Annals of Operations Research 135, 261–274, 2005

Discovering Credit Cardholders' Behavior by Multiple Criteria Linear Programming

GANG KOU
YI PENG
College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA

YONG SHI * yshi@unomaha.edu; yshi@gscas.ac.cn
College of Information Science and Technology, University of Nebraska at Omaha, Omaha, NE 68182, USA; Graduate School of Chinese Academy of Sciences, Beijing 100039, China

MORGAN WISE
First National Bank of Omaha, 1620 Dodge Street Stop 3103, Omaha, NE 68197, USA

WEIXUAN XU
Institute of Policy and Management, Chinese Academy of Sciences, Beijing, 100080, China

Abstract. In credit card portfolio management, predicting the cardholder's spending behavior is a key to reduce the risk of bankruptcy. Given a set of attributes for major aspects of credit cardholders and predefined classes for spending behaviors, this paper proposes a classification model by using multiple criteria linear programming to discover behavior patterns of credit cardholders. It shows a general classification model that can theoretically handle any class-size. Then, it focuses on a typical case where the cardholders' behaviors are predefined as four classes. A dataset from a major US bank is used to demonstrate the applicability of the proposed method.

Keywords: data mining, classification, multi-criteria linear programming, credit cardholders' behavior, SAS algorithms

1. Introduction

The history of credit card can be traced back to 1951 when the Diners' Club issued the first credit card in the US to 200 customers who could use it at 27 restaurants in New York (http://www.didyouknow.cd/creditcards.htm). At the end of fiscal 1999, there are 1.3 billion payment cards in circulation and Americans made $1.1 trillion credit purchases (http://www.nodebt.org/debt.htm). These statistics show that credit card business becomes a major power to stimulate the US economy growth in the last fifty years. However, the increasing credit card delinquencies and personal bankruptcy rates are causing plenty of headaches for banks and other creditors. The increase in personal bankruptcy

*Corresponding author.

rates was substantial. From 1980 to 2000, the number of individual bankruptcy filings in the US increased approximately 500% (Stavins, 2000). How to predict bankruptcy in advance and avoid huge charge-off losses becomes a critical issue of credit card issuers.

Since a credit card database can contain hundreds and thousands of credit transactions, it is impossible to discover or predict the cardholders' behaviors without using mathematical tools. In fact, the practitioners have tried a number of quantitative techniques to conduct the credit card portfolio management. Some examples of known approaches are (1) Behavior Score developed by Fair Isaac Corporation (FICO) (www.fairisaac.com); (2) Credit Bureau Score also developed by FICO (www.fairisaac.com); (3) First Data Resource (FDR)'s Proprietary Bankruptcy Score (http://www.firstdatacorp.com); (4) Multiple-criteria Score (Shi, 2001, 2002) and (5) Dual-model Score (Lin, 2002). A basic characteristic of these models is that they first consider the behaviors of the cardholders as two predefined classes: bankrupt accounts and non-bankrupt accounts according to their historical records. Then they use statistical methods, linear programming or neural networks to compute the Kolmogorov-Smirnov (KS) value that measures the largest separation of these two cumulative distributions of bankrupt accounts and non-bankrupt accounts in a training set (Conover, 1999). The resulting KS values from the learning process are applied to the real-life credit data warehouse to predict the percentage of bankrupt accounts in the future. Thus, these methods can be generally regarded as two-class models in credit card portfolio management.

In order to discover more knowledge for advanced credit card portfolio management, multi-class (class number is larger than two) data mining methods are needed. Comparing with two-class model, the multi-class model enlarges the difference between bankrupt accounts and non-bankrupt accounts. This enlargement increases not only the accuracy of separation, but also provides more useful information for credit card issuers or banks. From theoretical point of view, a general multi-class model is easy to construct. From practical point of view, the best control parameters (such as class boundaries) have to be identified through a learning process on a training data set. Therefore, finding the practical technology with certain type of multi-class model is not a trivial task. Peng (2002) and Shi (2002) have explored a three-class classification model. This model produces the prediction distribution for each behavior class so that credit card issuers can establish their credit limit policies for various cardholders.

The purpose of this paper is to build a four-class model by using charge-off status to separate credit cardholders' behavior. These four classes are predefined as: Bankrupt charge-off accounts, Non-bankrupt charge-off accounts, Delinquent accounts, and Current accounts. Bankrupt charge-off accounts are accounts that have been written off by credit card issuers because of cardholders' bankrupt claims. Non-bankrupt charge-off accounts are accounts that have been written off by credit card issuers due to reasons other than bankrupt claims. The charge-off policy may vary among authorized institutions. Delinquent accounts are accounts that haven't paid the minimum balances for more than 90 days. Current accounts are accounts that have paid the minimum balances or have no balances.

This paper proceeds as follows. Section 2 proposes a general multi-class model for credit card portfolio management. Section 3 reviews the previous results of two-class and three-class models. Section 4 develops a four-class model to capture charge-off behavior of cardholders and provides an algorithm to implement the model. Section 5 reports results of the empirical study for the four-class model on a real-life credit data warehouse from a major US bank. Section 6 concludes the paper with further research directions.

2. General model

Data mining for credit card portfolio management decisions is to classify the different cardholders' behavior in terms of their payment to the credit card companies, such as banks and mortgage loan firms. In realty, while all credit card companies share some common variables, some credit card companies have different variables to describe the cardholders' behavior. The examples of common categories of the variables are balance, purchase, payment and cash advance. Some credit card companies may consider residence state category and job security as special variables. In the case of FDR, there are 38 original variables (or attributes) from the common variables over the past seven months. Then, a set of 80 derived variables is internally generated from the 38 variables to perform a precise data mining task.

A general multi-class model by using multiple criteria linear programming can be proposed as:

Given a set of r variables about the credit cardholders' behavior $a = (a_1, \dots, a_r)$, let $A_i = (A_{i1}, \dots, A_{ir})$ be a development sample of credit data for the variables, where $i = 1, \dots, n$ and n is the sample size. We want to determine the coefficients of the variables, denoted by $\boldsymbol{X} = (x_1, \dots, x_r)$. If a given problem can be predefined as s different classes, then the boundary between the jth and $j + 1$th groups can be b_j, $j = 1, \dots, s-1$. The separation of these classes is:

$$\boldsymbol{A}_i\boldsymbol{X} \le b_1, \boldsymbol{A}_i \in G_1; \quad b_{k-1} \le \boldsymbol{A}_i\boldsymbol{X} \le b_k, \boldsymbol{A}_i \in G_k, k = 2, \dots, s-1;$$
$$\boldsymbol{A}_i\boldsymbol{X} \ge b_{s-1}, \boldsymbol{A}_i \in G_s.$$

Let α_i^j be the overlapping degree with respect of A_i within G_j and G_{j+1}, and β_i^j be the distance from A_i within G_j and G_{j+1} to its adjusted boundaries. The separation of these classes is modified as:

$$\boldsymbol{A}_i\boldsymbol{X} = b_1 - \alpha_i^1 + \beta_i^1, \boldsymbol{A}_i \in G_1; \quad b_{k-1} + \alpha_i^{k-1} - \beta_i^{k-1} = \boldsymbol{A}_i\boldsymbol{X} = b_k - \alpha_i^k$$
$$+ \beta_i^k, \boldsymbol{A}_i \in G_k, k = 2, \dots, s-1; \quad \boldsymbol{A}_i\boldsymbol{X} = b_{s-1} + \alpha_i^{s-1} - \beta_i^{s-1}, \boldsymbol{A}_i \in G_s;$$

The goal of this general problem is to reach the maximization of $\Sigma_j \beta_i^j$ and the minimization of $\Sigma_j \alpha_i^j$ simultaneously:

$$
\begin{aligned}
\text{(M1)}\quad & \text{Minimize}\quad \Sigma_i \Sigma_j \alpha_i^j \ \text{ and Maximize } \Sigma_i \Sigma_j \beta_i^j \\
& \text{Subject to:} \\
& \qquad A_i X = b_1 - \alpha_i^1 + \beta_i^1,\ A_i \in G_1; \\
& \qquad A_i X = b_{k-1} + \alpha_i^{k-1} - \beta_i^{k-1},\ A_i \in G_k,\ k = 2, \ldots, s-1; \\
& \qquad A_i X = b_k - \alpha_i^k + \beta_i^k,\ A_i \in G_k,\ k = 2, \ldots, s-1; \\
& \qquad A_i X = b_{s-1} + \alpha_i^{s-1} - \beta_i^{s-1},\ A_i \in G_s; \\
& \qquad b_{k-1} + \alpha_i^{k-1} \le b_k - \alpha_i^k,\ k = 2, \ldots, s-1;
\end{aligned}
$$

where $\boldsymbol{A}_i$ are given; $\boldsymbol{X}$ and b_j are unrestricted; and α_i^j, and $\beta_i^j \ge 0$, $j = 1, \ldots, s-1$.

The background of the above model is based on both linear discriminant analysis and multiple criteria linear programming. In linear discriminant analysis, the misclassification of data separation can be described by two opposite objectives in a linear system. The first one is to maximize the minimum distances (MMD) of observations from the critical value. The second objective separates the observations by minimizing the sum of the deviations (MSD) among the observations (Freed, 1981, 1986; Koehler, 1990). Comparing with the traditional mathematical tools in Classification, such as decision tree, statistics and neural networks, this approach is simple and direct, free of the statistical assumptions, and flexible by allowing decision makers to play an active part in the analysis (Joachimsthaler, 1988). However, because linear discriminant analysis uses only MMD, MSD, or a given combination of MMD and MSD to measure the misclassification, it could not find the best tradeoff of two measurements. This shortcoming has been coped with by the techniques of multiple criteria linear programming (MCLP) (Shi, 2002). Using MCLP, we can optimize MMD and MSD simultaneously to identify the best tradeoff of MMD and MSD. The resulting classification produces better data separation than linear discriminant analysis.

Theoretically speaking, the general model can classify data into any class-size of data ($s \ge 2$). But, the practical model for two, three, four or more classes of data has to be built individually since each model involves different control parameters. In addition, in the case of credit card portfolio management, classifying credit cardholders' behaviors into fewer than five categories is manageable and encouraged. This belief is not only based on business practice, but also comes from George Miller's classic article: "The Magical Number Seven, Plus or Minus Two" (Miller, 1956). Miller pointed out that human capacity for processing information is limited to the magic number seven, plus or minus two which fits the credit card portfolio situation correctly. In the following section, we review the preliminary research findings on two and three-class models.

3. Two class models

Two-class model based on multiple criteria linear programming is initiated as Shi (2001):

Given a set of r variables (attributes) $\boldsymbol{a} = (a_1, \ldots, a_r)$, let $\boldsymbol{A}_i = (A_{i1}, \ldots, A_{ir})$ be the development sample of data for the variables, where $i = 1, \ldots, n$ and n is the sample size. We want to determine the coefficients of the variables, denoted by $\boldsymbol{X} = (x_1, \ldots, x_r)$, and a boundary value b to separate two classes: G_1 (bankrupt accounts) and G_2 (non-bankrupt accounts); that is,

$$\boldsymbol{A}_i\boldsymbol{X} \le b, \boldsymbol{A}_i \in G_1 \text{ and } \boldsymbol{A}_i\boldsymbol{X} \ge b, A_i \in G_2.$$

To measure the separation of G_1 and G_2, we define:

α_i = the overlapping of two-group (classes) boundary for case $\boldsymbol{A}_i$ (external measurement);
β_i = the distance of case $\boldsymbol{A}_i$ from its adjusted boundary (internal measurement);

According to the general model (M1), we want to minimize the sum of α_i and maximize the sum of β_i simultaneously (M2) as:

(M2)
Minimize $\Sigma_i\alpha_i$ and Maximize $\Sigma_i\beta_i$
Subject to:

$$\boldsymbol{A}_i\boldsymbol{X} = b + \alpha_i - \beta_i, \quad \boldsymbol{A}_i \in G_1,$$
$$\boldsymbol{A}_i\boldsymbol{X} = b - \alpha_i + \beta_i, \quad \boldsymbol{A}_i \in G_2,$$

where $\boldsymbol{A}_i$ are given, $\boldsymbol{X}$ and b are unrestricted, and α_i and $\beta_i \ge 0$.

To facilitate the computation on the large-size of data warehouse, the compromise solution approach (Shi, 2001, 2001b) has been employed to reform the above model (M2) so that a commercial software package, such as SAS LP can be used to systematically identify the *best tradeoff* between $-\Sigma_i\alpha_i$ and $\Sigma_i\beta_i$ for an optimal solution. To explain this, we assume the "ideal value" of $-\Sigma_i\alpha_i$ be $\alpha^* > 0$ and the "ideal value" of $\Sigma_i\beta_i$

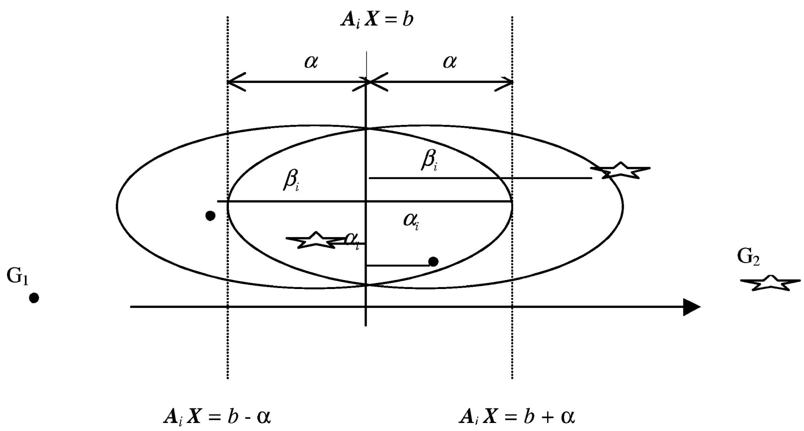

Figure 1. Two-class model.

be $\beta^* > 0$. Then, if $-\Sigma_i\alpha_i > \alpha^*$, we define the regret measure as $-d_\alpha^+ = \Sigma_i\alpha_i + \alpha^*$; otherwise, it is 0. If $-\Sigma_i\alpha_i < \alpha^*$, the regret measure is defined as $d_\alpha^- = \alpha^* + \Sigma_i\alpha_i$; otherwise, it is 0. Thus, we have (i) $\alpha^* + \Sigma_i\alpha_i = d_\alpha^- - d_\alpha^+$, (ii) $|\alpha^* + \Sigma_i\alpha_i| = d_\alpha^- + d_\alpha^+$, and (iii) $d_\alpha^-, d_\alpha^+ \geq 0$. Similarly, we derive $\beta^* - \Sigma_i\beta_i = d_\beta^- - d_\beta^+$, $|\beta^* - \Sigma_j\beta_i| = d_\beta^- + d_\beta^+$, and $d_\beta^-, d_\beta^+ \geq 0$. The two-class model has been gradually evolved as Shi (2001):

(M3)
Minimize $d_\alpha^- + d_\alpha^+ + d_\beta^- + d_\beta^+$
Subject to:

$$\begin{aligned}
\alpha^* + \Sigma_i\alpha_i &= d_\alpha^- - d_\alpha^+, \\
\beta^* - \Sigma_i\beta_i &= d_\beta^- - d_\beta^+, \\
\boldsymbol{A}_i\boldsymbol{X} &= b + \alpha_i - \beta_i, \quad \boldsymbol{A}_i \in G_1, \\
\boldsymbol{A}_i\boldsymbol{X} &= b - \alpha_i + \beta_i, \quad \boldsymbol{A}_i \in G_2,
\end{aligned}$$

where $\boldsymbol{A}_i$, α^*, and β^* are given, $\boldsymbol{X}$ and b are unrestricted, and $\alpha_i, \beta_i, d_\alpha^-, d_\alpha^+, d_\beta^-, d_\beta^+ \geq 0$.

Remark 1. We see that by using the compromise approach, the bi-criteria problem (M2) is transformed to a single criterion problem (M3). The optimal solution of (M3) is that of (M2), but not vice versa. Therefore, (M3) is a reduced problem of (M2). However, model (M3) allows us to utilize SAS LP for computing a large-scale of credit card databases. Although the pair of regret measures d_α^- and d_α^+ is stated in (M3), only one of two will have a non-zero value in reality. This is also true for d_β^- and d_β^+. The two-class model is known as "Good vs. Bad" in credit card portfolio management. The purpose of using this model is to produce a "black list" of Bads. The empirical study of this model based on real-life credit database from a major US Bank is reported by Shi (2001).

Example 1. As an illustration, a simple example is used to verify the feasibility of the model. The example, which was transformed from Freed (1981)'s first article of applying linear programming on discriminant problems, is about assigning credit applicants to different risk groups. An applicant is to be classified as a "poor", or "good" credit risk based on his/her responses to two questions appearing on a standard credit application. Table 1 shows previous experience with 9 customers.

According to the previously formulated model (Model 3), this classification problem can be recast as an MCLP problem:

Minimize $d_\alpha^- + d_\alpha^+ + d_\beta^- + d_\beta^+$
Subject to:

$$\begin{aligned}
\alpha^* + \Sigma_i\alpha_i &= d_\alpha^- - d_\alpha^+, \\
\beta^* - \Sigma_i\beta_i &= d_\beta^- - d_\beta^+, \\
x_1 + 3x_2 &= b + \alpha_1 - \beta_1 \\
2x_1 + 5x_2 &= b + \alpha_2 - \beta_2
\end{aligned}$$

Table 1
Small example of credit applicants risk classification.

	Credit customer	Responses		Training results
		Quest 1	Quest	
Group I (Poor risk)	A_1	1	3	Poor
	A_2	2	5	Poor
	A_3	3	4	Good
	A_4	4	6	Poor
Group II (Good risk)	A_5	5	7	Good
	A_6	6	9	Poor
	A_7	7	8	Good
	A_8	7	7	Good
	A_9	9	9	Good

$$3x_1 + 4x_2 = b + \alpha_3 - \beta_3$$
$$4x_1 + 6x_2 = b + \alpha_4 - \beta_4$$
$$5x_1 + 7x_2 = b - \alpha_5 + \beta_5$$
$$6x_1 + 9x_2 = b - \alpha_6 + \beta_6$$
$$7x_1 + 8x_2 = b - \alpha_7 + \beta_7$$
$$7x_1 + 7x_2 = b - \alpha_8 + \beta_8$$
$$9x_1 + 9x_2 = b - \alpha_9 + \beta_9$$

where $\alpha^* = 0.1, \beta^* = 30000$ and $b = 1$ are given, x_1 and x_2 are unrestricted, and $\alpha_i, \beta_i, d_\alpha^-, d_\alpha^+, d_\beta^-, d_\beta^+ \geq 0, i = 1, \ldots, 9$.

The results, coefficients values of

$$\boldsymbol{X} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3103.35632184 \\ -2068.79310345 \end{pmatrix},$$

are applied to 9 customers to produce optimal solution of linear programming for this classifier and are obtained as Column 4 of Table 1, where only cases $\boldsymbol{A}_3$ and $\boldsymbol{A}_6$ we misclassified.

4. Four-class model formulation

Although the results of two-class model can be used to handle two extreme conditions, it has nothing to do with the majority accounts that lie between bankrupt and non-bankrupt accounts. It cannot discover the process of any behavior change, for example, how a current account can become bankrupt. These changes are important for creditors to make a decision.

In the four-class separation, we use term: "charge-off" to predict the cardholders' behaviors. According to this idea, four classes are defined as: Bankrupt charge-off accounts, Non-bankrupt charge-off accounts, Delinquent accounts, and Current accounts. Bankrupt charge-off accounts are accounts that have been written off by credit card issuers because of cardholders' bankrupt claims. Non-bankrupt charge-off accounts are accounts that have been written off by credit card issuers due to reasons other than bankrupt claims. The charge-off policy may vary among authorized institutions. Normally, an account will be written off when the receivable has been overdue for more than 180 days or when the ultimate repayment of the receivable is unlikely (e.g., the cardholder cannot be located) (http://www.info.gov.hk/hkma/eng/press/2001/20010227e6.htm). Delinquent accounts are accounts that have not paid the minimum balances for more than 90 days. Current accounts are accounts that have paid the minimum balances or have no balances.

This separation is more precise than two-class and three-class models in credit card portfolio management. For instance, bankrupt charge-off and non-bankrupt charge-off accounts are probably both classified as "Bad" accounts in two or three-group separations. This model, however, will call for different handling against these accounts.

From the general model (M1), a four-class model has three boundaries, b_1, b_2, and b_3, to separate four classes (figure 2). Each class is represented by a symbol as follows:

✦ stands for G_1 (Bankrupt charge-off account),
⬮ stands for G_2 (Non-bankrupt charge-off account),
✲ stands for G_3 (Delinquent account), and
★ stands for G_4 (Current account).

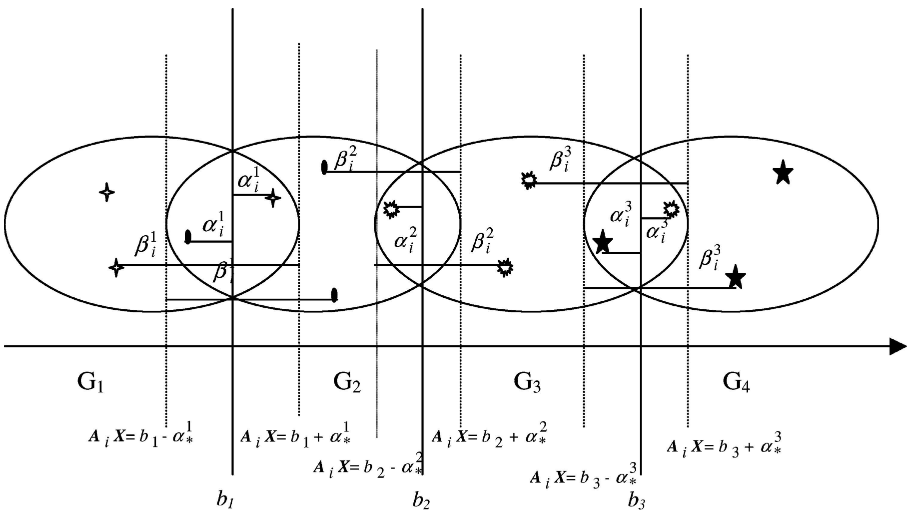

Figure 2. Four-class model.

Given a set of r variables about the cardholders $\boldsymbol{A} = (A_1, \ldots, A_i)$, let $A_t = (A_{i1}, \ldots, A_{ir})$ be the development sample of data for the variables, where $i = 1, \ldots, n$ and n is the sample size. We try to determine the coefficients of the variables, denoted by $\boldsymbol{X} = (x_1, \ldots, x_r)$, boundary b_1 to separate G_1 from G_2, G_3, and G_4, boundary b_3 to separate G_4 from G_1, G_2, and G_3, boundary b_2 to separate G_2 and G_3. This separation can be represented by:

$$\begin{aligned} &\boldsymbol{A}_i\boldsymbol{X} \leq b_1, \quad A_i \in G_1 \\ &\quad b_1 \leq \boldsymbol{A}_i\boldsymbol{X} \leq b_2, \quad \boldsymbol{A}_i \in G_2 \\ &\quad b_2 \leq \boldsymbol{A}_i\boldsymbol{X} \leq b_3, \quad \boldsymbol{A}_i \in G_3, \\ &\quad b_3 \leq \boldsymbol{A}_i\boldsymbol{X}, \quad \boldsymbol{A}_i \in G_4 \end{aligned}$$

Similar to two-class and three-class models, we apply two measurements for better separations. Let α_i^1 be the overlapping degree with respect of $\boldsymbol{A}_i$ within G_1 and G$_2$, α_i^2 be the overlapping degree with respect of $\boldsymbol{A}_i$ within G_2 and G_3, and α_i^3 be the overlapping degree with respect of $\boldsymbol{A}_i$, within G_3 and G_4. Let β_i^1 be the distance from $\boldsymbol{A}_i$ within G_1 and G_2 to its adjusted boundaries ($\boldsymbol{A}_i\boldsymbol{X} = b_1 - \alpha_*^1$, and $\boldsymbol{A}_i\boldsymbol{X} = b_1 + \alpha_*^1$), β_i^2 be the distance from $\boldsymbol{A}_i$ within G_2 and G_3 to its adjusted boundaries ($\boldsymbol{A}_i\boldsymbol{X} = b_2 - \alpha_*^2$, and $\boldsymbol{A}_i\boldsymbol{X} = b_2 + \alpha_*^2$), and β_i^3 be the distance from $\boldsymbol{A}_i$ within G_3 and G_4 to its adjusted boundaries ($\boldsymbol{A}_i\boldsymbol{X} = b_3 - \alpha_i^3$, and $\boldsymbol{A}_i X = b_3 + \alpha_i^3$). We want to reach the maximization of β_i^1, β_i^2, and β_i^3 and the minimization of $\alpha_i^1\alpha_i^2$ and α_i^3 simultaneously. After putting $\alpha_i^1, \alpha_i^2, \alpha_i^3, \beta_i^1, \beta_i^2$, and β_i^3 into the above four-class separation, we have:

(M4) Minimize $\Sigma_i\left(\alpha_i^1 + \alpha_i^2 + \alpha_i^3\right)$ and Maximize $\Sigma_i\left(\beta_i^1 + \beta_i^2 + \beta_i^3\right)$

Subject to:

G_1 : $\boldsymbol{A}_i\boldsymbol{X} = b_1 + \alpha_i^1 - \beta_i^1$, $\quad \boldsymbol{A} \in G_1$, (Bankrupt charge-off)

G_2 : $\boldsymbol{A}_i\boldsymbol{X} = b_1 + \alpha_i^1 - \beta_i^1$, $\quad \boldsymbol{A}_i X = b_2 + \alpha_i^2 - \beta_i^2$, $\boldsymbol{Ai} \in G2$, (Non-bankrupt charge-off)

G_3 : $\boldsymbol{A}_i\boldsymbol{X} = b_2 + \alpha_i^2 - \beta_i^2$, $\boldsymbol{A}_i X = b_3 - \alpha_i^3 + \beta_i^3$, $\quad \boldsymbol{Ai} \in G3$, (Delinquent)

G_4 : $\boldsymbol{A}_i\boldsymbol{X} = b_3 - \alpha_i^3 + \beta_i^3$, $\boldsymbol{A}_i \in G_4$ (Current)

$b_1 + \alpha_i^1 \leq b_2 - \alpha_i^2$,

$b_2 + \alpha_i^2 \leq b_3 - \alpha_i^3$

where $\boldsymbol{A}_i$, are given, X, b_1, b_2, b_3 are unrestricted, and $\alpha_i^1, \alpha_i^2, \alpha_i^3, \beta_i^1, \beta_i^2$, and $\beta_i^3 \geq 0$.

The constraints $b_1 + \alpha_i^1 \leq b_2 - \alpha_i^2$ and $b_2 + \alpha_i^2 \leq b_3 - \alpha_i^3$ guarantee the existence of four groups by enforcing b_1 lower than b_2, and b_2 lower than b_3. Then we apply the compromise solution approach (Shi, 1989, 2001) to reform the model. We assume the ideal value of $-\Sigma_i\alpha_i^1$ be $\alpha_*^1 > 0$, $-\Sigma_i\alpha_i^2$ be $\alpha_*^2 > 0$, $-\Sigma_i\alpha_i^3$ be $\alpha_*^3 > 0$, and the ideal value of $\Sigma_i\beta_i^1$ be $\beta_*^1 > 0$, $\Sigma_i\beta_i^2$ be $\beta_*^2 > 0$, $\Sigma_i\beta_i^3$ be $\beta_*^3 > 0$.

Then, the four-group model (M4) is transformed as:

(M5) Minimize $d_{\alpha 1}^{-} + d_{\alpha 1}^{+} + d_{\beta 1}^{-} + d_{\beta 1}^{+} + d_{\alpha 2}^{-} + d_{\alpha 2}^{+} + d_{\beta 2}^{-} + d_{\beta 2}^{+} + d_{\alpha 3}^{-} + d_{\alpha 3}^{+} + d_{\beta 3}^{-}$ $+d_{\beta 3}^{+}$ Subject to:

$$\alpha_*^1 + \Sigma_i \alpha_i^1 = d_{\alpha 1}^{-} - d_{\alpha 1}^{+}, \quad \beta_*^1 - \Sigma_i \beta_i^1 = d_{\beta 1}^{-} - d_{\beta 1}^{+},$$
$$\alpha_*^2 + \Sigma_i \alpha_i^2 = d_{\alpha 2}^{-} - d_{\alpha 2}^{+}, \quad \beta_*^2 - \Sigma_i \beta_i^2 = d_{\beta 2}^{-} - d_{\beta 2}^{+},$$
$$\alpha_*^3 + \Sigma_i \alpha_i^3 = d_{\alpha 1}^{-} - d_{\alpha 3}^{+}, \quad \beta_*^3 - \Sigma_i \beta_i^3 = d_{\beta 3}^{-} - d_{\beta 3}^{+},$$
$$b_1 + \alpha_i^1 \le b_2 - \alpha_i^2, \quad b_2 + \alpha_i^2 \le b_3 - \alpha_i^3,$$
$$G_1\colon \quad \boldsymbol{A}_i \boldsymbol{X} = b_1 + \alpha_i^1 - \beta_i^1, \boldsymbol{A}_i \ne G1 \quad \text{(Bankrupt charge-off)}$$
$$G_2\colon \quad \boldsymbol{A}_i \boldsymbol{X} = b_1 - \alpha_i^1 + \beta_i^1, \boldsymbol{A}_i \boldsymbol{X} = b_2 + \alpha_i^2 - \beta_i^2, A_i \in G2, \text{ (Non-bankrupt charge-off)}$$
$$G_3\colon \quad \boldsymbol{A}_i \boldsymbol{X} = b_2 + \alpha_i^2 - \beta_i^2, \boldsymbol{A}_i \boldsymbol{X} = b_3 - \alpha_i^3 + \beta_i^3, A_i \in G3, \text{ (Delinquent)}$$
$$G_4\colon \quad \boldsymbol{A}_i \boldsymbol{X} = b_3 - \alpha_i^3 + \beta_i^3, \boldsymbol{A}_i \in G_4, \quad \text{(Current)}$$

Where $\boldsymbol{A}_i$, are given, $b_1 \le b_2 \le b_3$, $\boldsymbol{X}$, b_1, b_2, b_3, are unrestricted, and α_i^1, α_i^2, α_i^3, β_i^1, β_i^2 and $\beta_i^3 \ge 0$.

A SAS-based algorithm of this four-class model is proposed as follows:

Algorithm 1.

Step 1. Use *ReadCHD.sas* to convert both the training and verifying data into SAS data sets.

Step 2. Use *GroupDef.sas* to divide the observations within the training data sets into four groups: G_1, G_2, G_3, and G_4.

Step 3. Use *4GModel.sas* to perform the separation task on the training data. Here, PROC LP in SAS is called to calculate the (M5) model for the best solution of the four-class separation given the values of control parameters (α_*^1, β_*^1, α_*^2, β_*^2, α_*^3, β_*^3, b_1, b_2, b_3).

Step 4. Use *Score.sas* to produce the graphical representations of the training results. Step 3–4 will not terminate until the best training result is found.

Step 5. Use *Predict, sas* to mine the four classes from the verifying data set.

5. Empirical study and managerial significance of four-class models

The credit data from a well-known major US bank is used to perform the Algorithm 1. A training set of 160 card account samples from 25,000 credit card records is used to test the control parameters of the model for the best class separation. A verifying data

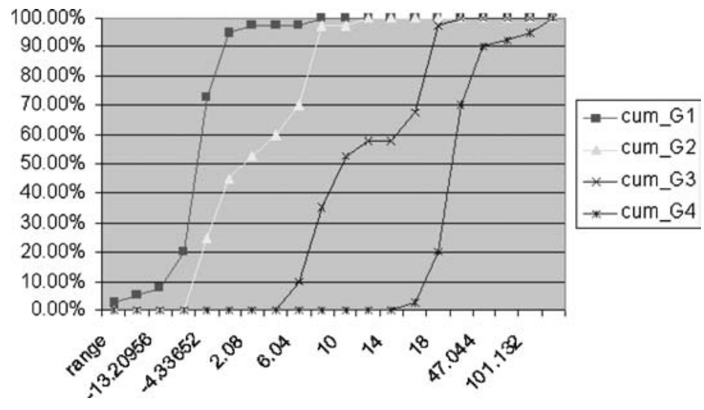

Figure 3. Four-class training data set (160) (X-axis is the score of each data and Y-axis is the percents that group is successfully classified).

set with 5,000 accounts is then applied. Four groups are defined as: Bankrupt charge-off accounts (The number of over-limits $\geq$ 13), Non-bankrupt charge-off accounts (7 $\leq$ The number of over-limits $\leq$12), Delinquent accounts (2 $\leq$ The number of over-limits $\leq$ 6), and Current accounts(0 $\leq$ The number of over-limits $\leq$ 2).

After several learning trials for different sets of boundary values, we found the values: $b_1 = 0.05, b_2 = 0.8, b_3 = 1.95$ (without changing $\alpha_*^1, \alpha_*^2, \alpha_*^3, \beta_*^1, \beta_*^2$, and β_*^3) brought the best separation as shown in figure 3, in which "cum_Gl" refers to cumulative percentage of G_1 (Bankrupt charge-off accounts); "cum_G2" refers to cumulative percentage of G_2 (Non-bankrupt charge-off accounts); "cum_G3" refers to cumulative percentage of G_3 (Delinquent accounts); and "cum_G4" refers to cumulative percentage of G_4 (Current accounts).

This Training set has total 160 samples. Gl has been correctly identified for 90% (36/40), G2 90% (36/40), G3 85% (34/40) and G4 97.5% (39/40). In addition to these absolute classifications criteria, the KS Score is calculated by KS value = max|Cum. distribution of Good - Cum. distribution of Badl.

The KS values are 50 for Gl vs. G2, 62.5 for G2 vs. G3 and 77.5 for G3 vs. G4.

We observe some relationships between boundary values and separation: (1) Optimal solution and better separation can be achieved by changing the value of b_j, α_*^j and β_*^j. (2) Once a feasible area is found, the classification result will be similar for the solution in that area. (3) The definition of group and data attributes will influence the classification result. When applying the resulting classifier to predict the verifying data set, we can predict the verifying set by the classifier as Gl for 43.4% (23/53), G2 for 51% (84/165), G3 for 28% (156/557) and G4 for 68% (2872/4225). The predicted KS values are 36.3 for Gl vs. G2, 21.6 for G2 vs. G3 and 50.7 for G3 vs. G4. These results indicate that the predicted separation between G3 and G4 is better than others.

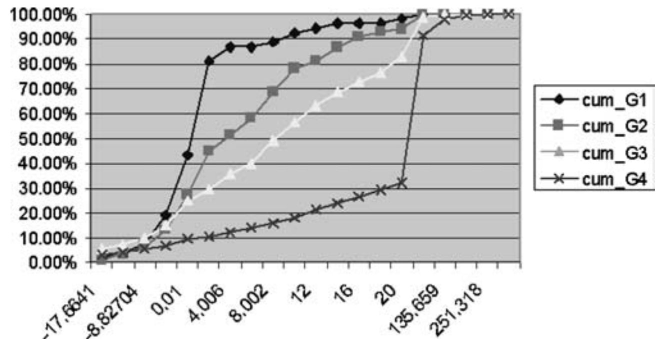

Figure 4. Four-class verifying data set.

In multi-group classification, a better result will be achieved in the separation between certain two groups. According to the definition of group and data attributes, the best KS score usually appears in the separation between the first group and second group or the separation between the last group and other groups. In the model, the last group is defined as individuals with perfect credit performance and other groups are defined as individuals who have some stains in their credit histories. As a result, the separations indicated that the distance between the last group (Good) and other groups is larger and a better KS score for that separation. It means that in practice it is easier to catch good ones or bad ones, but it is much more difficult to discriminant different levels of credit performance in between.

6. Concluding remarks

We have proposed a four-class model by using multiple criteria linear programming to discover the various behaviors of credit cardholders. In credit card portfolio management, predicting the cardholder's spending behavior is a key to reduce the risk of bankruptcy. Given a set of predicting variables (attributes) that describes all possible aspects of credit card holders, we have first built a general classification model that can theoretically handle any size of multiple-class cardholders' behavior problems. Then, we have focused on a special case where the cardholders' behaviors are predefined into four classes: (i) bankrupt charge-off; (ii) non-bankrupt charge-off; (iii) delinquent; and (iv) current. The algorithm of the four-group classification is developed and implemented by SAS. In this paper, a data set of 5000 card account samples from 25,000 credit card records in the data warehouse of a major US bank was used to test the control parameters of the model for the best class separation.

There are a number of research problems remaining unexplored. For example, the range of control parameters in the four-class model may be found through theoretical analyses on the model structure. This can help us to quickly learn the best values of the control parameters in developing practical classifiers. To promote the generality of multiple criteria linear programming classification in data mining, we may produce Linux codes by C++ so that we can utilize IBM or other computing platforms for any available credit data warehouses. These problems are currently under investigation. We will report the significant results in the near future.

Finally, we shall note even though the classification models discussed in this paper have been directly applied in discovering the knowledge for credit card portfolio management, they can be readily used in biomedical research, such as pharmaceutical and DNA analyses; in telecommunications and health cares industries for fraud management; and in retail industries for marketing analysis Han (2001) and Shi (2000).

Acknowledgments

This research has been partially supported by a grant under (DUE-9796243), the National Science Foundation of USA, a National Excellent Youth Fund under (#70028101), National Natural Science Foundation of China and the K.C. Wong Education Foundation (2001 and 2003), Chinese Academy of Sciences.

References

Conover, W.J. (1999). *Practical Nonparametric Statistics*. Wiley.

"Debt Statistics," http://www.nodebt.org/debt.htm.

"Enhancing human health through better decisions—from discovery through development to commercialization—and beyond", SAS Institute Inc, http://www.sas.com/industry/pharma/.

Freed, N. and F. Glover. (1981). "Simple but Powerful Goal Programming Models for Discriminant Problems." *European Journal of Operational Research* 7, 44–60.

Freed, N. and F. Glover. (1986), "Evaluating Alternative Linear Programming Models to Solve the Two-Group Discriminant Problem." *Decision Science* 17, 151–162.

Han, J. and M. Kamber. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers.

http://www.firstdatacorp.com

Joachimsthaler, E.A. and A. Stam. (1988). "Four Approaches to the Classification Problems in Discriminant Analysis: An Experimental Study." *Decision Sciences* 19, 322–333.

Koehler, G.J. and S.S. Erenguc. (1990). "Minimizing Misclassifications in Linear Discriminant Analysis." *Decision Science* 21, 63–85.

Lin, Y. (2002). "Improvement on Behavior Scores by Dual-Model Scoring System." *International Journal of Information Technology and Decision Making* 1, 153–164.

Miller, G.A. (1956). "The Magic Number seven, Plus or Minus Two." *The Psychological Review* 63(2).

Peng, Y., Y. Shi, and W. Xu. (2002). "Classification for Three-group of Credit Cardholders' Behavior Via A Multiple Criteria Approach." *Advanced Modeling and Optimization* 4, 39–56.

Quinlan, J.R. (1986). "Induction of Decision Tree." *Machine Learning* 1, 81–106.

Shi, Y. (2000). "Data Mining," In M. Zeleny (ed), *IEBM Handbook of Information Technology in Business*. International Thomson Publishing, pp. 490–495.

Shi, Y. (2001b). *Multiple Criteria Multiple Constraint-level Linear Programming: Concepts, Techniques and Applications*. World Scientific Publishing.

Shi, Y., Y. Peng, W. Xu and X. Tang. (2002). "Data Mining via Multiple Criteria Linear Programming: Applications in Credit Card Portfolio Management." *International Journal of Information Technology and Decision Making* 1, 131–151.

Shi, Y., M. Wise, M. Luo, and Y. Lin. (2001). "Data Mining in Credit card Portfolio Management: A Multiple Criteria Decision Making Approach." In M. Koksalan and S. Zionts (eds.), *Multiple Criteria Decision Making in the New Millennium*. Berlin: Springer, pp. 427–436.

Shi, Y. and P.L. Yu (1989). "Goal Setting and Compromise Solutions," In B. Karpak and S. Zionts (eds.), *Multiple Criteria Decision Making and Risk Analysis Using Microcomputers*. Berlin: Springer, pp. 165–203.

Stavins, J. (2000), "Credit Card Borrowing, Delinquency, and Personal Bankruptcy." *New England Economic Review,* July/August 2000, http://www.bos.frb.org/economic/neer/neer2000/neer400b.pdf.

"Survey on the Banking Sector's Credit Card Receivables," Hong Kong Monetary Authority, 2000, http://www.info.gov.hk/hkma/eng/press/2001/20010227e6.htm-annex.

"The First Credit Card was Issued in 1951." http://www.didyouknow.cd/creditcards.htm.

www. fairi saac. com