Incorporating logistic regression to decision-theoretic rough sets for classifications

Dun Liu\textsuperscript{a,*}, Tianrui Li\textsuperscript{b}, Decui Liang\textsuperscript{a}

\textsuperscript{a} School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, Sichuan, PR China
\textsuperscript{b} School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, Sichuan, PR China

\textbf{ARTICLE INFO}

\textbf{Article history:}
Available online 16 March 2013

\textbf{Keywords:}
Decision-theoretic rough sets
Binary logistic analysis
Multivariate logistic regression
Decision making

\textbf{ABSTRACT}

Logistic regression analysis is an effective approach to the classification problem. However, it may lead to high misclassification rate in real decision procedures. Decision-Theoretic Rough Sets (DTRS) employs a three-way decision to avoid most direct misclassification. We integrate logistic regression and DTRS to provide a new classification approach. On one hand, DTRS is utilized to systematically calculate the corresponding thresholds with Bayesian decision procedure. On the other hand, logistic regression is employed to compute the conditional probability of the three-way decision. The empirical studies of corporate failure prediction and high school program choices' prediction validate the rationality and effectiveness of the proposed approach.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Logistic regression analysis is a multivariate statistical method for classifications. As a representative discriminant analysis approach for classifying a set of observations into predefined classes with respect to several variables, it requires the data in an information system satisfy the conditions that the dependent variable is nonmetric and independent variables are metric [9,37,43,47]. In statistics, logistic regression is a classification method that fits data to a logistic function. It is used for predicting the outcome of a categorical criterion variable (a variable that can take on a limited number of categories) based on one or more predictor (independent or explanatory) variables. The probabilities describing the possible outcome of a single trial are modeled, as a function of explanatory variables, using a logistic function [49]. Because of no assumption regarding the distribution of the predictor variables, logistic regression is relatively robust, flexible and easily used, and it lends itself to a meaningful interpretation. In practice, logistic regression is usually used as a classifier, namely, logistic regression classifier, for probabilistic binary or multivariate classification. The criterion for selecting the category in logistic regression lies on the highest probability which is generated from a logistic function.

However, logistic regression may lead to high misclassification during the decision procedure [8]. For example, suppose an object $x$ with three categories $C_1$, $C_2$, $C_3$, where $P(C_1|x) = 0.34$, $P(C_2|x) = 0.33$ and $P(C_3|x) = 0.33$. We get $x \in C_1$ according to the maximum discriminant criterion, but this criterion may lead to the misclassification for $x$ with a high probability because of $P(C_1|x) = 0.34$ is not higher enough. For simplicity, we briefly list two drawbacks of logistic regression in real decision problems.

\textbf{Problem 1:} No consideration on the losses/costs of the misclassifications.

A decision is typically made under some risks and uncertainty. The misclassification may cause losses or costs. However, logistic regression does not consider losses or costs for misclassification.

\* Corresponding author.

E-mail addresses: newton83@163.com (D. Liu), trli@swjtu.edu.cn (T. Li), decuiliang@126.com (D. Liang).
**Problem 2:** No consideration on the deferment decision for classification.

In logistic regression, it only considers two actions (acceptance and rejection) in a real decision problem. As well, there exist two types of misclassification scenarios in logistic regression, namely, the incorrect acceptance and incorrect rejection. This method does not consider the deferment scenario and it is actually regarded as a two-way decision [60].

To solve the two problems in logistic regression, we introduce Decision-Theoretic Rough Sets (DTRS) in our following discussions. DTRS, proposed by Yao in the early 1990s [53,54], has become a powerful decision making method and attracted more and more attention in the last 5 years. As a quantitative probabilistic extension of the qualitative classical rough set model [64], DTRS introduces Bayesian decision procedure and loss functions to rough sets. In DTRS, the pair of thresholds $\alpha$ and $\beta$, which are used to describe the tolerance of errors in Probabilistic Rough Sets (PRS), can be directly calculated by minimizing the decision cost with Bayesian theory. It gives a sound semantics in practical applications with minimum decision risks. Based on DTRS, Yao [60,62] further proposed the concept of a three-way decision. Obviously, the thresholds $\alpha$ and $\beta$ can divide the universe into three pairwise disjoint regions: the positive region, boundary region and negative region. The three regions are viewed, respectively, as the regions of acceptance, rejection, and noncommitment in a ternary classification. The positive and negative regions can be used to induce rules of acceptance and rejection; whenever it is impossible to make an acceptance or a rejection decision, the boundary region can be used to induce rules of noncommitment or deferment [65]. The three-way decision is a common problem solving strategy and consistent with the idea of human decision, and it has been successfully applied to many fields, both in theories and methodologies [18,64,65], such as the extended models and their corresponding approaches on DTRS [1,10–12,14,17,21,22,25,40,41,51,61,70], attribute reduction in DTRS [13,15,30,31,58,68], applications on DTRS [2,3,6,7,19,20,24,26,28,29,33,42,46,50,52,66,67,69]. Although DTRS has achieved lots of achievements in many domains, it has met some big challenges yet. One of them is how to calculate both the conditional probability and the thresholds in DTRS.

For calculating the conditional probability, Yao and Zhou [61] proposed a naive Bayesian DTRS model. The conditional probability is estimated based on the Bayes’ theorem and the naive probabilistic independence assumption. In order to compute the thresholds $\alpha$ and $\beta$ in DTRS, Herbert and Yao [10–12] introduced a game-theoretic approach to DTRS for learning optimal parameter values. Measures of classification ability are interpreted as players in a game, each with a goal of optimizing its value by increasing or decreasing the size of the classification regions, and the optimal parameter values are generated by the Nash equilibrium in game theory. Li and Zhou [14] argued that the thresholds of probabilistic inclusion are calculated based on the minimization of risk cost under the optimistic decision, the pessimistic decision, or the equable decision. Deng and Yao [4] utilized an information-theoretic interpretation of thresholds in PRS. In this paper, we provide an alternative method for computing the conditional probability required by DTRS, which leads to an effective way to estimate the required conditional probability. We also generalize DTRS to the multiple category classification. The results enable us to apply DTRS in solving real-world classification problem.

In this paper, we propose an integrated classification approach by using of logistic regression and DTRS. On the one hand, DTRS considers the costs and the deferment decision in a real problem. In the view of semantics, DTRS is utilized to systematically calculate the corresponding thresholds by considering cost or loss. DTRS can reasonably explain the thresholds of logistic regression and supply the deferment strategy for the classification. On the other hand, as we stated, the three-way decision in DTRS depends on a pair of thresholds and conditional probability [19]. Observed by the continuous and discrete data may coexist in information systems in real applications, logistic regression provides a way to compute the conditional probability in this situation, and solves the mentioned challenge of DTRS.

The remainder of this paper is organized as follows: Section 2 provides the basic concepts of logistic regression and DTRS. In Section 3, we combine the logistic regression and DTRS, and propose two novel integrated classification models to solve the binary misclassification problem and multiple classification problem, respectively. Then, the case studies of corporate failure prediction and high school program choices’ prediction are given to illustrate our approaches in Section 4. Section 5 concludes the paper and outlines the future work.

### 2. Preliminaries

Basic concepts, notations and results of the logistic regression analysis and DTRS are briefly reviewed in this section [5,9,16,18,20,27,34–36,39,38,44,45,48,49,60,71,72].

#### 2.1. Logistic regression analysis

This subsection introduces the logistic regression analysis to calculate the conditional probability for objects in an information table. As we known, logistic regression can be bi- or multinomial. Binary or binomial logistic regression refers to the instance in which the observed outcome can have only two possible types (e.g., “dead” vs. “alive”, “success” vs. “failure”, or “yes” vs. “no”). Multivariate or multinomial logistic regression refers to cases where the outcome can have three or more possible types (e.g., “better” vs. “no change” vs. “worse”) [49]. As a common logistic regression analysis approach, binary logistic regression and multinomial logistic regression methods are utilized to deal with the classification problem because they can directly compute the probability of occurrence of an event.
2.11. Binary logistic regression

Binary logistic regression is a form of regression which is used when the dependent is a dichotomy and the independents are of any type. 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 \(Pr((d = 0)|x) = 1 - Pr((d = 1)|x)\) denotes non-occurrence. The logistic function transformation of \(Pr((d = 1)|x)\) is known as the logit transformation as follows.

\[
\theta(Pr((d = 1)|x)) = \logit(Pr((d = 1)|x)) = \ln \left( \frac{Pr((d = 1)|x)}{1 - Pr((d = 1)|x)} \right) \tag{1}
\]

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

\[
\theta(Pr((d = 1)|x)) = \ln \left( \frac{Pr((d = 1)|x)}{1 - Pr((d = 1)|x)} \right) = \eta_0 + \eta_1 a_1 + \eta_2 a_2 + \cdots + \eta_k a_k \tag{2}
\]

where \(\eta_0\) denotes the intercept, and \(\eta_1, \eta_2, \ldots, \eta_k\) denote the regression coefficients of \(a_1, a_2, \ldots, a_k\), respectively. With the above analysis, the probability of occurrence of event \(Pr((d = 1)|x)\) can be expressed as follows:

\[
Pr((d = 1)|x) = \frac{e^{\theta(Pr((d = 1)|x))}}{1 + e^{\theta(Pr((d = 1)|x))}} = \frac{e^{\eta_0 + \eta_1 a_1 + \eta_2 a_2 + \cdots + \eta_k a_k}}{1 + e^{\eta_0 + \eta_1 a_1 + \eta_2 a_2 + \cdots + \eta_k a_k}} \tag{3}
\]

The discriminant rules are directly generated as follows:

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

2.12. Multinomial logistic regression

Unlike a binary logistic regression, multinomial logistic regression is used when the dependent variable in question is nominal and consists of more than two categories. Support \(x\) is an event and the dependent variable has \(m\) categories \((m > 2)\). The logistic function transformation of \(Pr((d = j)|x)\) \((1 \leq j \leq m - 1)\) is expressed as follows:

\[
\theta(Pr((d = j)|x)) = \ln \left( \frac{Pr((d = j)|x)}{Pr((d = m)|x)} \right) = \eta_0^j + \eta_1^j a_1 + \eta_2^j a_2 + \cdots + \eta_k^j a_k, \tag{4}
\]

\[
\theta(Pr((d = 2)|x)) = \ln \left( \frac{Pr((d = 2)|x)}{Pr((d = m)|x)} \right) = \eta_0^2 + \eta_1^2 a_1 + \eta_2^2 a_2 + \cdots + \eta_k^2 a_k, \tag{5}
\]

\[
\vdots
\]

\[
\theta(Pr((d = m - 1)|x)) = \ln \left( \frac{Pr((d = m - 1)|x)}{Pr((d = m)|x)} \right) = \eta_0^{m-1} + \eta_1^{m-1} a_1 + \eta_2^{m-1} a_2 + \cdots + \eta_k^{m-1} a_k. \tag{6}
\]

where the category \(m\) is chosen as the reference category, \(\eta_0^j\) denotes the intercept, and \(\eta_1^j, \eta_2^j, \ldots, \eta_k^j\) denote the regression coefficients of \(a_1, a_2, \ldots, a_k\) in the \(j\)th category, respectively \((1 \leq j \leq m - 1)\). With the above analysis, the probability of occurrence of event \(Pr((d = j)|x)\) can be expressed as follows:

\[
Pr((d = j)|x) = \frac{e^{\theta(Pr((d = j)|x))}}{1 + \sum_{j=1}^{m-1} e^{\theta(Pr((d = j)|x))}} = \frac{e^{\eta_0^j + \eta_1^j a_1 + \eta_2^j a_2 + \cdots + \eta_k^j a_k}}{1 + \sum_{j=1}^{m-1} e^{\eta_0^j + \eta_1^j a_1 + \eta_2^j a_2 + \cdots + \eta_k^j a_k}}. \tag{7}
\]

For the probability of occurrence of event \(Pr((d = m)|x)\) can be expressed as follows:

\[
Pr((d = m)|x) = \frac{1}{1 + \sum_{j=1}^{m-1} e^{\theta(Pr((d = j)|x))}} = \frac{1}{1 + \sum_{j=1}^{m-1} e^{\eta_0^j + \eta_1^j a_1 + \eta_2^j a_2 + \cdots + \eta_k^j a_k}}. \tag{8}
\]
The discriminant rules are directly generated as follows:

$$\arg \max_{i=1}^n \Pr((d = t) | x)$$

Thus, we discriminate the maximum probability of occurrence of event as its category. Obviously, this discriminant criteria is both suiting for binary scenario and multinomial scenario.

2.2. Decision-theoretic rough sets (DTRS)

In this subsection, we briefly introduce original DTRS model, especially in the aspect of the thresholds of the three-way decision. According to the Bayesian decision procedure, the original DTRS model is composed of two states and three actions [55–57, 59–61, 63]. 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 $A = \{a_P, a_B, a_N\}$, where $a_P$, $a_B$, and $a_N$ represent the three actions in classifying an object $x$, namely, deciding $x \in$ POS($C$), deciding $x$ should be further investigated $x \in$ BND($C$), and deciding $x \in$ NEG($C$), respectively.

The loss function regarding the risk or cost of actions in different states is given by the $3 \times 2$ matrix:

$$\begin{array}{ccc} 
\text{C (P)} & \text{\neg C (N)} \\
\lambda_{PP} & \lambda_{PN} \\
\lambda_{BP} & \lambda_{BN} \\
\lambda_{NP} & \lambda_{NN} 
\end{array}$$

In the matrix, $\lambda_{PP}$, $\lambda_{BP}$ and $\lambda_{NP}$ denote the losses incurred for taking actions of $a_P$, $a_B$ and $a_N$, respectively, when an object belongs to $C$. Similarly, $\lambda_{PN}$, $\lambda_{BN}$ and $\lambda_{NN}$ 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:

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

(10)

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 $x \in$ POS($C$);

(B) If $R(a_B|\{x\}) \leq R(a_P|\{x\})$ and $R(a_B|\{x\}) \leq R(a_N|\{x\})$, decide $x \in$ BND($C$);

(N) If $R(a_N|\{x\}) \leq R(a_P|\{x\})$ and $R(a_N|\{x\}) \leq R(a_B|\{x\})$, decide $x \in$ 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 $x \in$ POS($C$);

(B) If $\Pr(C|\{x\}) \leq \alpha$ and $\Pr(C|\{x\}) \geq \beta$, decide $x \in$ BND($C$);

(N) If $\Pr(C|\{x\}) \leq \beta$ and $\Pr(C|\{x\}) \leq \gamma$, decide $x \in$ NEG($C$).

The thresholds values $\alpha$, $\beta$, $\gamma$ are given by:

$$\alpha = \frac{\lambda_{PP} - \lambda_{BN}}{\lambda_{PN} - \lambda_{BN} + (\lambda_{BP} - \lambda_{PP})},$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{\lambda_{PN} - \lambda_{BN} + (\lambda_{NP} - \lambda_{BP})},$$

$$\gamma = \frac{\lambda_{PN} - \lambda_{NN} - \lambda_{NP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{NN} + (\lambda_{NP} - \lambda_{PP})}.\quad (11)$$

In addition, as a well-defined boundary region, the conditions of rule (B) suggest that $\alpha > \beta$, that is,
3.1. An integrated classification approaches with binary logistic regression analysis and DTRS

two-category and multi-category, respectively. try to combine the two methodologies and propose two novel logistic regression analysis approaches with DTRS under the able interpretation for the thresholds in logistic regression. DTRS and logistic regression is complementary. Therefore, we logistic regression can compute the conditional probability in an information table and DTRS model can provide a reason-
regression is the former generates a three-way decision and the latter generates a two-way decision. On the other hand, the methodologies, we point out their differences in two aspects. On the one hand, the difference between DTRS and logistic

3. An integrated approach to two-category and multi-category classifications

As stated in Section 2, DTRS mainly focuses on the classification problem with two category and multi-category. Meanwhile, the logistic regression can also solve the two types of classification problems. 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 generates a three-way decision and the latter generates a two-way decision. On the other hand, the logistic regression can compute the conditional probability in an information table and DTRS model can provide a reasonable interpretation for the thresholds in logistic regression. DTRS and logistic regression is complementary. Therefore, we try to combine the two methodologies and propose two novel logistic regression analysis approaches with DTRS under the two-category and multi-category, respectively.

3.1. An integrated classification approaches with binary logistic regression analysis and DTRS

The binary classification corresponds to two states of C and ¬C in DTRS. Combined with logistic regression, C is replaced by d = 1 and ¬C is replaced by d = 0 or ¬(d = 1) simultaneously. Suppose an information table is composed of n objects O = {o1, o2, . . . , on} and k attributes A = {a1, a2, . . . , ak}. In logistic regression, an object oi (i = 1, 2, . . . , n) is described by it’s corresponding variable’s values a1i, a2i, . . . , aki. For simplicity, the conditional probability of the object oi belonging to d = 1, is straightforwardly expressed by Pr((d = 1)|oi). 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 oi, 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.

\[
\begin{array}{ccc}
   & d = 1 & \neg (d = 1) \\
   a_p & \lambda_{ip} & \lambda_{ip} \\
   a_B & \lambda_{iB} & \lambda_{iB} \\
   a_N & \lambda_{iN} & \lambda_{iN} \\
\end{array}
\]

where ap, aB, and aN represent the three actions, namely, accepting the result of oi belonging to d = 1, the result of oi belonging to d = 1 should be further investigated, and rejecting the result of oi belonging to d = 1, respectively. λip, λiB and λiN denote the losses incurred for taking actions of “accepted”, “further investigated” and “rejected”, respectively, when oj belongs to d = 1. Similarly, λip, λiB and λiN denote the losses incurred for taking the same actions when oj belongs to
Fig. 1. A new integrated classification approach with binary logistic regression analysis and DTRS.

\( \neg(d = 1) \). 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 \(o_i \in \text{POS}(d = 1)\);

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

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

The thresholds values \(\alpha_i, \beta_i\) are given by:

\[
\alpha_i = \frac{\lambda_{PN}^i - \lambda_{BN}^i}{(\lambda_{PN}^i - \lambda_{BN}^i) + (\lambda_{BP}^i - \lambda_{PP}^i)},
\]

\[
\beta_i = \frac{\lambda_{BN}^i - \lambda_{NN}^i}{(\lambda_{BN}^i - \lambda_{NN}^i) + (\lambda_{NP}^i - \lambda_{BP}^i)}.
\]

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 binary logistic regression equation as follows.

\[
Pr((d = 1)|o_i) = \frac{e^{\eta_0 + \eta_1 a_1 + \eta_2 a_2 + \cdots + \eta_k a_k}}{1 + e^{\eta_0 + \eta_1 a_1 + \eta_2 a_2 + \cdots + \eta_k a_k}}. \tag{12}
\]

In the following, a new integrated classification approach with binary logistic regression analysis and DTRS consists of five steps, which is shown in Fig. 1.

Step 1: Selection of the independent variables and dependent variables and then construction of 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., \( Pr((d = 1)|a_i) \) (i = 1, 2, . . . , n).

Step 4: Generation of decision rules for each object according to DTRS model. For object \( a_i \) (i = 1, 2, . . . , 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 \( \alpha_i \) and \( \beta_i \) according to (P11)-(N11).

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

3.2. An integrated classification approach with multinomial logistic regression analysis and DTRS

For clarity and simplicity, we directly adopt the new DTRS model with multi-category according to [70]. Note that decision rules generated from different classes exist and associated rule conflict problem. Following [70], we also do not discuss this issue. Suppose the multinomial logistic regression corresponds to \( m \) categories. An information table is composed of \( n \) objects \( O = \{o_1, o_2, \ldots, o_n\} \). In multinomial logistic regression, an object \( o_i \) (i = 1, 2, . . . , n) is described by \( k \) attributes \( A = \{a_1, a_2, \ldots, a_k\} \) and its corresponding variable’s values \( a_{i1}, a_{i2}, \ldots, a_{ik} \). The conditional probability of the object \( o_i \) belonging to the \( k \)th category, is straightforward expressed by \( Pr((d = t)|o_i) \) (1 ≤ t ≤ m). Considering the difference of every object, 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:

\[
\begin{array}{cccccc}
| \alpha_{i1} & \lambda_{i1}^d(a_{i1}(d = 1)) & \lambda_{i1}^2(a_{i1}(d = 2)) & \cdots & \lambda_{i1}^t(a_{i1}(d = t)) & \cdots \lambda_{i1}^m(a_{i1}(d = m)) \\
| \alpha_{i2} & \lambda_{i2}^d(a_{i2}(d = 1)) & \lambda_{i2}^2(a_{i2}(d = 2)) & \cdots & \lambda_{i2}^t(a_{i2}(d = t)) & \cdots \lambda_{i2}^m(a_{i2}(d = m)) \\
| \alpha_{i3} & \lambda_{i3}^d(a_{i3}(d = 1)) & \lambda_{i3}^2(a_{i3}(d = 2)) & \cdots & \lambda_{i3}^t(a_{i3}(d = t)) & \cdots \lambda_{i3}^m(a_{i3}(d = m)) \\
\end{array}
\]

where \( \alpha_{is} \), \( \lambda_{is}^d \), and \( \lambda_{is}^t \) represent the three actions, namely, accepting the result of \( o_i \) belonging to \( d = t \), the result of \( o_i \) belonging to \( d = t \) should be further investigated, and rejecting the result of \( o_i \) belonging to \( d = t \) respectively. \( \lambda_{ip}^t, \lambda_{ip}^t \) and \( \lambda_{iu}^t \) denote the losses incurred for taking actions of “accepted”, “further investigated” and “rejected”, respectively, when \( o_i \) belongs to the category \( d = t \). By considering a reasonable kind of loss functions with

\[
\lambda_{ip}^t \leq \lambda_{it}^t < \lambda_{iu}^t \quad \lambda_{ip}^t \leq \lambda_{ip}^t < \lambda_{ip}^t \quad \text{for all } u, u \neq t.
\]

Furthermore, according to the condition of (P1)-(N1) in Section 2, we reconsider it as follows:

\[
\frac{(\lambda_{ip}^t - \lambda_{ip}^t)}{(\lambda_{ip}^t - \lambda_{ip}^t)} < \frac{(\lambda_{iu}^t - \lambda_{iu}^t)}{(\lambda_{iu}^t - \lambda_{iu}^t)}.
\]

With above conditions, the decision rules (P’ti) — (N’ti) of the object \( o_i \) can be expressed as:

(P’ti) If \( Pr((d = t)|o_i) \geq \alpha_{ti} \), decide \( o_i \in \text{POS}(d = t) \);

(B’ti) If \( \beta_{ti} < Pr((d = t)|o_i) < \alpha_{ti} \), decide \( o_i \in \text{BND}(d = t) \);

(N’ti) If \( Pr((d = t)|o_i) \leq \beta_{ti} \), decide \( o_i \in \text{NEG}(d = t) \).

The thresholds values \( \alpha_{ti}, \beta_{ti} \) are given by:

\[
\alpha_{ti} = \frac{\sum_{u=1,u \neq t}^{m} Pr((d = u)|o_i)(\lambda_{ip}^t - \lambda_{ip}^t)}{\lambda_{ip}^t - \lambda_{ip}^t},
\]

\[
\beta_{ti} = \frac{\sum_{u=1,u \neq t}^{m} Pr((d = u)|o_i)(\lambda_{ip}^t - \lambda_{ip}^t)}{\lambda_{ip}^t - \lambda_{ip}^t}.
\]
The conditional probability \( Pr((d = t)|o_i) \) can be calculated according to the multinomial logistic regression as follows:

\[
Pr((d = t)|o_i) = \begin{cases} \frac{e^{\beta_0 + \beta_1a_1 + \beta_2a_2 + \ldots + \beta_ka_k}}{1 + \sum_{t=1}^{m-1} e^{\beta_0 + \beta_1a_1 + \beta_2a_2 + \ldots + \beta_ka_k}} & 1 \leq t \leq m - 1; \\ \frac{1}{1 + \sum_{t=1}^{m} e^{\beta_0 + \beta_1a_1 + \beta_2a_2 + \ldots + \beta_ka_k}} & t = m. \end{cases}
\]

Similarly, a new integrated classification approach with multinomial logistic regression analysis and DTRS also consists of five steps as follows:

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

Step 2: Application of multinomial 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 every category according to logistic regression equation, i.e., \( Pr((d = t)|o_i) \) \((i = 1, 2, \ldots, n; \ t = 1, 2, \ldots, m)\).

Step 4: Generation of decision rules for each object in the every category. For object \( o_i \) \((i = 1, 2, \ldots, n)\), we set the loss functions of different actions according to expert experience and prior information, and calculate the thresholds \( a_{ti} \) and \( b_{ti} \) according to \( (P' ti) - (N' ti) \).

Step 5: Determination of concrete action for each object. For object \( o_i \), we compare the conditional probability \( Pr((d = t)|o_i) \) with the thresholds values \( a_{ti} \) and \( b_{ti} \), where \( t \) changes from 1 to \( m \). Then, the concrete action of \( o_i \) is determined.

Summarily, as our discussions in Sections 3.1 and 3.2, logistic regression analysis approach brings us a new way on estimating the conditional probability of the three-way decision. DTRS is utilized to calculate the thresholds for the logistic regression analysis approach and forms the three-way decision.

4. Illustration examples

Logistic regression can discriminate the category of every object according to the logistic regression equation. Different discriminative rules of classic logistic regression, we use the DTRS to explain the threshold of logistic regression. DTRS not only gives a semantic explanation corresponding to some practical problems, e.g., risk or cost, but also adds a new deferred decision rule. In fact, the DTRS can be used to revise the outcome of logistic regression. Besides, logistic regression provides an approach to compute the conditional probability of DTRS. In this section, we illustrate the new logistic discriminate approach by two examples. One is used to illustrate the new binary logistic regression with DTRS, the other is to illustrate the new multinomial logistic regression with DTRS.

**Example 1.** A case study of corporate failure prediction [3] is given to explain our approach proposed in Section 3.1. The data of this case includes 30 failed and 30 non-failed UK companies. Beynon and Peel [3] 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 [3]. 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. The coefficients in logistic regression equation is shown in Table 2.

From Table 2, we find the coefficient of independent variables and constant are significant. The binary logistic equation is shown in (13).

\[
Pr((d = 1)|o_i) = \frac{e^{-3.861 + 3.566a_1 + 0.010a_2 - 1.370a_3}}{1 + e^{-3.861 + 3.566a_1 + 0.010a_2 - 1.370a_3}}, \tag{13}
\]

where \( Pr((d = 1)|o_i) \) denotes the conditional probability of firm \( o_i \) belonging to a failed company. In Eq. (13), 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 \( a_2, a_1 \) and \( a_3 \) according to the Wald values of the three variables. The predicted conditional probabilities and predictions for 60 UK companies based on Eq. (13) and the discriminant rules in Section 2.1.1 are shown in Table 3, and \( d \) denotes the true state for these companies.
The underlined values in Table 3 denote the inconsistent issue between predicted classification and true state for a firm. A natural idea from these inconsistencies illustrates that the predictions with binary logistic regression need to reconsider.
For the firm o

The dependent variable is program choice denoted as 3.2. Entering a high school, each student makes program choices among general program, vocational program and academic program. Example 2.

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

Calculating results for the four firms are listed in Table 4.

For the firm o

We utilize SPSS 16.0 to run multinomial logistic regression for the original information table in [48]. The overall effect of multinomial logistic regression is significant. The coefficients in multinomial logistic regression equations are shown in Table 6.
From Table 6, the category \( d = 3 \) is chosen as the reference category. The results have two parts, labeled with the categories of dependence variable. \( a_4 \) is also as the reference category of \( a_2 \) and \( a_3 \). It is redundant and its coefficient is zero. For the students who select the general program \((d = 1)\), the difference between \( a_2 \) and \( a_4 \) is significant and the difference between \( a_2 \) and \( a_4 \) is not significant. For the students who select the vocational program \((d = 2)\), the difference between \( a_2 \) and \( a_4 \) is not significant and the difference between \( a_3 \) and \( a_4 \) is significant. According to (4)–(6), the multinomial logistic equations are shown in (14) and (15).

\[
\theta(Pr((d = 1) | o_i)) = \ln \left( \frac{Pr((d = 1) | o_i)}{Pr((d = 3) | o_i)} \right) = 1.689 - 0.058a_1 + 1.163a_2 + 0.630a_3; \quad (14)
\]

\[
\theta(Pr((d = 2) | o_i)) = \ln \left( \frac{Pr((d = 2) | o_i)}{Pr((d = 3) | o_i)} \right) = 4.236 - 0.114a_1 + 0.983a_2 + 1.274a_3. \quad (15)
\]

The conditional probabilities of student \( o_i \) belonging to every category are further expressed based on (7) and (8):

\[
Pr((d = 1) | o_i) = \frac{e^{1.689 - 0.058a_1 + 1.163a_2 + 0.630a_3}}{1 + e^{1.689 - 0.058a_1 + 1.163a_2 + 0.630a_3} + e^{4.236 - 0.114a_1 + 0.983a_2 + 1.274a_3}}, \quad (16)
\]

\[
Pr((d = 2) | o_i) = \frac{e^{4.236 - 0.114a_1 + 0.983a_2 + 1.274a_3}}{1 + e^{1.689 - 0.058a_1 + 1.163a_2 + 0.630a_3} + e^{4.236 - 0.114a_1 + 0.983a_2 + 1.274a_3}}, \quad (17)
\]

\[
Pr((d = 3) | o_i) = \frac{1}{1 + e^{1.689 - 0.058a_1 + 1.163a_2 + 0.630a_3} + e^{4.236 - 0.114a_1 + 0.983a_2 + 1.274a_3}}. \quad (18)
\]

The predicted conditional probabilities and predictions for 200 students based on Eqs. (16)–(18) and the discriminant rules in Section 2.1.2 are shown in Table 7. \( d \) denotes the true category for these students.

The underlined values in Table 7 denote the inconsistent issue between predicted classification and true category for students. According to the statistic results of SPSS, the prediction correct rate of multinomial logistic regression is 61%. Our new multinomial logistic regression approach is similar to Example 1 and also provides an approach to revise the predictions of original multinomial logistic regression with DTRS. For a student \( o_i \), we discriminate his or her program choice by comparing the conditional probability \( Pr((d = t) | o_i) \) generated by multinomial logistic regression and the thresholds \((\alpha_t, \beta_t)\) \((i = 1, 2, \ldots, n; \ t = 1, 2, \ldots, m)\). According to the decision rules \((P^t(o_i)) - (N^t(o_i))\), DTRS is utilized to construct the corresponding thresholds.

For simplicity, we select four firms \( o_{92}, o_{98}, o_{109} \) and \( o_{197} \) to illustrate our new multinomial logistic regression approach. The loss functions are carefully estimated by corresponding experts according to their experience and prior information. The loss functions and calculating results for the four students are listed in Table 8.

In Table 8, we set \( \lambda_{i,t} = 0 \) \((i = 1, 2, \ldots, m)\) by considering the fact that there is no cost when doing a right decision. Here, we compare the predicted probabilities in Table 7 with their corresponding thresholds in Table 8 among the four students. For the student \( o_2 \), \( Pr((d = 1) | o_2) = 0.1806, Pr((d = 2) | o_2) = 0.6992 \) and \( Pr((d = 3) | o_2) = 0.1202 \). The pair of thresholds for the three categories are: \((1) \alpha_1 = 0.6628, \beta_1 = 0.3037; \ (2) \alpha_2 = 0.4059, \beta_2 = 0.1444; \ (3) \alpha_3 = 0.8011, \beta_3 = 0.1486. Pr((d = 1) | o_2) = 0.1806 < \beta_1 = 0.3037, Pr((d = 2) | o_2) = 0.6992 > \alpha_2 = 0.4059 \) and \( Pr((d = 3) | o_2) = 0.1202 < \beta_3 = 0.1486 \). Hence, we take the action \( o_2 \) and accept the prediction. In this case, we agree the predicted result with multinomial logistic regression, namely, \( o_2 \) can choose the vocational program and we disagree the original category of \( o_2 \) in Table 5.

For the student \( o_{98} \), \( Pr((d = 1) | o_{98}) = 0.2302, Pr((d = 2) | o_{98}) = 0.3094 \) and \( Pr((d = 3) | o_{98}) = 0.4605 \). The pair of thresholds for the three categories are: \((1) \alpha_1 = 0.4911, \beta_1 = 0.3274; \ (2) \alpha_2 = 0.3011, \beta_2 = 0.1919; \ (3) \alpha_3 = 0.5689, \beta_3 = 0.0977\).
student at present. We need to collect more information to make a final decision.

The predicted conditional probabilities and predictions for the students.

| $O$ | $d$ | Predicted category | $Pr((d = 1)|o_1)$ | $Pr((d = 2)|o_1)$ | $Pr((d = 3)|o_1)$ | $O$ | $d$ | Predicted category | $Pr((d = 1)|o_2)$ | $Pr((d = 2)|o_2)$ | $Pr((d = 3)|o_2)$ |
|-----|-----|-------------------|-------------------|-------------------|-------------------|-----|-----|-------------------|-------------------|-------------------|-------------------|
| $o_1$ | 2 | 2 | 0.3382 | 0.5135 | 0.1483 | $o_{101}$ | 3 | 3 | 0.2302 | 0.3094 | 0.4605 |
| $o_2$ | 1 | 2 | 0.1806 | 0.6992 | 0.1202 | $o_{102}$ | 2 | 3 | 0.2302 | 0.3094 | 0.4605 |
| $o_3$ | 2 | 3 | 0.2368 | 0.3445 | 0.4187 | $o_{3}$ | 2 | 3 | 0.2214 | 0.2517 | 0.5269 |
| $o_4$ | 2 | 2 | 0.3508 | 0.4765 | 0.1727 | $o_{104}$ | 2 | 2 | 0.2204 | 0.5465 | 0.2331 |
| $o_5$ | 2 | 2 | 0.1689 | 0.7309 | 0.1001 | $o_{95}$ | 1 | 3 | 0.2342 | 0.3720 | 0.3938 |
| $o_6$ | 1 | 2 | 0.2378 | 0.4088 | 0.3534 | $o_{106}$ | 3 | 3 | 0.3165 | 0.1263 | 0.5572 |
| $o_7$ | 2 | 3 | 0.1974 | 0.6464 | 0.1563 | $o_{107}$ | 1 | 3 | 0.3515 | 0.1853 | 0.4632 |
| $o_8$ | 2 | 2 | 0.1689 | 0.7309 | 0.1001 | $o_{108}$ | 1 | 3 | 0.2248 | 0.2703 | 0.5049 |
| $o_9$ | 2 | 2 | 0.2204 | 0.5465 | 0.2331 | $o_{109}$ | 3 | 3 | 0.2036 | 0.1853 | 0.6110 |
| $o_{10}$ | 2 | 2 | 0.2026 | 0.6275 | 0.1699 | $o_{1010}$ | 3 | 3 | 0.2214 | 0.2517 | 0.5269 |

$\beta_3 = 0.4777$. $Pr((d = 1)|o_{98}) = 0.2302 < \beta_1 = 0.3274$, $Pr((d = 2)|o_{98}) = 0.6992 > \alpha_2 = 0.3011$ and $Pr((d = 3)|o_{98}) = 0.1202 < \beta_3 = 0.4777$. We take the action $a_{1}$ and disagree the prediction. In this case, we agree $o_{98}$ can choose the vocational program.

For the student $o_{109}$, $Pr((d = 1)|o_{109}) = 0.2036$, $Pr((d = 2)|o_{109}) = 0.1853$ and $Pr((d = 3)|o_{109}) = 0.6110$. The pair of thresholds for the three categories are: (1) $\alpha_1 = 0.4496$, $\beta_1 = 0.3272$; (2) $\alpha_2 = 0.7128$, $\beta_2 = 0.1527$; (3) $\alpha_3 = 0.8468$, $\beta_3 = 0.6112$. $Pr((d = 1)|o_{109}) = 0.2036 < \beta_1 = 0.3272$, $\beta_2 = 0.1527 < Pr((d = 2)|o_{109}) = 0.1853 < \alpha_2 = 0.7128$ and $Pr((d = 3)|o_{109}) = 0.6110$. We take the action $a_{1}$. In this case, we disagree the predicted result and original result, namely, $o_{109}$ may choose the vocational program. We cannot discriminate the program category of the student at present. We need to collect more information to make a final decision.

### Table 7
The predicted conditional probabilities and predictions for the students.

### Table 8
The loss functions and calculating results of the four students ($u$ is a unit cost).
For the student $\alpha_{197}$, $\Pr((d = 1)|\alpha_{197}) = 0.1814$, $\Pr((d = 2)|\alpha_{197}) = 0.1322$ and $\Pr((d = 3)|\alpha_{197}) = 0.6864$. The pair of thresholds for the three categories are: (1) $\alpha_1 = 0.4014$, $\beta_1 = 0.3169$; (2) $\alpha_2 = 0.2276$, $\beta_2 = 0.1119$; (3) $\alpha_3 = 0.8337$, $\beta_3 = 0.2218$. $\Pr((d = 1)|\alpha_{197}) = 0.1814 < \beta_1 = 0.3169$, $\beta_2 = 0.1119 < \Pr((d = 2)|\alpha_{197}) = 0.1322 < \alpha_2 = 0.2276$ and $\beta_3 = 0.2218 < \Pr((d = 3)|\alpha_{197}) = 0.6864 < \alpha_3 = 0.8337$. We take the action $\delta_{B_1}$ and $\delta_{B_3}$. In this case, we disagree the predicted result and original result, namely, $\alpha_{109}$ may choose the vocational program or the academic program. Similar with the student $\alpha_{197}$, we also cannot discriminate the program category of the student at present. We need to collect more information to make a final decision.

5. Conclusions

In logistic regression, the chosen category for an event is depended on the maximum probability of occurrence, but this criterion may lead to misclassifications when the maximum probability is not higher enough. In DTRS, it has met two big challenges yet, namely, how to determine the loss functions and how to calculate both the conditional probability and the thresholds. Specially, the latter one is a key point in applications of DTRS. In this paper, we combined logistic regression and DTRS together, and proposed two integrated classification approaches to solve the two classification problem (by using binary logistic regression) and multiple classification problem (by using multinomial logistic regression), respectively. Firstly, DTRS provides a semantic interpretation for the thresholds acquisition in logistic regression. Secondly, logistic regression provides a reasonable way to compute the conditional probability in DTRS. Furthermore, the conditional probability which is generated by logistic regression comes from logistic regression function, and it is machine-made; On the contrary, the thresholds which generate by DTRS is lying on loss functions, and it is man-made. Hence, our proposed approaches further provide a human–machine viewpoint in problem solving, and we believe our researches can make a solid contributions on the applications of DTRS. Our future research work will focus on the extension of the proposed method with behavior strategies or psychological methods. The conflict phenomena in multiple classification problem should be further investigated. In addition, other classification or discriminant methodologies, e.g., linear discriminant analysis, distance discriminant analysis, Fisher discriminant analysis and stepwise discriminant analysis, will be introduced to our future studies.

Acknowledgements

This paper is an extended version of the paper published in the proceedings of RSKT2011. This work is partially supported by the National Science Foundation of China (Nos. 71201133, 61175047, 71090402/G1), the Youth Social Science Foundation of the Chinese Education Commission (No. 11YJC630127), the Research Fund for the Doctoral Program of Higher Education of China (No. 20120184120028), the China Postdoctoral Science Foundation (No. 2012MS520310) and the Fundamental Research Funds for the Central Universities of China (No. SWJTU12CX117).

References
