「정보시스템연구」 제16권 제1호
한국정보시스템학회
2007년 3월, pp. 1~21

Consumer Credit Scoring Model with Two-Stage Mathematical Programming*

Sung-Wook Lee** · Tae-Hyup Roh***

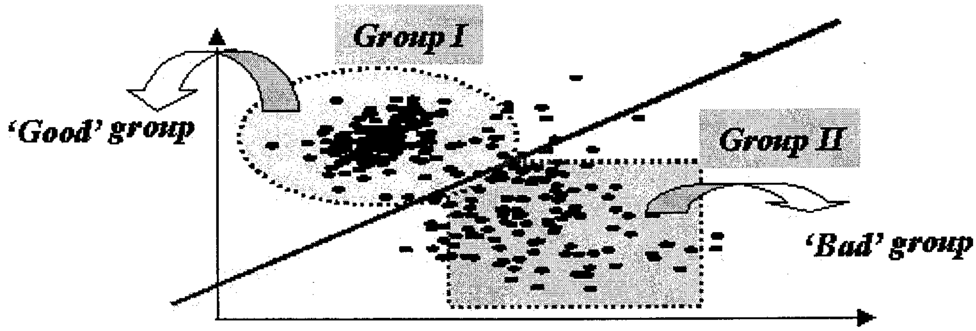
〈목 차〉

Ⅰ. Introduction

Consumer credit is granted by various other lending institutions including banks, building societies, retailers and mail order companies and is a sector of the economy that has seen grown rapidly. Traditional methods of deciding whether to grant credit to a particular individual use human judgment of the risk of default, based on the experience of previous decisions. However, economic pressures resulting from increased demand for credit, allied with greater commercial competition and the emergence of new computer technology, have led to the development of sophisticated statistical models to aid the credit granting decision.

Credit scoring is the name used to describe the process of determining how likely an applicant is to default with repayments. Statistical models which give estimates of these default probabilities are referred to as

* 이 논문은 2007학년도 서울여자대학교 교내특별과제 연구비의 지원을 받았음
** CitiBank Risk Management Group, Researcher, statopia@dreamwiz.com
*** Seoul Women's University, Dept. of Business, Professor, Corresponding Author, rohth@swu.ac.kr



<Figure 1> Graphical illustration of credit scoring

scorecards or classifiers. Standard methods used for developing scorecards include discriminant analysis, logistic regression, decision trees and mathematical programming. An accept/reject decision can then be taken on a particular applicant by comparing the estimated good/bad probability with a suitable threshold. Figure 1 represents graphical illustration of credit scoring.

Credit scoring has two types of decisions that should be made to lend to applicants. The first type of decision is whether they grant credit to a new applicant. Techniques that aid this decision are called application scoring. The second one is how to deal with existing customers. If an existing customer wants to increase his credit limit, should the firm agree to that? If the customer starts to fall behind in his repayments, what actions should the firm take? The tools that help with these decisions are called behavioral scoring. This study will focus on the classification of applicants into good or bad risk classes (application scoring) based on their initial application characteristics.

Credit scoring is one of classification problems whose objective is to predict the group membership of a new observation by using measured values on a set of relevant variables or attributes. Fisher's linear discriminant function and the quadratic discriminant function have long been the standard techniques for establishing discriminant rules in classification analysis (Ragsdale and Stam, 1991). However, both of these discriminant functions are based on the assumption of multivariate normality of the measured variables (attributes). In many situations involving real data, these assumptions are seriously violated, for instance, in the case of binary variables and when outliers are present in the data set.

In credit scoring field, the modern data mining techniques, which have made a significant contribution to the field of information science (Chen & Huang, 2003), can be adopted to construct the credit scoring models. Practitioners and researchers have developed a variety of traditional statistical models and datamining tools for credit scoring, which involve linear discriminant models, logistic regression models, k-nearest neighbor models, decision tree models, neural network models, and genetic programming models.

Desai et al. (1996) investigated neural networks, linear discriminant analysis and logistic regression for scoring credit decision. They concluded that neural networks outperform linear discriminant analysis in classifying loan applicants into good and bad credits, and logistic regression is comparable to neural networks. West (2000) investigated the credit scoring accuracy of several neural networks. Results were benchmarked against traditional statistical methods such as linear discriminant analysis, logistic regression, k-nearest neighbor and decision trees. Malhotra and Malhotra (2002) applied neuro-fuzzy models to analyze consumer loan applications and compared the advantages of neuro-fuzzy systems over traditional statistical techniques in credit-risk evaluation. Hoffmann, Baesens, Martens, Put, and Vanthienen (2002) applied a genetic fuzzy and a neuro-fuzzy classifier for credit scoring. Baesens et al. (2003) benchmarked state-of-the-art classification algorithms for credit scoring.

Recently, researchers have proposed the hybrid data mining approach in the design of an effective credit scoring model. Hsieh (2005) proposed a hybrid system based on clustering and neural network techniques; Lee and Chen (2005) proposed a two-stage hybrid modeling procedure with artificial neural networks and multivariate adaptive regression splines. Lee, Chiu, Lu, and Chen (2002) integrated the backpropagation neural networks with traditional discriminant analysis approach. Chen and Huang (2003) present a work involving two interesting credit analysis problems and resolve them by applying neural networks and genetic algorithms techniques.

As the useful non-parametric techniques, a number of researchers have introduced and investigated mathematical programming (MP) formulations to solve the classification problem, resulting in a number of non-parametric techniques which have been shown to perform well under various conditions (Bajgier and Hill, 1982; Freed and Glover, 1981a, b, 1986a; Gehrlein, 1986; Joachimsthaler and Stam 1988; Koehler and Erenguc, 1990; Stam and Joachimsthaler, 1990). The most common mathematical programming approaches suggested in the literature are the MSD (minimize the sum of the deviations), the MMD (minimize the maximum deviation), and hybrid models which seek to minimize external deviations and maximize internal deviations. Mixed-integer programming (MIP) models have also been suggested to minimize directly the number of misclassified observations (Koehler and Erenguc, 1990; Stam and Joachimsthaler, 1990).

Mathematical programming methods have certain advantages over the parametric methods (Erenguc and Koehler,1990):

(1) Mathematical programming methods are free from parametric assumptions;
(2) Varied objectives and more complex problem formulations are easily accommodated;
(3) Individual weights to each of the data points and misclassification costs, either fixed or depending on the extent of misclassification, are easily incorporated;
(4) Some mathematical programming methods, especially linear programming, lend themselves to sensitivity analysis.

Although the classification performance of these methods is promising, several researchers have pointed out that a number of these mathematical programming formulations suffer from theoretical shortcomings (Markowski and Markowski, 1985; Freed and Glover, 1986b; Koehler, 1989a). These include unacceptable solutions (if a discriminant function of zeros results, in which case all observations will be classified in the same group), improper solutions (if all observations fall exactly on the separating hyperplane), and unbounded solutions (if the objective function can be improved without limit). The outcomes can lead to useless or erroneous results and interpretation (Koehler, 1989b). Of course, the MIP formulations also can require extensive computational resources that may be prohibitive for large data sets.

Therefore, it is necessary for mathematical programming formulations to overcome unacceptable solution, improper solution and computational requirement. This provides the authors with the motivation to propose a two-phase mathematical programming approach which explicitly considers the classification gap associated with a gap constraint formulation. This classification gap can be viewed as a fuzzy area between the groups which requires special consideration in establishing the final classification rule. The effectiveness of two-phase approach is compared with Fisher's linear discriminant function (FLDF), logistic regression, MSD and MIP using empirical data sets.

The reminder of this study is organized as follows. In Chapter 2, literature review of mathematical programming approaches presents. Considerations of the existing mathematical programming approaches and mathematical formulation of the proposed two-phase mathematical programming approach are represented in Chapter 3. Chapter 4 presents the results of computational experiments in order to show the performance of this approach and shows that two-phase mathematical programming approach rivals or outperforms other approaches in terms of the relative classification performance. Finally, some concluding remarks are discussed in Chapter 5.

Ⅱ. Literature review

Among the mathematical programming approaches, the MSD model and the MIP model have been most widely used for discriminant problem in the literature. Typical models of these models are introduced.

2.1 MSD model

One of the first and most widely used mathematical programming models of the discriminant problem is the MSD (minimize the sum of the deviations) model (Freed and Glover, 1981b). In general, the MSD model tries to find a hyperplane that minimizes the weighted sum of exterior deviations. Suppose there are n_k observations in group k ($k=1,2$) on p independent (measured) variables (attributes). The MSD model is given in (P1):

$$\text{(P1)}\quad \text{Min}\ \ z = \mathbf{1}'\mathbf{d}_1 + \mathbf{1}'\mathbf{d}_2 \qquad (1)$$

subject to

$$\mathbf{X}_1\mathbf{w} + \mathbf{d}_1 > c\mathbf{1}, \qquad (2)$$

$$\mathbf{X}_2\mathbf{w} - \mathbf{d}_2 \le c\mathbf{1}, \qquad (3)$$

$$\mathbf{d}_1, \mathbf{d}_2 \ge \mathbf{0} \qquad (4)$$

$$c\ ,\ \mathbf{w}\ \text{unrestricted,} \qquad (5)$$

, where $\mathbf{X}_k$ is an ($n_k \times p$) matrix of observations in group k, the $\mathbf{d}_k$ are ($n_k \times 1$) vectors of deviational variables ($k=1,2$), $\mathbf{1}$ is an appropriately dimensioned column vector of ones, $\mathbf{0}$ is an appropriately dimensioned column vector of zeros, $\mathbf{w}$ is a ($p \times 1$) vector of attribute weights, and c is a scalar variable. In the following discussion, let $\mathbf{x}_{ki}$ represent a ($1 \times p$) vector corresponding to i th observation in group k (i.e., the i th row of $\mathbf{X}_k$), and let d_{ki} represent the i th component of $\mathbf{d}_k$. The value of the variable d_{ki} represents the extent to which observation $\mathbf{x}_{ki}$ is misclassified. For instance, if observation i in group 1 is correctly classified, then $\mathbf{X}_1\mathbf{w} > c\mathbf{1}$, in (2) and the objective in (1) of minimizing the sum of the undesirable deviations implies $d_{1i} = 0$. Similarly, a correctly classified observation belong to group 2 will satisfy $\mathbf{X}_2\mathbf{w} \le c\mathbf{1}$, in (3), and the corresponding deviational variable d_{2i} will equal zero by (1). However, if observation i in group 1 is misclassified then $\mathbf{X}_1\mathbf{w} \le c\mathbf{1}$,, which, by (2), forces d_{2i} to assume a strictly positive value that is penalized in (1). Likewise, (3) ensure that $d_{2i}>0$ for any observation i in group 2 that is misclassified (i.e., $d_{2i}>0$ if and only if $\mathbf{X}_2\mathbf{w}>c\mathbf{1}$,).

The formulation (P1) has considerably intuitive appeal, as its optimal solution ($\mathbf{w}^*, c^*$) identifies a hyperplane in R^k which minimizes the extent of misclassification as measured by the sum of the undesirable deviations from the separating hyperplane for all observations. It is important to note that minimizing the extent of misclassification is not necessarily the same as minimizing the number of misclassification observations. For instance, the MSD model makes no preferential distinction between solutions with $z=100$ and $\mathbf{d}_1=(0,0,0,100)$ or $\mathbf{d}_1=(25,25,25,25)$, even though the first solution has one misclassification and the second has four.

2.2 MIP model

In general, mixed integer programming (MIP) models try to find a separating hyperplane that minimizes the number of misclassifications. The MIP model that was suggested by several authors (Freed and Glover, 1986b; Glover, 1988) is given in (P2):

(P2) $$\text{Min } z=\mathbf{1}'\mathbf{I}_1+\mathbf{1}'\mathbf{I}_2 \qquad (6)$$

subject to

$$\mathbf{X}_1\mathbf{w}+M\cdot\mathbf{I}_1>c\mathbf{1}, \qquad (7)$$

$$\mathbf{X}_2\mathbf{w}-M\cdot\mathbf{I}_2\le c\mathbf{1}, \qquad (8)$$

c , $\mathbf{w}$ unrestricted, (9)

, where M is a large positive number, $\mathbf{I}_1, \mathbf{I}_2$ are zero-one vectors and other notations are same as those of (P1). Let I_{ki} represent the i th component of $\mathbf{I}_k$. The value of the deviational variable I_{ki} represents the extent to which observation $\mathbf{x}_{ki}$ is misclassified. For instance, if observation i in group 1 is correctly classified, then $\mathbf{X}_1\mathbf{w}>c\mathbf{1}$, in (7) and the objective in (6) of minimizing the number of the misclassifications implies $I_{1i}=0$. Similarly, a correctly classified observation which belongs to group 2 will satisfy $\mathbf{X}_2\mathbf{w}\le c\mathbf{1}$, in (8), and the corresponding deviational variable I_{2i} will equal to zero by (7). However, if observation i in group 1 is misclassified then $\mathbf{X}_1\mathbf{w}\le c\mathbf{1}$, which, by (8), forces I_{1i} to assume a strictly positive value that is penalized in (1). Likewise, (3) ensures that $I_{2i}=1$ for any observation i in group 2 that is misclassified (i.e., $I_{i2}=1$ if and only if $\mathbf{X}_2\mathbf{w}>c\mathbf{1}$,).

As mentioned previously, various MIP models have been proposed for directly minimizing the number of misclassifications in the training sample (Koehler and Erenguc, 1990; Stam and Joachimsthaler, 1990). Such methods are inherently insensitive to outliers, since all misclassified observations are weighted equally, irrespective of their distance from the separating hyperplane.

In Chapter 3, this study refers to a number of problems/concerns that researchers should consider in the study of mathematical programming approaches for determining linear discriminant functions. And two-phase mathematical programming approach will be proposed to solve these problems to some degree and outperform the existing approaches.

Ⅲ. Mathematical formulations and solution approach

3.1 Considerations for mathematical programming approaches

The greater part of the literature associated with mathematical programming methods for determining linear discriminant functions falls into two groups: those that give empirical comparisons of the performance of one or more models versus parametric methods and those that point out problems in earlier models. In the following, a number of problems and issues including unacceptable solution, gaps and computational efforts (efficiency) that have appeared or been raised in the literature are explained.

3.1.1 Unacceptable solution

A system of equation of the form

$$\mathbf{X}_1\mathbf{w} \geq c\mathbf{1},$$

$$\mathbf{X}_2\mathbf{w} \leq c\mathbf{1},$$

has a trivial solution of ($\mathbf{w}=\mathbf{0}, c=0$), which gives an unacceptable discrimination - every observation will be classified both groups 1 and 2. However, LP formulations may generate this type of solution (Koehler, 1989a, b).

Many techniques have been suggested to prevent a zero solution. These include:

(1) Adding a linear constraint to prevent $\mathbf{w}=\mathbf{0}$,

(2) Adding a non-convex constraint to prevent $\mathbf{w}=\mathbf{0}$,

(3) Translating the data to prevent $\mathbf{w}=\mathbf{0}$ and

(4) Adding a redundant constraint to prevent $\mathbf{w}=\mathbf{0}$.

All of the above except method (4) have side effects.

A linear equality constraint used to prevent a zero solution takes the form of $\mathbf{a}'\mathbf{w}=1$, where $\mathbf{a}$ is any $p\times 1$ vector. This certainly prevents a zero solution but also prevents any $\mathbf{w}$ in the set $\{\mathbf{w}:\mathbf{a}'\mathbf{w}=0\}$. This overkill is potentially detrimental. So, another normalization constraint is required to solve troublesome above mentioned.

A typical non-convex constraint is $\mathbf{w}'\mathbf{w}>0$.

This constraint only restricts $\mathbf{w}=\mathbf{0}$ so that it is superior to any type of linear constraint to prevent a zero solution. If $(\mathbf{w},c)$ gives n misclassifications, so does $(\lambda\mathbf{w},\lambda c)$ for any $\lambda>0$. Since $(\mathbf{w},c)$ gives the same hyperplane as $(\lambda\mathbf{w},\lambda c)$, one can simplify the above to $\mathbf{w}'\mathbf{w}=1$ without any loss of generality. So, although $\mathbf{w}'\mathbf{w}=1$ restricts $\mathbf{w}=\mathbf{0}$, it dose not restrict consideration of any hyperplanes. (This is not the case with linear constraints. If $\mathbf{w}$ is non-zero and is in $\{\mathbf{w}:\mathbf{a}'\mathbf{w}=0\}$, any scalar multiple of $\mathbf{w}$ is also in $\{\mathbf{w}:\mathbf{a}'\mathbf{w}=0\}$. Hence, a linear constraint with non-zero $\mathbf{w}$ necessarily restricts consideration of some hyperplanes.)

A constraint similar to $\mathbf{w}'\mathbf{w}=1$ is $\|\mathbf{w}\|=1$, where $\|\mathbf{w}\|$ denotes a norm of $\mathbf{w}$. While both type of constraints prevent a $\mathbf{w}=\mathbf{0}$ solution, they change a linear program into a non-convex programming problem, and these very hard to solve.

Constraints above mentioned are non-linear normalization constraints or detrimental to solve classification problem, so appropriate linear normalization constraints are included in the proposed two-phase mathematical programming approach to solve unacceptable solution problem.

3.1.2 Gaps

LP formulations cannot directly handle strict inequality constraints. As seen already, one really wants a solution to

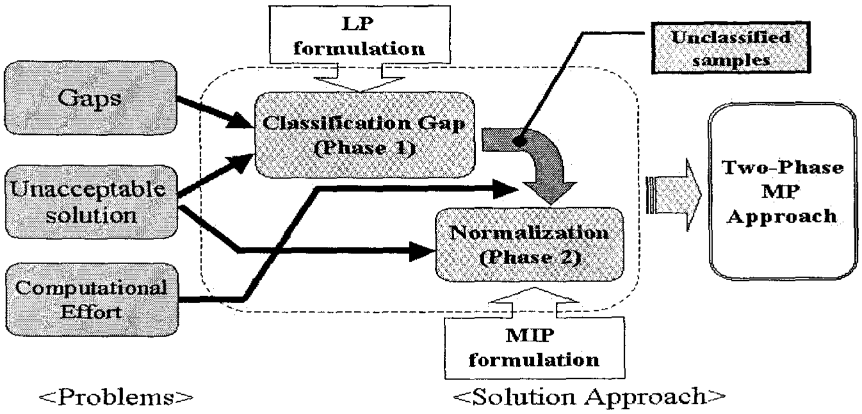
$$\mathbf{X}_1\mathbf{w}>c\mathbf{1},$$
$$\mathbf{X}_2\mathbf{w}\le c\mathbf{1},$$

Most approaches have relaxed the > to ≥. Gehlein (1986) and Glover (1991) replaced the constraint by $\mathbf{X}_1\mathbf{w}>(c+\varepsilon)\mathbf{1}$, where $\varepsilon>0$ and small. They introduced a gap where observations may fall and be unclassified. Because of existence of unclassified observations, classification gap is considered as undesirable in the literature. However, this study does not view classification as undesirable, but merely as an area where additional analysis is required to determine the appropriate classification rule.

3.1.3 Computational effort

Real-world linear discriminant problems typically have a large number of observations ($n=n_1+n_2$) and a small number of attributes (p is usually relatively small). As such, many of linear programming formulations typically have a large number of constraints and a small number of variables (The dual has the opposite properties and accordingly might be the preferred problem to solve.) In either case, polynomial methods exist to solve these problems.

When one considers mixed-integer programming approaches, there are at least n zero-one integer variables. This is a major problem. For this reason, in order to reduce



<Figure 2> Illustration of two-phase mathematical programming approach

the number of observations applied to mixed-integer programming approach, two-phase mathematical programming approach is proposed in this study. After all observations are filtered through phase 1, remaining observations that are not classified in phase 1 are applied to phase 2.

3.2 Two-phase mathematical programming approach.

As mentioned previously in section 3.1, classification gap has been considered as undesirable, however, this study views classification gap as a merely region where the classification decision is not clear and additional analysis is required to determine the appropriate classification rules. Moreover, to prevent unacceptable solution and reduce computational efforts, appropriate normalization constraints are presented in two-phase mathematical programming approach.

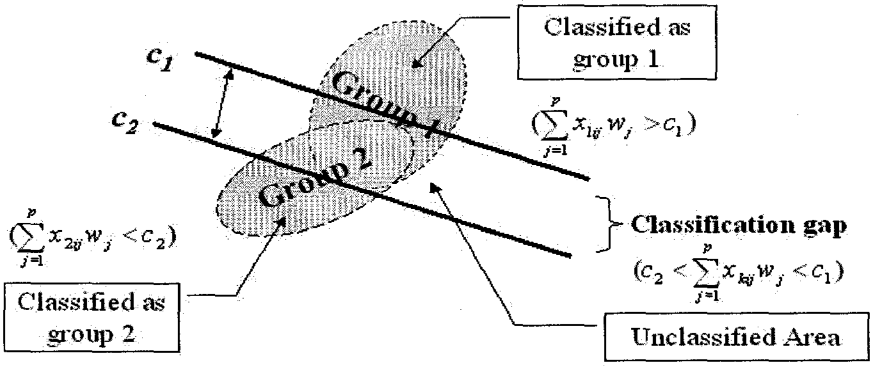
Simple illustration of composition of two-phase mathematical programming approach is presented in Figure 2.

In phase 1, the classification gap is identified, while in phase 2 the explicit focus is to analyze the fuzzy area of observations defined by the classification gap. Problem description and mathematical formulation of Phase 1 and Phase 2 are following.

3.2.1 Phase 1

The objective of phase 1 is to minimize sum of deviations from each classification score. The objective of phase 1 is to minimize sum of deviations from each classification score. Phase 1 is illustrated in Figure 3.

Suppose one has a sample of n_1 in group 1 ('good') and n_2 in group 2 ('bad') $(n=n_1+n_2)$ and a set of p attributes from the application form, x_{kij} is the value of attribute j



<Figure 3> Graphical illustration of phase 1

$(j=1,2,\ldots p)$ in observation i, $(i=1,2,\ldots n_k)$ from group $k(k=1,2)$. Let w_j be the jth attribute weight, c_1 be cut-off value for group 1, c_2 be cut-off value for group 2, d_{1i} be deviation from c_1 when any observation i from group 1 is misclassified and d_{2i} be deviation from c_2 when any observation i from group 2 is misclassified. Phase 1 can be formulated as follows:

(P3) $$\text{Min } z_1 = \sum_{i=1}^{n_1} d_{1i} + \sum_{i=n_1+1}^{n} d_{2i} \quad (10)$$

subject to

$$\sum_{j=1}^{p} x_{1ij} w_j + d_{1i} \geq c_1, \; i=1,2,\ldots,n_1 \quad (11)$$

$$\sum_{j=1}^{p} x_{2ij} w_j - d_{2i} \leq c_2, \; i=n_1+1,\ldots,n \quad (12)$$

$$\sum_{j=1}^{p} x_{1ij} w_j \geq c_2, \; i=1,2,\ldots,n_1 \quad (13)$$

$$\sum_{j=1}^{p} x_{2ij} w_j \leq c_1, \; i=n_1+1,\ldots,n \quad (14)$$

$$c_1 - c_2 \geq 1, \quad (15)$$

$$d_{1i}, d_{2i} \geq 0, \quad (16)$$

$$w_j, c_1, c_2 \text{ unrestricted} \quad (17)$$

Formulation of phase 1 explicitly considers the classification gap and facilitates a useful interpretation of this gap to find appropriate classification scores. If a classification score of any observation is over c_1 or under c_2, then it would be considered as group 1 or group 2, otherwise it would not decide on phase 1. Constraints (13) and (14) are bound constraints for upper and lower classification score. Constraint (13) restricts classification score of observations in group 1 to above c_2. Similarly, constraint (14) classification score of observations in group 2 to under c_1. By constraint (13) and (14), the objective function may be provided a good separation solution. That is, constraints (13) and (14) try to enforce the classification score of observations in group k to be c_k, where $k=1,2$. Constraint (15) is gap constraint. The relative difference between c_1 and c_2 affects the scaling of the parameter estimates for w_j. Therefore, constraint (15) is also normalization constraint. In (P3), c_1 and c_2 are simply defined as two decision variables. Phase 1 has an objective function of minimizing the sum of deviations from each group classification score (c_1, c_2). All observations are filtered through phase 1 and remaining observations, which are not classified in phase 1, are applied to phase 2.

3.2.2 Phase 2

After solving (P3) in phase 1, phase 2 considers only observations that are not classified in phase 1. In phase 2, the objective function is to minimize the weighted sum of misclassified observations. Graphical illustration of phase 2 is represented in Figure 4,

Let m_1 be the number of observations in group 1 which are unclassified in phase 1, m_2 be the number of observations in group 2 which are unclassified in phase 1 ($m=m_1+m_2$), C_{12} be the cost of misclassifying group 1 as group 2 and C_{21} be the cost of misclassifying group 2 as group 1.

In phase 2, the cut-off value c can be determined by solving the following mixed-integer programming (P4):

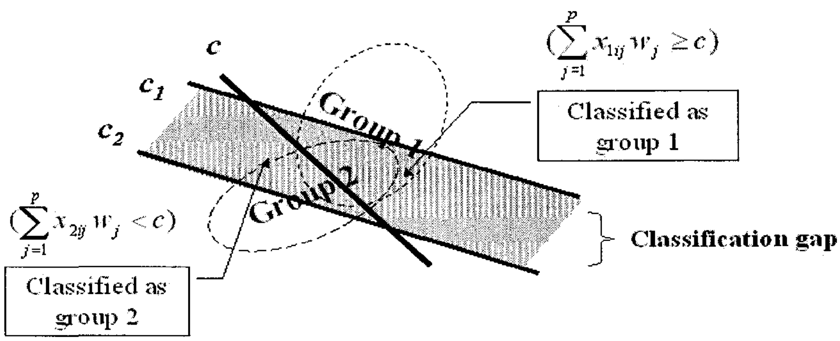
$$\text{(P4)} \quad \text{Min } z_2 = C_{12}\sum_{i=1}^{m_1} I_{1i} + C_{21}\sum_{i=m_1+1}^{m} I_{2i} \qquad (18)$$

subject to

$$\sum_{j=1}^{p} x_{1ij}(w_j^+ - w_j^-) + M \cdot I_{1i} \geq c,$$

$$i = 1,2,\ldots,m_1 \qquad (19)$$

$$\sum_{j=1}^{p} x_{2ij}(w_j^+ - w_j^-) - M \cdot I_{2i} \leq c,$$



<Figure 4> Graphical illustration of Phase 2

$$i = m_2 + 1, \ldots, m \quad (20)$$

$$\sum_{j=1}^{p} (w_j^+ + w_j^-) = 1, \quad (21)$$

$$w_j^+ - \varepsilon \alpha_j \ge 0, \quad j = 1,2,\ldots,p \quad (22)$$

$$w_j^+ - \alpha_j \le 0, \quad j = 1,2,\ldots,p \quad (23)$$

$$w_j^- - \varepsilon \beta_j \ge 0, \quad j = 1,2,\ldots,p \quad (24)$$

$$w_j^- - \beta_j \le 0, \quad j = 1,2,\ldots,p \quad (25)$$

$$\alpha_j + \beta_j \le 1, \quad j = 1,2,\ldots,p \quad (26)$$

c unrestricted (27)

$$w_j^+, w_j^- \ge 0, \ (w_j = w_j^+ - w_j^-) \quad (28)$$

, where I_{ki} is binary variable in observation i from group k, that is, if an observation is misclassified, then $I_{ki} = 1$, otherwise, $I_{ki} = 0$, M is a large positive number.

To prevent unacceptable solution which is mentioned previously in section 3.1, normalization constraints are added to the formulation (P4) (refer to Glen (1999)). Constraints (21)-(26) are normalization constraint. α_j and β_j are binary variables such that $\alpha_j = 1 \Leftrightarrow w_j^+ \ge \varepsilon$ and $\beta_j = 1 \Leftrightarrow w_j^- \ge \varepsilon$, where ε is a small positive number. $\alpha_j = 1$ if and only if w_j^+ is positive and $\beta_j = 1$ if and only if w_j^- is negative.

Procedure of two-phase mathematical programming approach can be summarized as follows:

Step 1: Solve formulation (P3),

Step 2: According to each classification score from (P3), classify observations into each group.

Step 3: Solve formulation (P4) with observations which are not classified into group 1 or group 2 in step 2.

Step 4: According to the final cut-off value (classification score), classify unclassified observations into each group.

Through this procedure, two-phase mathematical programming approach can classify loan applicants more accurate than the existing mathematical programming approaches.

In Chapter 4, the performance of two-phase mathematical programming approach is compared with that of other existing approaches by experimenting on real managerial problems computationally. And the results show that two-phase mathematical programming approach may be a good alternative to other statistical approaches.

Ⅳ. Computational experiments

To test the effectiveness of the proposed approach, this study compared it with other approaches. Two-phase mathematical programming approach was implemented by CPLEX mathematical programming solver. This solver was also used to solve the MSD and MIP formulations discussed in chapter 2.

In two-phase mathematical programming approach, in order to solve formulation (P4) of phase 2, each parameter was set up as following: $C_{12}=1$, $C_{21}=5$, $M=1000$, $\varepsilon=0.001$. Cost matrix for all dataset (Michie, Spiegelhalter, and Taylor, 1994) is represented in Table 1. These costs are what are called "opportunity costs". The columns are the predicted class and the rows the true class.

<Table 1> Cost Matrix for the datasets

	Good (1)	Bad (2)
Good (1)	0	1
Bad (2)	5	0

Logistics regression, Fisher's linear discriminant function, MSD, MIP and two-phase mathematical programming approach were applied to two data sets. The Fisher's linear discriminant function and logistic regression were used to calculate discriminant function by using SAS. All computations were carried out on a Pentium-III computer.

The data set, German credit data is from the Department of Statistics, University of Munich (http://www.stat.uni-muenchen.de). The qualitative attributes are given a score that is based on the assessment of experienced bank specialists dealing with credits (see Appendix). German credit data set contains 400 applicants with 280 being accepted and 120 being rejected. Usually the information needed by the decision maker is given on the application form. The data set consists of twenty attributes which are listed as follows:

1) x_1: Status of checking account (qualitative),
2) x_2: Duration in months (numerical),
3) x_3: Credit history (qualitative),
4) x_4: Purpose (qualitative),
5) x_5: Credit amount (numerical),
6) x_6: Savings account/bonds (qualitative),
7) x_7: Present employment since (qualitative),
8) x_8: Installment rate in percentage of disposable income (numerical),
9) x_9: Personal status and sex (qualitative),
10) x_{10}: Other debtors/ guarantors (qualitative),
11) x_{11}: Present residence since (numerical),
12) x_{12}: Property (qualitative),
13) x_{13}: Age in years (numerical),
14) x_{14}: Other installment plans (qualitative),
15) x_{15}: Housing (qualitative),
16) x_{16}: Number of existing credits at this bank (numerical),
17) x_{17}: Occupation (qualitative),
18) x_{18}: Number of people being liable to provide maintenance for (numerical),
19) x_{19}: Telephone (binary),
20) x_{20}: Foreign worker (binary),

<Table 2> Hit ratio of five approaches to German credit data

Method	Sample	Correctly accepted	Erroneously accepted	Correctly rejected	Erroneously rejected	Hit-ratio
Logistic Regression	Training samples	122	29	34	15	0.7800
	Validation samples	118	32	28	25	0.7150
FLDF	Training samples	105	15	48	32	0.7650
	Validation samples	104	16	41	39	0.7250
MSD	Training samples	107	35	28	30	0.6750
	Validation samples	103	34	23	40	0.6300
MIP	Training samples	104	28	35	33	0.6950
	Validation samples	105	39	18	38	0.6150
Two-Phase MP	Training samples	113	13	50	24	0.8150
	Validation samples	105	22	35	38	0.7000

For German credit data, to test the predictive power of the classification techniques, 200 applicants are chosen as the training samples and the remaining 200 applicants are used as validation samples. Among the attributes (variables), status of checking account (x_1), credit history (x_3), other debtors/guarantors (x_{10}), property (x_{12}), job (x_{17}) and telephone (x_{19}) are considered as common important factors in five approaches.

The hit ratios and misclassification cost of five approaches to German credit data are reported in Table 2 and Table 3, respectively.

<Table 3> Misclassification cost of five approaches to the German credit data

Sample \ Method	Logistic regression	FLDF	MSD	MIP	Two-phase MP
Training samples	160	107	205	173	89
Validation samples	185	119	210	233	148

In Table 2, two-phase mathematical programming shows the higher hit ratio than other approaches in training sample. Moreover, misclassification cost of two-phase mathematical programming is less than that of any other approaches in training sample in Table 3.

In case of German credit data, 13 attributes are qualitative or binary variables among attributes, therefore the multivariate normality assumption underlying parametric statistical technique such as Fisher' linear discriminant function is being violated. Under this situation, two-phase mathematical approach may be also a good alternative to parametric statistical techniques. Similarly bankruptcy firm data, two-phase mathematical programming also outperformed other existing mathematical programming approaches.

Overall, experimental results conclude that two-phase mathematical programming approach may be a good alternative to other statistical approaches and an improving approach of the existing mathematical programming approaches.

V. Conclusions

In this study, two-phase mathematical programming approach is introduced for solving the credit scoring problem. This approach differs from previous formulations in that it explicitly considers the classification gap and provides a means for classifying observations which fall within this area. By using linear programming (LP) considering classification gap, phase 1 makes decision to grant credit, deny credit, or to seek additional information before making a decision. Phase 2 finds a cut-off value, which minimizes the misclassification cost of granting credit to 'bad' or denying credit to 'good' by using mixed-integer programming (MIP). However, the assumption of the cost matrix could be changeable depending on the policy of organization, which can be affected to the results.

The purpose of this study has been tested whether this approach perform as well as other statistical approaches do. In the empirical test carried out here on German credit data, this approach outperformed the existing mathematical programming approaches and other statistical approaches.

This study concludes that two-phase mathematical programming approach can be a good or better alternative to statistical approaches and traditional mathematical approaches to credit scoring and other discriminant problems.

References

Baesens, B., T. Van Gestel, S. Viaene, M.

Stepanova, J. Suykens and J. Vanthienen. (2003). Benchmarking State-of-the-art Classification Algorithms for Credit Scoring, *Journal of the Operational Research Society*, Vol. 54, No.6, pp. 627-635.

Bajgier, S.M. and Hill, A.V. (1982), An Experimental Comparison of Statistical and Linear Programming Approaches to the Discriminant Problem *Decision Science*, Vol. 13, pp. 604-618.

Chen, M.C. and S.H. Huang. (2003). Credit Scoring and Rejected Instances Reassigning Through Evolutionary Computation Techniques, *Expert Systems with Applications*, Vol. 24 No. 4, pp. 433-441.

Desai, V.S., J.N. Crook and G.A. Overstreet. (1996). A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment, *European Journal of Operational Research*, Vol. 95, No. 1, pp. 24-37.

Erenguc, S.S. and Koehler, G.J. (1990), Survey of Mathematical Programming Models and Experimental Results for Linear Discriminant Analysis, *Managerial and Decision Economics*, Vol. 11, pp. 215-225.

Freed, N. and Glover, F. (1981a), A LP Approach to the Discriminant Problem *Decision Science*, Vol. 12, pp. 68-74.

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

Freed, N. and Glover, F. (1986a), Evaluating Alternative LP Models to Solve the Two-Group Discriminant Problem, *Decision Science*, Vol. 17, pp. 151-162.

Freed, N. and Glover, F. (1986b), Resolving Certain Difficulties and Improving the Classification Power of LP Discriminant Analysis Formulations, *Decision Science*, Vol. 17, pp. 589-595.

Gehrlein, W.V. (1986), General Mathematical Programming Formulations for the Statistical Classification Problem, *Operations Research Letters*, Vol. 5, pp. 299-304.

Glen J.J. (1999), IP Methods for Normalization and Variable Selection in Mathematical Programming Discriminant Analysis Models. *JORS*, Vol. 50, pp. 1043-1053.

Glover, F. (1991), Improved Linear and Integer Programming Models for Discriminant Analysis, *Creative and Innovative Approaches to the Science of Management*, RGK Foundation Press, pp. 187-215.

Glover, F., Keene, S. and Duea, B. (1988), A New Class of Models for the Discriminant Problem *Decision Science*, Vol. 19, pp. 269-280.

Hoffmann, F., B. Baesens, J. Martens, F. Put and J. Vanthienen. (2002). Comparing a Genetic Fuzzy and a Neurofuzzy Classifier for Credit Scoring, *International Journal of Intelligent*

Systems, Vol. 17 No.11, pp. 1067-1083.

Hsieh, N.-C. (2005). Hybrid Mining Approach in the Design of Credit Scoring Models, *Expert Systems with Applications,* Vol. 28, No.4, pp. 655-665.

Joachimsthaler, E.A. and Stam, A. (1988), Four Approaches to the Classification Problem in Discriminant Analysis: An Experiment Study, *Decision Science,* Vol. 19, pp. 322-333.

Koehler, G.J. (1989a), Characterization of Unacceptable Solutions in LP Discriminant Analysis, *Decision Science,* Vol. 20, pp. 239-257.

Koehler, G.J. (1989b), Unacceptable Solutions and the Hybrid Discriminant Model, *Decision Science,* Vol. 20, pp. 844-848.

Koehler, G.J. and Erenguc, S.S. (1990), Minimizing Misclassification in Linear Discriminant Analysis, *Decision Science,* Vol. 21, pp. 63-85.

Lee T.-S. and I.-F. Chen. (2005). A Two-stage Hybrid Credit Scoring Mmodel Using Artificial Neural Networks and Multivariate Adaptive Regression Splines, *Expert Systems with Applications,* Vol. 28, No.4, pp. 743-752.

Lee, T.-S. C.-C. Chiu, C.-J. Lu and I.-F. Chen. (2002). Credit Scoring Using the Hybrid Neural Discriminant Technique, *Expert Systems with Applications,* Vol. 23, No.3, pp. 245-254.

Malhotra, R. and D.K. Malhotra. (2002). Differentiating between Good Credits and Bad Credits Using Neuro-fuzzy Systems, *European Journal of Operational Research,* Vol. 136, No.1, pp. 190-211.

Markowski, E.P. and Markowski, C.A. (1985), Some Difficulties and Improvements in Applying LP Formulations to the Discriminant Problem, *Decision Science,* Vol. 16, pp. 237-247.

Michie, D., Spiegelhalter, D.J. and Taylor, C.C. (1994), Machine Learning, Neural and Statistical Classification, P-H Gen Re Press.

Ragsdale,C.T. and Stam, A. (1991), Mathematical Programming Formulations for the Discriminant Problem: An Old Dog Does New Tricks, *Decision Science,* Vol. 22, pp. 296-307.

Stam, A. and Joachimsthaler, E.A. (1990), A Comparison of a Robust Mixed-Integer Approach to Existing Methods for Establishing Classification Rules for the Discriminant Problem, *European Journal of Operational Research,* Vol. 46, pp. 113-122.

The Department of Statistics, University of Munich, http//www.stat.uni-muenchen.de/

West, D. (2000). Neural Network Credit Scoring Models, *Computers and Operations Research,* Vol. 27 No.11~12, pp. 1131-1152.

Appendix. German credit data

The following given score for the categorical (qualitative/binary) variables is based on the assessment of experienced bank specialists dealing with credits.

Variable	Description	Categories	Score
Status	Balance of current account	No balance or debit	2
		0 <= … < 200 DM	3
		… >= 200 DM or checking account for at least 1 year	4
		No running account	1
Credit history	Payment of previous credits	No previous credits / paid back all previous credits	2
		Paid back previous credits at this bank	4
		No problems with current credits at this bank	3
		Hesitant payment of previous credits	0
		Problematic running account / there are further credits running but at other banks	1
Purpose	Purpose of credit	New car	1
		Used car	2
		Items of furniture	3
		Radio / TV	4
		Household appliances	5
		Repair	6
		Education	7
		Vacation	8
		Retaining	9
		Business	10
		Others	0
Savings	Value of savings or stocks	< 100, DM	2
		100, <= … < 500, DM	3
		500, <= … < 1000, DM	4
		>= 1000, DM	5
		Not available / no savings	1
Employ ment	Has been employed by current employer for	Unemployed	1
		<= 1 year	2
		1 <= … < 4 years	3

		4 <= … < 7 years	4
		>= 7 years	5
Marital	Marital Status / Sex	Male: divorced / living apart	1
		Female: divorced / living apart / married/ male: single	2
		Male: married / widowed	3
		Female: single	4
Debtor	Further debtors / Guarantors	None	1
		Co-Applicant	2
		Guarantor	3
Property	Most valuable available assets	Ownership of house or land	4
		Savings contract with a building society / Life insurance	3
		Car / Other	2
		Not available / no assets	1
Credits	Further running credits	At other banks	1
		At department store or mail order house	2
		No further running credits	3
Housing	Type of apartment	Rented flat	2
		Owner-occupied flat	3
		Free apartment	1
Job	Occupation	Unemployed / unskilled with no permanent residence	1
		Unskilled with permanent residence	2
		Skilled worker / skilled employee / minor civil servant	3
		Executive / self-employed / higher civil servant	4
Phone	Telephone	No	1
		Yes	2
Foreign	Foreign worker	Yes	1
		No	2

이성욱(Sung-Wook Lee)

연세대학교 응용통계학과에서 학부를 마쳤으며, 석사는 한국과학기술원 산업공학과에서 공학석사를 취득하였다. 한국씨티은행에서 소비자금융 및 리스크관리 업무를 수행하였으며, 현재 한국시티은행에서 개인여신리스크관리부 차장으로 활동하고 있다.

노태협(Tae-Hyup Roh)

한국과학기술원에서 경영공학박사와 석사를 취득하였으며, 연세대학교 응용통계학과에서 학부를 졸업하였다. 현재 서울여자대학교 경영학과에서 교수로 재직 중이며, 한국경영정보학회, 한국지능정보시스템학회, 한국정보시스템학회 등에서 활동하고 있다. 주요 관심분야는 재무/회계정보시스템, 고객관계관리, 데이터마이닝 등이다.

<요 약>

통합 수리계획법을 이용한 개인신용평가모형

이성욱 · 노태협

신용평점을 위한 부도예측의 분류 문제를 다루는데 있어서 통계적 판별분석 및 인공신경망 및 유전자알고리즘 등을 이용한 데이터 마이닝의 방법들이 일반적으로 고려되어왔다. 이 연구에서는 수리계획법을 응용하여 classification gap을 고려한 이단계 수리계획 접근방법을 신용평가에 적용하는 방법론을 제안하여 수리계획법을 통한 신용평가모형 구축의 가능성을 제시한다. 1단계에서는 선형계획법을 이용해서 대출 신청자에게 대출을 허가할 것인지의 여부를 결정하게 되는 대출 심사 filtering으로의 적용단계이고, 2단계에서는 정수계획법을 이용하여 오분류 비용이 최소가 되도록 하는 판별점수를 찾는 과정으로 모형을 구성한다.

개인 대출 신청자의 데이터(German Credit Data)에 대하여 피셔의 선형 판별함수, 로지스틱 회귀모형 및 기존의 수리계획 기법들과의 비교를 통해서 제안된 모델의 성능을 평가한다. 이단계 수리계획 접근법의 평가 결과를 통하여 신용평가모형에의 적용가능성을 기존 통계적인 접근방법 및 수리계획 접근법과 비교하여 제시하고 있다.

Keywords: 개인신용평가, 수리계획법, 판별분석, 다단계분석법

* 이 논문은 2006년 6월 30일 접수하여 2차 수정을 거쳐 2006년 11월 30일 게재 확정되었습니다.