

Cross-Validation and Ensemble Analyses on Multiple-Criteria Linear Programming Classification for Credit Cardholder Behavior^{*}

Yi Peng¹, Gang Kou¹, Zhengxin Chen¹, and Yong Shi^{1,2**}

¹College of Information Science and Technology
University of Nebraska at Omaha
Omaha, NE 68182, USA

{ypeng, gkou, zchen, yshi}@mail.unomaha.edu

²Graduate School of Chinese Academy of Sciences, Beijing 100039, China

Abstract. In credit card portfolio management, predicting the cardholders' behavior is a key to reduce the charge off risk of credit card issuers. As a promising data mining approach, multiple criteria linear programming (MCLP) has been successfully applied to classify credit cardholders' behavior into two or multiple-groups for business intelligence. The objective of this paper is to study the stability of MCLP in classifying credit cardholders' behavior by using cross-validation and ensemble techniques. An overview of the two-group MCLP model formulation and a description of the dataset used in this paper are introduced first. Then cross-validation and ensemble methods are tested respectively. As the results demonstrated, the classification rates of cross-validation and ensemble methods are close to the rates of using MCLP alone. In other words, MCLP is a relatively stable method in classifying credit cardholders' behavior.

Keywords: Credit Card Portfolio Management, Data Mining, Classification, Multi-criteria Linear Programming, Cross-Validation, and Ensemble

1 Introduction

Mining useful information or discovering knowledge from large databases becomes a cutting-edge information technology tool in today's competitive business world. Data mining techniques help organizations and companies to discover previously unknown and actionable information from various and large databases for decision-making. Recently, Multiple-Criteria Linear Programming (MCLP) has been applied to credit cardholders' behavior classification for business decision [1, 2, 3, 4]. However, the stability of MCLP classification in credit card portfolio management remains unexplored. In order to respond this challenge, this paper conducts cross-validation and ensemble analysis using real-life credit card data from a large US bank. In this study, we intend to investigate whether the classification results generated by MCLP method can be improved. The k-fold cross-validation is first applied on the credit card data to acquire a number of global optimal solutions of MCLP method.

^{*} This research has been partially supported by a grant of US Air Force Research Laboratory (PR No. E-3-1162) and a grant from the K.C. Wong Education Foundation (2003), Chinese Academy of Sciences.

^{**} The corresponding Author

Then, part of these global optimal solutions is aggregated to form a single ensemble. The output ensemble is used to justify and improve the MCLP classification. Our findings indicate that the classification rates of cross-validation and ensemble analyses are close to that of MCLP alone. In other words, MCLP is a relatively stable method in classifying credit cardholders' behavior.

This paper is organized as follows. Next section is an overview of two-group MCLP model formulation. The third section describes the characteristics of the credit card dataset. The fourth section discusses the process and results of cross validation. The fifth section describes the procedure of ensemble analysis and examination of the resulting classification. The last section concludes the paper with some remarks.

2 Two-Group Multiple-Criteria Linear Programming Model

This section describes the two-group MCLP model briefly. For more details of MCLP and MCLP model formulation, please refer to [1, 2].

Often linear classification models use a linear combination of the minimization of the sum of overlapping (represented by α_i) and maximization of the sum of distance (represented by β_i) to reduce the two criteria problem into a single criterion. A two-criterion linear programming model is stated as:

(Model 1) Minimize $\sum_i \alpha_i$ and Maximize $\sum_i \beta_i$

Subject to:

$$A_i X = b + \alpha_i - \beta_i, A_i \in G1,$$

$$A_i X = b - \alpha_i + \beta_i, A_i \in G2,$$

Where A_i are given, X and b are unrestricted, and α_i and $\beta_i \geq 0$. The advantage of this conversion is to easily utilize all techniques of LP for separation, while the disadvantage is that it may miss the scenario of trade-offs between these two separation-criteria.

Applying the techniques of MCLP and the compromise solution, we want to minimize the sum of α_i and maximize the sum of β_i simultaneously. We assume the "ideal value" of $-\sum_i \alpha_i$ be $\alpha^* > 0$ and the "ideal value" of $\sum_i \beta_i$ be $\beta^* > 0$. Then, if $-\sum_i \alpha_i > \alpha^*$, we define the regret measure as $-d_\alpha^+ = \sum_i \alpha_i + \alpha^*$; otherwise, it is 0. If $-\sum_i \alpha_i < \alpha^*$, the regret measure is defined as $d_\alpha^- = \alpha^* + \sum_i \alpha_i$; otherwise, it is 0. Thus, we have (i) $\alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+$, (ii) $|\alpha^* + \sum_i \alpha_i| = d_\alpha^- + d_\alpha^+$, and (iii) $d_\alpha^-, d_\alpha^+ \geq 0$. Similarly, we derive $\beta^* - \sum_i \beta_i = d_\beta^- - d_\beta^+$, $|\beta^* - \sum_i \beta_i| = d_\beta^- + d_\beta^+$, and $d_\beta^-, d_\beta^+ \geq 0$ (see Figure 1). A two-group MCLP model has been gradually evolved as:

(Model 2) Minimize $d_\alpha^- + d_\alpha^+ + d_\beta^- + d_\beta^+$

Subject to:

$$\alpha^* + \sum_i \alpha_i = d_\alpha^- - d_\alpha^+,$$

$$\beta^* - \sum_i \beta_i = d_\beta^- - d_\beta^+,$$

$$A_i X = b + \alpha_i - \beta_i, A_i \in G1,$$

$$A_i X = b - \alpha_i + \beta_i, A_i \in G2,$$

where A_i , α^* , and β^* are given, X and b are unrestricted, and α_i , β_i , d_{α}^- , d_{α}^+ , d_{β}^- , $d_{\beta}^+ \geq 0$.

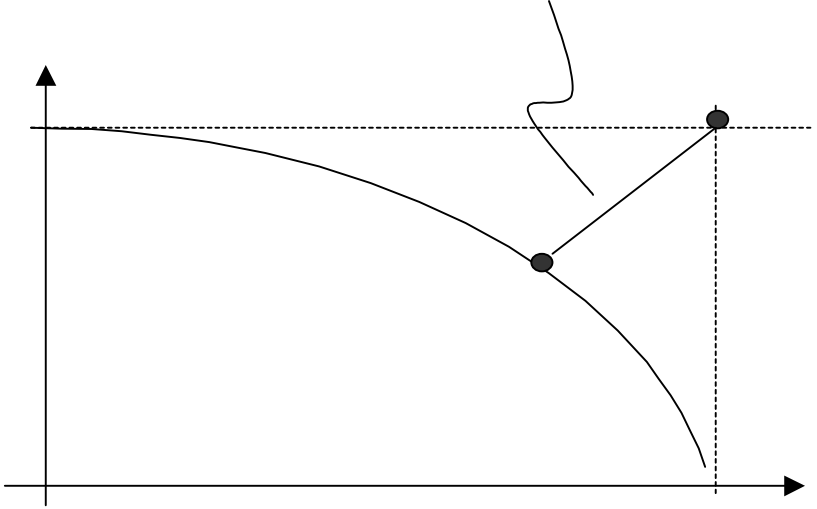


Fig. 1. Compromise Formulation

3 Credit Card Dataset Description

In order to understand the data mining process, it is important to comprehend the dataset. This section presents the nature and structure of the credit card data in details.

One of important data mining applications in banking industry is credit card bankruptcy analysis. Given a set of attributes, such as monthly payment, balance, purchases, and cash advance, the purpose is to find the optimized classifier through a training set and then use the classifier to predict future customers' spending behaviors [1, 3]. The common practice in credit card portfolio management is to separate credit cardholders' behaviors into two classes: bankruptcy or current. This is also known as the method of making "black list." Popular methods include Behavior Score, Credit Bureau Score, FDC Bankruptcy Score, and Set Enumeration Decision Tree Score [3]. These methods were developed by either statistics or decision tree.

The 5000 credit card records used in this paper were randomly selected from 25,000 real-life credit card records of a major US bank. Each record has 113 columns or variables (38 original variables and 65 derived variables) which are used to describe cardholders' behaviors. The 38 original variables can be divided into five categories: balance, purchase, payment, cash advance, and related variables. Balance, purchase,

payment, and cash advance categories each have six variables that represent raw data of six consecutive months. Other related variables include interest charges, date of last payment, times of cash advance, account open date, and so on. The 65 derived variables (CHAR01-CHAR65) are derived from original 38 variables using simple arithmetic methods to reinforce the comprehension of cardholders' behaviors, such as times of over limit in last two years, calculated interest rate, cash as percentage of balance, purchase as percentage to balance, payment as percentage to balance, purchase as percentage to payment, and so forth. These variables are not static; rather, they are evolving. New variables which are considered important can be added and variables which are proved to be trivia or irrelative in separating can be removed. Within the 5000 records, 815 accounts are bankrupted and 4185 are current.

After the discussion of MCLP model formulation and dataset structure, the basis of classification of credit cardholders' behaviors has established. The next section will introduce the concept of cross-validation and report the experimental steps and results.

4 Empirical Studies of Cross-Validation

Cross-validation is frequently used for estimating generalization error, model selection, experimental design evaluation, training exemplars selection, or pruning outliers [5]. By definition, cross-validation is the practice of partitioning a sample of data into sub samples such that analysis is initially performed on a single sub sample, while further sub samples are retained "blind" in order for subsequent use in confirming and validating the initial analysis [6]. The basic idea is to remove some of the data before training. After training is done, the data that was removed is used to test the performance of the model.

Three kinds of cross validation forms: holdout cross validation, k-fold cross validation, and leave-one-out cross validation are widely used. In this paper, k-fold cross validation is used. In the k-fold method, the data set is divided into k subsets and the holdout method is repeated k times. Each time, one of the k subsets is used for testing and the other k-1 subsets are used for training. The advantage is that all the examples in the dataset are eventually used for both training and testing. The error estimate is obtained as the average error rate on test examples. The variance of the resulting estimate is reduced as k is increased [7]. The disadvantage of this method is that it required high computation cost.

One of the important questions of cross validation is to decide the number of folds. With a large number of folds, the bias of the true error rate estimator will be small and the computational time will be large. On the other hand, when the number of folds is small, the bias of the estimator will be large and the computational time will be reduced. In practice, the choice of the number of folds depends on the size of the dataset. A common choice for k-fold cross-validation is around 10 [7]. Due to the structure of our dataset, k is decided to be 7. As stated in section 3, the total bankruptcy accounts are 815 and current accounts are 4185. Since bankruptcy class has smaller number of records, records of each class in the training dataset should be

calculated using bankruptcy data: $815 \times 6/7 = 699$. For easier computation, 700 is used instead of 699. Thus, the training dataset is formulated with 1400 records (700 bankruptcy and 700 current).

The procedure to select training datasets is described as follows: first, the bankruptcy dataset (815 records) is divided into 100 intervals (each interval has 8 records). Within each interval, 7 records are randomly selected. Thus the total of 700 bankruptcy records are obtained after this selection was repeated 100 times. Second, the current dataset (4185 records) is divided into 100 intervals (each interval has 41 records). Within each interval, 7 records are randomly selected. Thus the total of 700 current records are obtained after this selection was repeated 100 times. Third, the 700 bankruptcy and 700 current records are combined to form a single training dataset. Finally, the remaining 115 bankruptcy and 3485 current accounts become the testing dataset. According to this procedure, the total possible combination of this selection equals to $(C_8^7 \times C_{41}^7)^{100}$, which is infinitely large. That is, we can consider the possibility to get identical training or testing datasets is approximately zero. Considering the limited data availability in this study, we set the across-the-board threshold of 65% for absolute catch rate of the Bankruptcy class and 70% for the Current class to select the experimental results from training and test processes. This criterion is also applied to the ensemble analysis. Under the conditions stated above, the following steps are designed to carry out cross-validation:

Algorithm 1 (Cross Validation)

Step1 Generate the Training set (700 Bankruptcy data+700 Current Data) and Testing set (115 Bankruptcy data+3485 Current Data) from the credit card data set.

Step2 Apply the two-group MCLP model to compute the compromise solution $X^* = (x_1^*, x_2^*, \dots, x_{65}^*)$ as the best weights of all 65 variables with given values of control parameters (b, α^*, β^*) .

Step3 The classification score $MCLP_i = A_i X^*$ against of each observation has been calculated against the boundary b to check the performance measures of the classification.

Step4 If the classification result of Step 2 is acceptable (i.e., the given performance measure is larger or equal to the given threshold), go to Step 5. Otherwise, choose different values of control parameters (b, α^*, β^*) and go to Step 1.

Step5 Use $X^* = (x_1^*, x_2^*, \dots, x_{65}^*)$ to calculate the MCLP scores for all A_i in the test set and conduct the performance analysis. If it produces a satisfying classification result, go to the next step. Otherwise, go back to Step 1 to reformulate the Training Set and Testing Set.

Step6 Repeat the whole process until a preset number (e.g. 999) of different X^* are generated.

Some samples of the Cross-Validation tests based on Algorithm 1 are summarized in Tables 1.

The training and testing datasets have been computed using the above procedure. A part (20 out of the total 467 cross-validation results) of the results against the threshold is summarized. The worst and best classification catch rates for training set are 78.00% and 81.29% for Bankruptcy, 77.14% and 81.14% for the Current. The worst and best classification catch rates for testing set are 65.22% and 73.04% for Bankruptcy, 71.68% and 76.13% for the Current. As shown in table 1, the absolute catch rates of the Bankruptcy class are all above 65% and the absolute catch rates of the Current class are all above 70% for the selected experimental results. The result indicates that a good separation of Bankruptcy and Current is observed with this method.

Table 1. A Sample of the Cross-Validation Results

Cross Validation	Training Set (700 Bankruptcy data + 700 Current data)				Testing Set (115 Bankruptcy data + 3485 Current data)			
	Bank- ruptcy	Catch Rate	Cur- rent	Catch Rate	Bank- ruptcy	Catch Rate	Cur- rent	Catch Rate
DataSet 1	563	80.43%	557	79.57%	78	67.83%	2575	73.89%
DataSet 2	546	78.00%	546	78.00%	75	65.22%	2653	76.13%
DataSet 3	564	80.57%	560	80.00%	75	65.22%	2550	73.17%
DataSet 4	553	79.00%	553	79.00%	78	67.83%	2651	76.07%
DataSet 5	548	78.29%	540	77.14%	78	67.83%	2630	75.47%
DataSet 6	567	81.00%	561	80.14%	79	68.70%	2576	73.92%
DataSet 7	556	79.43%	548	78.29%	77	66.96%	2557	73.37%
DataSet 8	562	80.29%	552	78.86%	79	68.70%	2557	73.37%
DataSet 9	566	80.86%	557	79.57%	83	72.17%	2588	74.26%
DataSet 10	560	80.00%	554	79.14%	80	69.57%	2589	74.29%
DataSet 11	548	78.29%	540	77.14%	79	68.70%	2592	74.38%
DataSet 12	554	79.14%	546	78.00%	79	68.70%	2521	72.34%
DataSet 13	571	81.57%	568	81.14%	83	72.17%	2498	71.68%
DataSet 14	560	80.00%	552	78.86%	77	66.96%	2598	74.55%
DataSet 15	549	78.43%	535	76.43%	77	66.96%	2637	75.67%
DataSet 16	569	81.29%	563	80.43%	75	65.22%	2586	74.20%
DataSet 17	560	80.00%	555	79.29%	75	65.22%	2580	74.03%
DataSet 18	562	80.29%	557	79.57%	80	69.57%	2619	75.15%
DataSet 19	564	80.57%	560	80.00%	84	73.04%	2572	73.80%
DataSet 20	550	78.57%	550	78.57%	81	70.43%	2575	73.89%

5 An Ensemble Analysis

An ensemble consists of a set of individually trained classifiers whose predictions are combined when classifying novel instances. There are two fundamental elements of ensembles: a set of properly trained classifiers and an aggregation mechanism that organizes these classifiers into the output ensemble. Normally, the aggregation process will be an average or a simple majority vote over the output of the ensembles [8]. Previous research has shown that an ensemble can help to increase accuracy and stability [9, 10, 11].

There are two basic criterion based on which the ensemble is chosen: first, voters of an ensemble have to satisfy the across-the-board threshold of 65% for absolute catch rate of the Bankruptcy class and 70% for the Current class as mentioned previously. Second, as the majority vote method is used here, the number of voters in any ensemble must be odd.

From previous k-fold cross-validation, we have computed some optimal solutions of MCLP. Part of these optimal solutions is selected to form an ensemble. Each solution will have one vote for each credit card record and final classification result is determined by the majority votes. The reason for choosing the specific numbers of voters to form ensembles is that we have about 700 optimal solutions available from cross-validation studies. In order to utilize these results, number of voters is determined to be 9, 199, 299, and 399. The test did not go further because we observed that the catch rates remain stable when the number of voters equals to 399. Actually, the number of voters in an ensemble should be determined by the particular context.

The following steps describe the whole process:

Algorithm 2 (Ensemble Analysis)

Step1 A committee of certain odd number (e.g. 9, 99, 199) of classifiers X^* is formed.

Step2 The classification score $MCLP_i = A_i X^*$ against of each observation has been calculated against the boundary b by every member of the committee. The performance measures of the classification will be decided by majorities of the committee. If more than half of the committee members succeed in the classification, then the prediction for this observation is successful, otherwise, the prediction is failed.

Step3 The catch rate for each group will be computed by the percentage of successful classification in all observations.

Several results of Algorithm 2 for different ensemble committee size are summarized in Table 2. The results point out three findings: (1) the classification rates are close to cross-validation process; (2) the number of voters does not affect the classification results significantly; (3) although the catch rates of ensembles do not outperform the best results of cross-validation, they are more steady than cross-validation.

Table 2. Ensemble Results

No. of Voters	Bank-ruptcy	Catch Rate	Cur-rent	Catch Rate	Bank-ruptcy	Catch Rate	Cur-rent	Catch Rate
9	553	79.00%	544	77.71%	79	68.70%	2605	74.75%
99	552	78.86%	542	77.43%	78	67.83%	2595	74.46%
199	553	79.00%	544	77.71%	78	67.83%	2596	74.49%
299	552	78.86%	545	77.86%	78	67.83%	2597	74.52%
399	553	79.00%	544	77.71%	78	67.83%	2600	74.61%

6 Conclusions

Classification of credit cardholders' behavior is an important data mining application in banking industry. According to prior research, MCLP method exhibits promising results in credit cardholders' behavior classification. This paper explored the stability of MCLP method on credit card dataset by using cross-validation and ensemble analyses. The experimental results have shown that there is little effect on the MCLP classification with cross-validation and ensemble techniques. In other words, this indicates that MCLP is a reasonably stable classification method in this specific application. However, the comprehensive understanding on the general impact of using cross-validation and ensemble on MCLP performance needs to be further investigated. We shall report the related findings in the near future.

References

1. Shi, Y., Wise, M., Luo, M. and Lin, Y. (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*, Springer, Berlin, 427-436.
2. Shi, Y., Peng, Y., Xu, W. and Tang, X.: Data Mining via Multiple Criteria Linear Programming: Applications in Credit Card Portfolio Management, *International Journal of Information Technology and Decision Making*. 1 (2002) 131-151.
3. Peng, Y. (2002). Data Mining in Credit Card Portfolio Management: Classifications for Card Holder Behavior. Master Thesis, College of Information Science and Technology, University of Nebraska at Omaha.
4. Kou, G., X. Liu, Y. Peng, Y. Shi, M. Wise and W. Xu, "Multiple Criteria Linear Programming to Data Mining: Models, Algorithm Designs and Software Developments" *Optimization Methods and Software*, Vol. 18, 453-473, 2003.
5. Plutowski, M.E. (1996). "Survey: Cross-Validation in Theory and in Practice." Unpublished manuscript. Available online at: <http://www.emotivate.com/CvSurvey.doc>.
6. From Wikipedia, the free encyclopedia, available online: <http://en2.wikipedia.org/wiki/Cross-validation>.

7. Ricardo Gutierrez-Osuna, "Selected topics in computer science", Texas A&M University. Available online:
http://faculty.cs.tamu.edu/rgutier/courses/cs790_wi02/.
8. Gabriele Zenobi and Pádraig Cunningham, An Approach to Aggregating Ensembles of Lazy Learners That Supports Explanation, *Lecture Notes in Computer Science*, Vol. 2416, p. 436-447, 2002.
9. David Opitz and Richard Maclin, "Popular ensemble methods: an empirical study", *Journal of Artificial Intelligence Research* 11 (1999) 169-198.
10. Dietterich, T.G. (2000). Ensemble methods in machine learning. First international workshop on multiple classifier systems. pp.1-15. New York, 2000. Springer Verlag.
11. Jinseog Kim, Ensemble methods for data mining, Probability and data mining lab, Feb 4, 2002. Available online:
<http://srccs.snu.ac.kr/VerII/Activity/Tutorial/ensemble.pdf>