대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일~4일

두 단계 수리계획 접근법에 의한 신용평점 모델

(Credit Score Modelling in A Two-Phase Mathematical Programming)

Chang Sup Sung, Sung Wook Lee
Department of Industrial Engineering, KAIST
373-1 Guseong-dong Youseong-gu, Daejon 305-701, Korea

Abstract

This paper proposes a two-phase mathematical programming approach by considering classification gap to solve the proposed credit scoring problem so as to complement any theoretical shortcomings. Specifically, by using the linear programming (LP) approach, phase 1 is to make the associated decisions such as issuing grant of credit or denial of credit to applicants, or to seek any additional information before making the final decision. Phase 2 is to find a cut-off value, which minimizes any misclassification penalty (cost) to be incurred due to granting credit to 'bad' loan applicant or denying credit to 'good' loan applicant by using the mixed-integer programming (MIP) approach. This approach is expected to find appropriate classification scores and a cut-off value with respect to deviation and misclassification cost, respectively.

Statistical discriminant analysis methods have been commonly considered to deal with classification problems for credit scoring. In recent years, much theoretical research has focused on the application of mathematical programming techniques to the discriminant problems. It has been reported that mathematical programming techniques could outperform statistical discriminant techniques in some applications, while mathematical programming techniques may suffer from some theoretical shortcomings.

The performance of the proposed two-phase approach is evaluated in this paper with firm data and loan applicants data, by comparing with three other approaches including Fisher's linear discriminant function, logistic regression and some other existing mathematical programming approaches, which are considered as the performance benchmarks. The evaluation results show that the proposed two-phase mathematical programming approach outperforms the aforementioned statistical approaches. In some cases, two-phase mathematical programming approach marginally outperforms both the statistical approaches and the other existing mathematical programming approaches

1. Introduction

1.1 Background and Motivation

Consumer credit is granted by various other lending institutions including banks, building societies, retailers and mail order companies, which is a sector of the economy that has been grown rapidly. Traditional methods of deciding whether to grant credit to a particular individual have used human judgment on 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 *scorecards* or *classifiers*. Standard methods used for developing scorecards include discriminant analysis, logistic regression, decision tree 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 a graphical illustration of credit scoring.

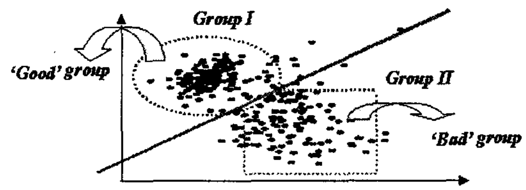

Fig. 1. A Graphical illustration of Credit Scoring

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.

Recently, a number of researchers have introduced and investigated mathematical programming (MP) formulations to solve the classification problem, resulting in a number of useful 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), Stam and Joachimsthaler (1990), Koehler and Erenguc (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 (1989)). 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). Therefore, the outcomes can lead to useless or erroneous results and interpretation (Koehler (1989)). Of course, the MIP formulations also can require extensive computational resources that may be prohibitive for large data sets.

Thus, 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 the two-phase approach is then compared with the aforementioned other methods including Fisher's linear discriminant function (FLDF), logistic regression, MSD and MIP using empirical data sets.

1.2 Organization of The paper

The remainder of this paper is organized as follows. In Section 2, the literature review on mathematical programming approaches is presented. Considerations of the existing mathematical programming approaches and a mathematical formulation of the proposed two-phase mathematical programming approach are represented in Section 3. Section 4 presents the results of computational experiments to show that the two-phase mathematical programming approach outperforms or is as good as the other approaches in the literature in terms of the relative classification performance. Finally, some concluding remarks are discussed in Section 5.

2. Literature Review

Among the mathematical programming approaches in the literature, two typical models including the MSD model and the MIP model have been most widely used for discriminant problems. The two typical models are now introduced.

2.1 MSD model

One of the most widely used mathematical programming models for discriminant problems 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 any associated exterior deviations. Suppose that there are observations in group (,) on independent (measured) variables (attributes). The MSD model is then given as in Problem **(P1)**.

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

subject to

(2)

(3)

(4)

, unrestricted, (5)

where is an ( ) matrix of observations in group , are the ( ) vectors of deviational variables (,), represents an appropriately dimensioned column vector of ones, represents an appropriately dimensioned column vector of zeros, denotes a ( ) vector of attribute weights, and is a scalar variable. Let represent a ( ) vector corresponding to the th observation in group (i. e., the th row of), and let represent the th component of . The value of the variable represents the extent to which observation is misclassified. For instance, if observation in group 1 is correctly classified, then the relation will hold in Eq. (2) and the objective in Eq. (1) of minimizing the sum of any undesirable deviations will imply that . Similarly, a correctly classified observation belonging to group 2 will satisfy the relation in Eq. (3), and the corresponding deviation variable will be equal to zero by Eq. (1). However, if observation in group 1 is misclassified, then the relation holds, which, by Eq. (2), forces to be a strictly positive value that is penalized as in Eq. (1). Likewise, Eq. (3) ensures that the relation holds for any observation in group 2 that is misclassified (i. e., if and only if).

Problem **(P1)** appeals intuitively, as its optimal solution ( ) identifies a separating hyperplane in which minimizes the extent of misclassification as measured by the sum of any 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 and $\mathbf{d}_1 = (0,0,0,100)$ or $\mathbf{d}_1 = (25,25,25,25)$, even though the first solution has one misclassification but the second one has four misclassifications.

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 has been suggested by several authors (Freed and Glover (1986b), Glover (1988)) is given as in Problem (**P2**):

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

subject to

(7)

(8)

, unrestricted, (9)

where represents a large positive number, denote zero-one vectors and the other notations are the same as those of Problem (**P1**). Let represent the th component of . The value of the deviational variable represents the extent to which observation is misclassified. For instance, if observation in group 1 is correctly classified, then the relation holds in Eq. (7) and the objective in Eq. (6) of minimizing the number of misclassifications implies the relation . Similarly, a correctly classified observation belonging to group 2 will satisfy the relation in Eq. (8), and the corresponding deviational variable will be equal to zero by Eq. (7). However, if observation in group 1 is misclassified, then the relation holds, which, by Eq. (8), forces to be a strictly positive value that is penalized in Eq. (1). Likewise, Eq. (3) ensures that the relation holds for any observation in group 2 that is misclassified (i. e., if and only if).

As mentioned above, various MIP models have been proposed for minimizing directly 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 Section 3, this paper refers to a number of problems/concerns that researchers should consider in the study of mathematical programming approaches for determining linear discriminant functions. Thereupon, a two-phase mathematical programming approach will be proposed to solve those problems to some degree and to show its outperforming the existing approaches in the literature.

3. Mathematical Formulations and Solution Approach

3.1 Considerations of mathematical programming approaches

Most of the literatures considering mathematical programming methods for determining linear discriminant functions are classified into two groups: one group that gives empirical performance comparisons between their models and parametric methods and the other group that points out problems from any other earlier models. Therewith, a number of problems and issues including unacceptable solution, gap and computational effort (efficiency) that have appeared or been raised in the literature are now explained.

3.1.1 Unacceptable solution

A system of equations of the form

has a trivial solution of ( ), which gives an unacceptable discrimination, so that every observation will be classified into both groups 1 and 2. For instance, LP formulations may generate this type of solution (Koehler (1989a, b)).

A variety of different techniques has been suggested to prevent it from having a zero solution. These include

(1) Adding a linear constraint to prevent ,

(2) Adding a non-convex constraint to prevent ,

(3) Translating the data to prevent and

(4) Adding a redundant constraint to prevent .

All of them, except method (4), have side effects.

A linear equality constraint used to prevent a zero solution takes the form of , where is a vector. This certainly prevents a zero solution but also prevents any from being included in the set . This overkill is potentially detrimental. Therefore, another normalization constraint is required to solve such troublesome.

A typical non-convex constraint is . This constraint is only restricted to have so that it is superior to any other type of linear constraint to prevent a zero solution. If ( ) gives misclassifications, so does ( ) for any . Since ( ) gives the same hyperplane as ( ), one can simplify the above constraint to be without any loss of generality. Therefore, although the constraint is restricted to be , it dose not restrict considering any hyperplanes. (This is not the case with linear constraints. If is non-zero and included in the set , then any scalar multiple of is also included in the set . Hence, a linear constraint with non-zero is necessarily to restrict considering some hyperplanes.)

A constraint similar to is , where denotes a norm of . While both types of constraints prevent a solution, they change a linear program into a non-convex programming problem, which is very hard to solve.

The above-mentioned constraints are non-linear normalization constraints or detrimental to solve classification problems, so that it is necessarily to include appropriate linear normalization constraints in the proposed two-phase mathematical programming approach to solve any unacceptable solution problems.

3.1.2 Gap

LP formulations cannot directly handle any strict inequality constraints. As seen above, it is interesting to find a solution to the constraints

Most approaches have relaxed the inequality (>) constraint to the equal-or-greater-than ( ) constraint. Gehlein (1986) and Glover (1991) have replaced the constraint by , , where and is small. They have introduced a gap where observations may fall into the gap and be unclassified. Because of existence of unclassified observations, such classification gap has been considered as being undesirable in the literature. However, this paper does not want to view such classification gap as being undesirable, but merely view it 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 ( ) and a small number of attributes (is usually relatively small). For example, many linear programming formulations typically have a large number of constraints and a small number of variables, which its dual has the opposite structure so as to be more preferable to handle. In either case, polynomial methods exist to solve them though.

When one considers mixed-integer programming approaches, at least zero-one integer variables will be involved, which is a major problem. For this reason, in order to reduce the number of observations applied to any mixed-integer programming approach, a two-phase mathematical programming approach is proposed in this paper. After all observations are filtered through Phase 1, any remaining observations that are not classified yet 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 being undesirable, while, however, this paper views classification gap as a merely region such that the classification decision is not clear in the region and so 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 a two-phase mathematical programming approach.

A simple illustration of the composition of the proposed two-phase mathematical programming approach is depicted as in Figure 2.

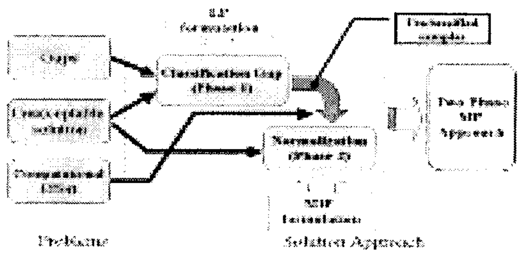

Fig. 2. Illustration of the Proposed Two-Phase Mathematical Programming Approach

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. The problem descriptions and mathematical formulations of Phase 1 and Phase 2 are now discussed.

3.2.1 Phase 1

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

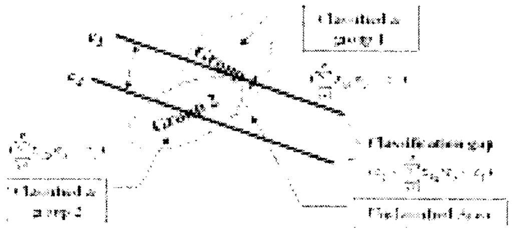

Fig. 3. A Graphical illustration of Phase 1

Suppose that one has a sample of size in Group 1 ('good') and in Group 2 ('bad') ( ) and an element of the set of attributes from the application form, say ,denotes the value of attribute ($j=1,2,\cdots,p$) in observation ($i=1,2,\cdots,n_k$) from Group . Let be the th attribute weight, be the cut-off value for Group 1, be the cut-off value for Group 2, be deviation from when any observation from Group 1 is misclassified and be deviation from when any observation from Group 2 is misclassified. Phase 1 can be formulated as Problem **(P3)**.

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

subject to

$\sum^{p} x_{1ij} w_j + d_{1i} \geq c_1, \quad i=1,2,\cdots,n_1$ (11)

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

$$\sum_{=1}^{p} x_{1} \; w \; \geq c_2, \qquad i=1,2,\cdots,n_1 \quad (13)$$

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

(15)

(16)

unrestricted (17)

The formulation of Phase 1 explicitly considers the classification gap so as to facilitate a useful interpretation of the gap in finding appropriate classification scores. If a classification score of any observation is over or under , then it would be considered as Group 1 or Group 2, respectively; otherwise, it would not decide at Phase 1. Constraints (13) and (14) are bound constraints for upper and lower classification scores. Constraint (13) restricts the classification score of observations in Group 1 to be above . *Similarly, constraint (14) restricts the classification* score of observations in Group 2 to be under . By constraints (13) and (14), the objective function may provide a good separation solution. That is, constraints (13) and (14) try to enforce the classification score of observations in Group to be , for . Constraint (15) is a gap constraint. The relative difference between and affects the scaling of parameter estimates for . Therefore, constraint (15) is also a normalization constraint. In Problem **(P3)**, and are simply defined as two decision variables. Phase 1 has an objective function of minimizing the sum of deviations from each group classification score (,). All observations are filtered through Phase 1 and the remaining observations, which are not classified in Phase 1, are applied to Phase 2.

3.2.2 Phase 2

After solving Problem **(P3)** in Phase 1, Phase 2 will consider only the observations that are not classified yet in Phase 1. In Phase 2, the objective function is to minimize the weighted sum of misclassified observations. A Graphical illustration of Phase 2 is depicted as in Figure 4

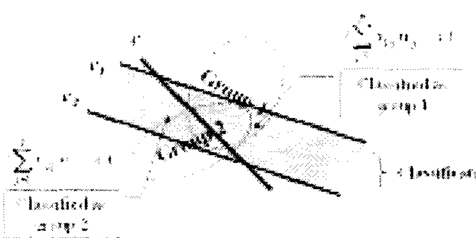

Fig. 4. A Graphical illustration of Phase 2

Let be the number of observations in Group 1 which are unclassified in Phase 1, be the number of observations in Group 2 which are unclassified in Phase 1 ($m=m_1+m_2$), be the cost of misclassifying Group 1 as Group 2 and be the cost of misclassifying Group 2 as Group 1.

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

$$\textbf{(P4)} \quad \text{Min}\, z = C_1 \sum_{=1}^{m_1} I_1 + C_1 \sum_{=m_1+1}^{m} I \qquad (18)$$

subject to

$$\sum_{j=1}^{p} x_{1ij}(w_j^+ - w_j^-) + M \cdot I_{1i} \geq c, \quad i=1,2,\cdots,m_1 \quad (19)$$

$$\sum_{j=}^{p} x_{2ij} \; w_j^+ - w_j^- \; - M \cdot I_{2i} \leq c, \quad i=m_1+1,\cdots,m \quad (20)$$

$$\sum^{p} (w_j^+ + w_j^-) = 1, \qquad (21)$$

$$j=1,2,\cdots,p \qquad (22)$$

$$j=1,2,\cdots,p \qquad (23)$$

$$j=1,2,\cdots,p \qquad (24)$$

$$j=1,2,\cdots,p \qquad (25)$$

$$j=1,2,\cdots,p \qquad (26)$$

unrestricted (27)

(28)

where is a large positive number, is the binary variable indicating whether observation from group is misclassified or not, that is, if an observation is misclassified, then , otherwise, .

To prevent any unacceptable solution, mentioned previously as in Section 3.1, some normalization constraints are added to Problem (P4) (referring to Glen (1999)). Constraints (21)—(26) are normalization constraints. and are binary variables such that and , where is a small positive number. if and only if is positive and if and only if is negative.

The procedure of the proposed two-phase mathematical programming approach is summarized as follows:

Step 1: Solve Problem **(P3)**,

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

Step 3: Solve Problem **(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, the proposed two-phase mathematical programming approach can classify loan applicants more accurately than any existing mathematical programming approaches in the literature.

In Paper 4, the performance of the proposed two-phase mathematical programming approach is compared with that of the other existing approaches by experimenting with real managerial problems computationally. The comparison results show the proposed that two-phase mathematical programming approach is at least as good as other statistical approaches.

4. Computational Experiments

To test the effectiveness of the proposed approach, this Paper compares it with the other approaches in the literature. The proposed Two-phase mathematical programming approach was implemented by the CPLEX mathematical programming solver. This solver was also used to solve the MSD and MIP formulations discussed as in Section 2. In the two-phase mathematical programming approach, to solve Problem **(P4)** of Phase 2, each parameter was set up as [illegible]. The cost matrix for all the data sets (Michie, D., Spiegelhalter, D.J. and Taylor, C.C. (1994)) is given in Table 1. These costs are what are called "opportunity costs". The columns are the predicted classes and the rows are the true classes.

Table 1. Cost Matrix for the data sets

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

All the approaches including 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 approach were used to calculate the discriminant function by using SAS. All computation was carried out on a Pentium-III computer.

Bankruptcy firm data and German credit data were taken as examples for computational experiments. These approaches were compared one against another in hit-ratio that is the ratio of the correctly classified observations to the classified observations, and also in cost that is the sum of misclassification costs based on the cost matrix.

4.1 Bankruptcy firm data

Bankruptcy firm data set (Johnson and Wichern (1988)) represents financial information for 46 firms on four indices from *Moody's Industrial Manuals* regarding firms that went bankrupt or remained solvent during a two-year interval following the measurement of the attributes. This dataset consists of 21 bankrupt firms collected over a period prior to their bankruptcy and for 25 non-bankrupt firms collected over a period of the same duration. The data set consists of four numeric attributes.

For bankruptcy firm data, the whole sample is used for implementation. And in order to test the predictive power of the classification techniques, 32 firms are chosen as the training samples and the remaining 14 firms are used as validation samples.

The hit ratios and misclassification costs of the five approaches to bankruptcy firm data are listed in Table 2 and Table 3, respectively.

Table 2. Hit ratio of the five approaches to the bankruptcy firm data

Method	Sample	Correctly accepted	Wrongly accepted	Correctly rejected	Wrongly rejected	Hit Ratio
LR*	Whole	23	3	18	2	0.8913
	Training	13	2	14	3	0.9063
	Validation	5	0	6	3	0.7857
FLDF	Whole	24	3	18	1	0.9130
	Training	14	2	13	3	0.8438
	Validation	6	0	6	2	0.8571
MSD	Whole	23	4	17	2	0.8696
	Training	15	1	14	2	0.9063
	Validation	6	0	6	2	0.8571
MIP	Whole	23	3	18	2	0.9111
	Training	15	1	14	2	0.9063
	Validation	6	4	2	2	0.5714
Two Phase	Whole	25	1	18	2	**0.9348**
	Training	15	1	14	2	**0.9063**
	Validation	7	0	6	1	**0.9286**

Table 3. Misclassification cost of the five approaches to the bankruptcy dat

Method / Sample	LR*	FLDF	MSD	MIP	Two-phase MP
Whole sample	17	16	22	17	7
Training sample	13	13	7	7	7
Validation sample	3	2	2	22	1

(* LR : Logistic regression)

In Table 2, the proposed two-phase mathematical programming approach shows higher hit-ratios than the other four approaches in the whole, training and validation samples. In Table 3, the two-phase mathematical programming approach shows costs less than the other approaches on the basis of cost matrix.

In case of bankruptcy firm data, sample size is small and all attributes are numerical. Although multivariate normality assumptions were accepted, the proposed two-phase mathematical programming approach outperformed the other statistical approaches in terms of classification accuracy and cost. It is also natural that the two-phase mathematical programming approach outperformed MSD model and MIP model, because it is designed to complement any defect of MSD model and MIP model

4.2. German credit data

The data set is obtained 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. 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

The hit ratios and misclassification costs of the five approaches to German credit data are listed in Table 4 and

Table 5, respectively.

Table 4. Hit ratio of the five approaches to the German Credit data

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

Table 5. Misclassification cost of the five approaches to the German credit data

Method / Sample	LR	FLDF	MSD	MIP	Two-phase MP
Training sample	160	107	205	173	**89**
Validation sample	185	119	210	233	**148**

In Table 4, the proposed two-phase mathematical programming approach shows higher hit ratio than the other approaches in the training sample. Moreover, the misclassification cost of the two-phase mathematical programming approach is less than that of any other approaches in the training sample in Table 5.

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

The overall conclusion is made based on the experimental results that the proposed two-phase mathematical programming approach may be a good alternative to other statistical approaches and be an improving approach of the existing mathematical programming approaches.

5. Conclusions

In this paper, a two-phase mathematical programming approach is proposed for solving the proposed credit scoring problem. This approach differs from the previous formulations in the literature such that it explicitly considers the classification gap and provides a means for classifying observations which fall within the gap. By using a linear programming (LP) to consider the classification gap, Phase 1 makes decision to grant credit, deny credit, or to seek additional information before making a decision. On the other hand, Phase 2 finds a cut-off value, which minimizes the misclassification cost of granting credit to 'bad' or denying credit to 'good' by using the technique of mixed-integer programming (MIP).

The proposed approach has been tested to see if it performs as well as other statistical approaches. In the empirical test carried out with bankruptcy firm data and German credit data, the proposed approach has shown that it outperforms the existing mathematical programming approaches and other statistical approaches in the literature.

Based on the test results, it is concluded that the proposed two-phase mathematical programming approach may be at least as good as statistical approaches and traditional mathematical programming approaches for credit scoring and discriminant problems

References

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.

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*, 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*, 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. (1990), Improved LP Models for Discriminant Analysis, *Decision Science*, Vol. 21, pp. 771-785.

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.

Hand, D.J. and Henley W.E. (1997), Statistical Classification Methods in Consumer Credit Scoring: a Review, *Journal of Royal Statistics Society A* Vol. 160, pp. 523-541.

Henley W.E. (1995), Statistical Aspects or Credit Scoring, PhD thesis, Open University.

Johnson, R.A and Wichern D.W (1988), Applied Multivariate Statistical Analysis, Prentice Hall, NJ.

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. (1990), Considerations for Mathematical Programming Models in Discriminant Analysis, *Managerial and Decision Economics*, Vol. 11, pp. 227-234.

Koehler, G.J. (1991), Improper linear discriminant classifiers, *European Journal of Operational Research*, Vol. 50, No. 2, pp. 188-198

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

Lam, K.F., Choo, E.U. and Moy, J.W. (1996), Minimizing Deviations from the Group Mean: A New LP Approach for the Two-Group Classification Problem, *European Journal of Operational Research*, Vol. 88, pp. 358-367.

Lewis, E.M. (1992), An Introduction to Credit Scoring, Athena Press: San Rafael, CA.

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.

Rosenberg, E. and Gleit, A. (1994), Quantitative Methods of in Credit Management: A Survey, *Operations Research*, Vol. 42, pp. 589-613.

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.

Thomas, L.C. (2000), A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumers, *International of Journal of Forecasting*, Vol. 16, pp. 149-172.

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