1212	197	0	0	o_{32}	0.7081	265	1	1	o_{47}	0.2986	303	1	1
o_3	0.3307	64	1	0	o_{18}	0.2761	303	1	0	o_{33}	1.0139	301	0	1	o_{48}	0.7437	303	0	1
o_4	0.4829	286	0	0	o_{19}	0.5014	302	1	0	o_{34}	0.7585	309	1	1	o_{49}	0.4976	362	1	1
o_5	0.3786	301	1	0	o_{20}	0.406	421	1	0	o_{35}	0.56	243	0	1	o_{50}	0.6672	289	0	1
o_6	0.3548	314	1	0	o_{21}	0.5458	77	0	0	o_{36}	0.7772	393	0	1	o_{51}	0.5726	301	0	1
o_7	0.7018	249	1	0	o_{22}	0.3983	192	0	0	o_{37}	1.2024	360	0	1	o_{52}	0.7586	301	0	1
o_8	0.4347	55	1	0	o_{23}	0.7412	304	1	0	o_{38}	0.7005	296	1	1	o_{53}	0.6596	293	1	1
o_9	0.6715	298	0	0	o_{24}	0.3426	191	0	0	o_{39}	0.6377	295	0	1	o_{54}	0.5916	179	0	1
o_{10}	0.1401	241	1	0	o_{25}	0.1538	117	1	0	o_{40}	0.4767	233	0	1	o_{55}	0.5261	283	0	1
o_{11}	0.4533	168	1	0	o_{26}	0.6198	160	1	0	o_{41}	1.4865	301	0	1	o_{56}	0.7112	207	1	1
o_{12}	0.5297	36	0	0	o_{27}	0.4034	264	1	0	o_{42}	0.3883	265	1	1	o_{57}	0.5742	301	0	1
o_{13}	0.7866	212	1	0	o_{28}	1.1831	183	1	0	o_{43}	0.8516	86	1	1	o_{58}	0.8618	153	1	1
o_{14}	0.6301	188	1	0	o_{29}	0.6965	117	0	0	o_{44}	0.7585	297	0	1	o_{59}	0.7444	243	1	1
o_{15}	0.5072	268	0	0	o_{30}	0.7883	147	1	0	o_{45}	0.656	144	0	1	o_{60}	0.732	301	0	1

From Table 2, we find the coefficients of independent variables and constant are significance. The binary logistic equation is shown in (6).

$$Pr((d = 1)|o_i) = \frac{e^{-3.861+3.566a_1+0.010a_2-1.370a_3}}{1 + e^{-3.861+3.566a_1+0.010a_2-1.370a_3}}. \tag{6}$$

where $Pr((d = 1)|o_i)$ denotes the conditional probability of firm o_i belonging to a failed company. In equation (6), the values of coefficients of a_1 and a_2 are positive, that is, the larger values of a_1 and a_2, the higher probability the firm will fail. The opposite situation happens in a_3 because the value of coefficient of a_3 is negative. The importance of the three independent variables is ordered by

Table 2. The binary logistic regression results

Coefficients	B	$S.E.$	$Wald$	df	$Sig.$	$Exp(B)$
a_1	3.566	1.574	5.137	1	0.023	35.375
a_2	0.010	0.004	5.817	1	0.016	1.010
a_3	-1.370	0.629	4.745	1	0.029	0.254
Constant	-3.861	1.490	6.718	1	0.010	0.021

a_2, a_1 and a_3 according to the Wald values of the three variables. The predicted conditional probabilities and predictions for 60 UK companies based on the equation (6) and the discriminant rules in Section 2.1 are shown in Table 3, and d denotes the true state for these companies.

Table 3. The predicted conditional probabilities and predictions for 60 UK companies

O	d	Predicted group	Predicted $Pr((d=1)\|o_i)$	O	d	Predicted group	Predicted $Pr((d=1)\|o_i)$	O	d	Predicted group	Predicted $Pr((d=1)\|o_i)$
o_1	0	0	0.3368	o_{21}	0	0	0.2416	o_{41}	1	1	0.9885
o_2	0	1	0.8806	o_{22}	0	0	0.3731	o_{42}	1	0	0.2325
o_3	0	0	0.0319	o_{23}	0	1	0.6118	o_{43}	1	0	0.2086
o_4	0	1	0.6734	o_{24}	0	0	0.3257	o_{44}	1	1	0.8601
o_5	0	0	0.2956	o_{25}	0	0	0.0290	o_{45}	1	0	0.4799
o_6	0	0	0.3051	o_{26}	0	0	0.1947	o_{46}	1	1	0.6925
o_7	0	0	0.4412	o_{27}	0	0	0.2404	o_{47}	1	0	0.2435
o_8	0	0	0.0419	o_{28}	0	1	0.6941	o_{48}	1	1	0.8610
o_9	0	1	0.8200	o_{29}	0	0	0.4486	o_{49}	1	1	0.5415
o_{10}	0	0	0.0895	o_{30}	0	0	0.2791	o_{50}	1	1	0.8039
o_{11}	0	0	0.1264	o_{31}	1	1	0.5912	o_{51}	1	1	0.7674
o_{12}	0	0	0.3519	o_{32}	1	0	0.4866	o_{52}	1	1	0.8649
o_{13}	0	0	0.4245	o_{33}	1	1	0.9409	o_{53}	1	1	0.5134
o_{14}	0	0	0.2493	o_{34}	1	1	0.6379	o_{54}	1	1	0.5101
o_{15}	0	1	0.6525	o_{35}	1	1	0.6383	o_{55}	1	1	0.7001
o_{16}	0	0	0.2511	o_{36}	1	1	0.9450	o_{56}	1	0	0.3491
o_{17}	0	0	0.1889	o_{37}	1	1	0.9825	o_{57}	1	1	0.7684
o_{18}	0	0	0.2290	o_{38}	1	1	0.5571	o_{58}	1	0	0.3483
o_{19}	0	0	0.3964	o_{39}	1	1	0.7967	o_{59}	1	0	0.4639
o_{20}	0	1	0.6060	o_{40}	1	1	0.5426	o_{60}	1	1	0.8535

The underlined values in Table 3 denote the inconsistent issue between predicted classification and true state for a firm. A natural idea from these inconsistences illustrates that the predictions with binary logistic regression need to reconsider. In our discussions, DTRS is utilized to construct the corresponding thresholds through setting the loss functions. For every firm, we have three scenarios for determining the prediction by comparing the conditional probability generated by logistic regression and the thresholds (α, β), namely, accept the prediction, reject it or defer it. Our new discriminant approach provides a DTRS based strategy to revise the predictions with binary logistic regression.

For simplicity, we select four firms o_2, o_3, o_{32} and o_{54} to illustrate our discriminant approach. The loss functions are carefully estimated by corresponding experts according to the experience and prior information. The calculating results for the four firms are listed in Table 4.

Table 4. The loss functions and calculating results of the four firms (u is a unit cost)

O	λ^i_{PP}	λ^i_{BP}	λ^i_{NP}	λ^i_{PN}	λ^i_{BN}	λ^i_{NN}	α_i	β_i
o_2	0	$8u$	$14u$	$16u$	$4u$	0	0.6000	0.4000
o_3	0	$4u$	$10u$	$25u$	$7u$	0	0.8182	0.5385
o_{32}	0	$6.5u$	$13u$	$18u$	$5u$	0	0.6667	0.4348
o_{54}	0	$5.5u$	$12.5u$	$19.5u$	$5.5u$	0	0.7179	0.4400

In Table 4, we set $\lambda^i_{PP} = 0$, $\lambda^i_{NN} = 0$ $(i = 1, 2, \cdots, n)$ by considering the fact that there is no cost when doing a right decision. Since the importance of independent variables in (6) is ordered by a_2, a_1 and a_3, and the values in o_2 play higher in a_1 and a_2 among four firms, we put higher estimations in λ^2_{BP} and λ^2_{NP}, but lower estimations in λ^2_{PN} and λ^2_{BN}. However, the opposite situation happens in o_3.

Finally, we compare the predicted probabilities in Table 3 with their corresponding thresholds in Table 4 among the four firms. For the firm o_2, $Pr((d = 1)|o_2) = 0.8806 > 0.6000 = \alpha_2$, we take the action a_P and accept the prediction. In this case, we agree the predicted result with binary logistic regression, namely, o_2 is a failed company and we disagree the state of o_2 in Table 1. Similarly, we can discriminate the three other firms according to these criteria in our new approach. For the firm o_3, $Pr((d = 1)|o_3) = 0.0319 < 0.538 = \beta_3$, we take the action a_N and accept the prediction. In this case, we both agree the predicted result with binary logistic regression and the state in Table 1. We discriminate o_3 is a non-failed company. For the firm o_{32} and o_{54}, the values of $Pr((d = 1)|o_{32})$ and $Pr((d = 1)|o_{54})$ are between their corresponding two thresholds. Hence, we need take the action a_B for these two firms and cannot discriminate whether they are failed companies or not at present. We need to collect more information for the two companies to make a final decision. The temporary deferment decision for o_{32} and o_{54} may reduce the losses or costs generated by the misclassifications in real-life applications.

5 Conclusions

In this paper, a new discriminant analysis approach was proposed by combining with DTRS and binary logistic regression. The two thresholds α and β generated by DTRS were utilized to revise the predictions with discriminant analysis (binary logistic regression). A case study of corporate failure prediction validated the rationality of the proposed method. Our future research work will focus on the extension of the proposed method to multiple-category discriminant analysis.

Acknowledgements This work is partially supported by the National Science Foundation of China (No. 60873108), and the Scientific Research Foundation of Graduate School of Southwest Jiaotong University (200906), China.

References

1. Beynon, M.J., Peel, M.J.: Variable Precision Rough Set Theory and Data Discretisation: an Application to Corporate Failure Prediction. Omega 29, 561–576 (2001)
2. Dai, G., Yeung, D.Y., Qian, Y.T.: Face Recognition Using a Kernel Fractional-step Discriminant Analysis Algorithm. Pattern Recognition 40(1), 229–243 (2007)
3. Gu, Z.H., Yang, J., Zhang, L.: Push-Pull Marginal Discriminant Analysis for Feature Extraction. Pattern Recognition Letters 31(15), 2345–2352 (2010)
4. He, X.Q.: Mutivariable Statistics Analysis. Renmin University Press (2007)
5. Lingras, P., Chen, M., Miao, D.Q.: Rough Cluster Quality Index Based on Decision Theory. IEEE Transactions on Knowledge and Data Engineering 21(7), 1014–1026 (2009)
6. Liu, D., Li, H.X., Zhou, X.Z.: Two Decades' Research on Decision-theoretic Rough Sets. In: Proceeding of ICCI 2010, pp. 968–973. IEEE Press, New York (2010)
7. Liu, D., Yao, Y.Y., Li, T.R.: Three-way Investment Decisions with Decision-theoretic Rough Sets. Inter. J. of Comput. Intell. Sys. 4, 66–74 (2011)
8. Liu, D., Li, T.R., Ruan, D.: Probabilistic Model Criteria with Decision-theoretic Rough Sets. Information Sciences 181, 3709–3722 (2011)
9. Pai, P.F., Chen, T.C.: Rough Set Theory with Discriminant Analysis in Analyzing Electricity Loads. Expert Systems with Applications 36(5), 8799–8806 (2009)
10. Pawlak, Z.: Rough Sets. International Journal of Computer and Information Science 11, 341–356 (1982)
11. Pawlak, Z.: Rough Sets. In: Theoretical Aspects of Reasoning about Data. Kluwer Academic Publishers Press, Dordrecht (1991)
12. Pires, A.M., Branco, J.A.: Projection-pursuit Approach to Robust Linear Discriminant Analysis. Journal of Multivariate Analysis 101, 2464–2485 (2010)
13. Sueyoshi, T.: DEA-Discriminant Analysis: Methodological Comparison among Eight Discriminant Analysis Approaches. European Journal of Operational Research 169, 247–272 (2006)
14. Discriminant Analysis, `http://www.referenceforbusiness.com/encyclopedia/Dev`
15. Yao, Y.Y.: Three-way Decisions with Probabilistic Rough Sets. Information Sciences 180, 341–353 (2010)
16. 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)