Science in China Series F: Information Sciences

www.scichina.com
info.scichina.com
www.springerlink.com

© 2008 SCIENCE IN CHINA PRESS
Springer

Regularized Multiple Criteria Linear Programs for Classification

Yong Shi[1], Yingjie Tian[1], Xiaojun Chen[2], and Peng Zhang[1,3]

[1] Research Center on Fictitious Economy & Data Science, Chin. Academy of Sci., Beijing 100190, China;
[2] Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong;
[3] School of Inform. Science and Eng., Graduate Univ. of Chin. Academy of Sciences, Beijing 100190, China

Although multiple criteria mathematical program (MCMP), as an alternative method of classification, has been used in various real-life data mining problems, its mathematical structure of solvability is still challengeable. This paper proposes a regularized multiple criteria linear program (RMCLP) for two classes classification problems . It first adds some regularization terms in the objective function of the known multiple criteria linear program (MCLP) model for possible existence of solution. Then the paper describes the mathematical framework of the solvability. Finally, a series of experimental tests are conducted to illustrate the performance of the proposed RMCLP with the existing methods: MCLP, multiple criteria quadratic program (MCQP), and support vector machine (SVM). The results of four publicly available datasets and a real-life credit dataset all show that RMCLP is a competitive method in classification. Furthermore, this paper explores an Ordinal RMCLP (ORMCLP) model for ordinal multi-group problems. Comparing ORMCLP with traditional methods such as One-Against-One, One-Against-The rest on large-scale credit card dataset, experimental results show that both ORMCLP and RMCLP perform well.

multiple criteria mathematical program, regularized multiple criteria mathematical program, classification, data mining.

1 Introduction

For the last decade, the researchers have extensively applied a quadratic program, known as Vapnik's Support Vector Machine (SVM) [1−8], into classification as well as various data analysis. However, using optimization techniques to deal with data separation and data analysis goes back to more than forty years ago [9−12]. According to Mangasarian [13], his group has formulated linear program as a large margin classifier in 1960's. In 1970's, Charnes and Cooper initiated Data Envelopment Analysis where a fractional programming is used to evaluate decision making units, which is economic representative data in a given training dataset [14]. From 1980's to 1990's, Glover proposed a number of linear programming models to solve discriminant problems with a small sample size of data [15,16]. Then, since 1998 Shi and his colleagues ex-

Received December 30, 2007; accepted July 29, 2008
doi:
†Corresponding author (email: yshi@gucas.ac.cn)
This research has been partially supported by a grant from National Natural Science Foundation of China (#70621001, #70531040, #70501030, #10601064, #70472074), National Natural Science Foundation of Beijing #9073020, 973 Project #2004CB720103, Ministry of Science and Technology, China, the Research Grants Council of Hong Kong and BHP Billiton Co., Australia.

tended such a research idea into classification via multiple criteria linear programming (MCLP) and multiple criteria quadratic programming (MQLP), which differ from statistics, decision tree induction, and neural networks [17−21]. These mathematical programming approaches to classification have been applied to handle many real-world data mining problems, such as credit card portfolio management [22−23], bioinformatics [24,25], fraud management [26], information intrusion and detection [27,28], firm bankruptcy [29], etc.

However, the structure of the MCLP models cannot ensure there is always a solution. To overcome this shortcoming, the objective of this paper is to propose regularized multiple criteria linear programs (RMCLP) with existence of solution for classification. Rest of the paper proceeds as follows. Section 2 introduces the basic notions and formulation of MCLP. Then section 3 describes the mathematical framework of the solvability. Section 4 uses a series of experimental tests to illustrate the performance of the proposed RMCLP with the existing methods: MCLP, MCQP, and SVM. The experimental results of four publicly available datasets and a real-life credit dataset all show that RMCLP is a competitive method in classification. Based on the above sections, Section 5 constructs an ordinal RMCLP model (ORMCLP) for multi-group classification problems, and the model also shows its efficiency through real-life credit dataset. Finally Section 6 gives the conclusions.

2 Regularized MCLP for Data Mining

Given an matrix $A \in R^{m\times n}$ and vectors $d, c \in R^m_+$, the multiple criteria linear programming (MCLP) has the following version

$$
\begin{aligned}
\min_{u,v} \quad & d^T u - c^T v, && (1)\\
\text{s.t.} \quad & a_i x - u_i + v_i = b, \quad i = 1,2,\ldots,l, \\
& a_i x + u_i - v_i = b, \quad i = l+1, l+2, \ldots, m, \\
& u, v \geqslant 0,
\end{aligned}
$$

where a_i is the ith row of A which contains all given data.

The MCLP model is a special linear program, and has been successfully used in data mining for a number of applications with large data sets [21,22,24−29]. However, we cannot ensure this model always has a solution. Obviously the feasible set of MCLP is nonempty, as the zero vector is a feasible point. For $c \geqslant 0$, the objective function may not have a lower bound on the feasible set. In this paper, to ensure the existence of solution, we add regularization terms in the objective function, and consider the following regularized MCLP

$$
\begin{aligned}
\min_{z} \quad & \frac{1}{2}x^T Hx + \frac{1}{2}u^T Qu + d^T u - c^T v, && (2)\\
\text{s.t.} \quad & a_i x - u_i + v_i = b, \quad i = 1,2,\ldots,l, \\
& a_i x + u_i - v_i = b, \quad i = l+1, l+2, \ldots, m, \\
& u, v \geqslant 0,
\end{aligned}
$$

where $z = (x, u, v, b) \in R^{n+m+m+1}$, $H \in R^{n\times n}$ and $Q \in R^{m\times m}$ are symmeteric positive definite matrices. The regularized MCLP is a convex quadratic program. Although the objective function

$$f(z) := \frac{1}{2}x^T Hx + \frac{1}{2}u^T Qu + d^T u - c^T v$$

is not a strictly convex function, we can show that (2) always has a solution. Moreover, the solution set of (2) is bounded if H, Q, d, c are chosen appropriately.

Let $I_1 \in R^{l\times l}, I_2 \in R^{(m-l)\times(m-l)}$ be identity matrices,

$$A_1 = \begin{pmatrix} a_1 \\ \vdots \\ a_l \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_{l+1} \\ \vdots \\ a_m \end{pmatrix},$$

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \quad E = \begin{pmatrix} -I_1 & \\ & I_2 \end{pmatrix},$$

and $e \in R^m$ be the vector whose all elements are 1. Let

$$B = \begin{pmatrix} A & E & -E & -e \end{pmatrix}.$$

The feasible set of (2) is given by

$$\mathcal{F} = \{z \mid Bz = 0, u \geqslant 0, v \geqslant 0\}.$$

Since (2) is a convex program with linear constraints, the known KKT condition is a necessary and sufficient condition for optimality. To show that $f(z)$ is bounded on $\mathcal{F}$, we will consider the the KKT system of (2).

3 Solution set of RMCLP

Without loss of generality, we assume that $l > 0$ and $m - l > 0$.

Theorem 1. The solution set of RMCLP (2) is nonempty.

Proof. We show that under the assumption that $l > 0, m - l > 0$, the objective function has a lower bound. Note that the first terms in the objective function are nonnegative. If there is sequence z^k in $\mathcal{F}$ such that $f(z^k) \to$

$-\infty$, then there is i such that $v_i^k \to \infty$, which, together with the constraints of (2), implies that there must be j such that $|x_j^k| \to \infty$ or $u_j^k \to \infty$. However, the objective function has quadratic terms in x and u which are larger than the linear terms when $k \to \infty$. This contradicts $f(z^k) \to -\infty$. Therefore, by the Frank-Wolfe Theorem, the regularized MCLP (2) always has a solution. We complete the proof.

Now we show that the solution set of problem (2) is bounded if parameters H, Q, d, c are chosen appropriately.

Theorem 2. Suppose that $AH^{-1}A^T$ is nonsingular. Let $G = (AH^{-1}A^T)^{-1}$, $\mu = 1/e^TGe$ and

$$M = \begin{pmatrix} Q + EGE - \mu EGee^TGE & \mu EGee^TGE - EGE \\ -EGE + \mu EGee^TGE & EGE - \mu EGee^TGE \end{pmatrix},$$

$$q = \begin{pmatrix} d \\ -c \end{pmatrix} \qquad y = \begin{pmatrix} u \\ v \end{pmatrix}.$$

Then problem (2) is equivalent to the linear complementarity problem

$$My + q \geqslant 0, \quad y \geqslant 0, \quad y^T(My + q) = 0. \tag{3}$$

If we choose Q and H such that M is a positive semidefinite matrix and c, d satisfy

$$d + 2Qe > (\mu EGee^TGE - EGE)e > c, \tag{4}$$

then problem (2) has a nonempty and bounded solution set [30].

Proof. Let us consider the KKT condition of (2)

$$\begin{aligned} &Hx + A^T\lambda = 0 \\ &-c - E\lambda - \beta = 0 \\ &Qu + E\lambda + d - \alpha = 0 \\ &Bz = 0 \\ &e^T\lambda = 0 \\ u \geqslant 0, \quad \alpha \geqslant 0, \quad &\alpha^Tu = 0 \\ v \geqslant 0, \quad \beta \geqslant 0, \quad &\beta^Tv = 0. \end{aligned}$$

From the first three equalities in the KKT condition, we have

$$x = -H^{-1}A^T\lambda$$

$$\beta = -c - E\lambda$$

$$\alpha = Qu + E\lambda + d.$$

Substituting x in the 4th equality in the KKT condition gives

$$\lambda = G(Eu - Ev - eb).$$

Furthermore. from the 5th equality in the KKT condition, we obtain

$$b = \mu e^TGE(u - v).$$

Therefore, β and α can be defined by u, v as

$$\begin{aligned} \beta &= -c - EG(Eu - Ev - eb) \\ &= -c - EG(Eu - Ev - \mu ee^TGE(u - v)) \end{aligned}$$

and

$$\begin{aligned} \alpha &= d + Qu + EG(Eu - Ev - eb) \\ &= d + Qu + EG(Eu - Ev - \mu ee^TGE(u - v)). \end{aligned}$$

This implies that the KKT condition can be written as the linear complementarity problem (3). Since problem (2) is a convex problem, it is equivalent to the linear complementarity problem (3).

Let $u = 2e$, $v = e$ and $y_0 = (2e, e)$. Then from (4), we have

$$\begin{aligned} &My_0 + q \\ = &\begin{pmatrix} 2Qe + EGEe - \mu EGee^TGHe + d \\ \mu EGee^TGEe - EGEe - c \end{pmatrix} > 0 \end{aligned}$$

which implies that y_0 is a trictly feasible point of (3). Therefore, when M is a positive semidefinite matrix, the solution set of (3) is nonempty and bounded [30].

Let $y^* = (u^*, v^*)$ be a solution of (3), then $z^* = (x^*, u^*, v^*, b^*)$ with

$$b^* = \mu e^TGE(u^* - v^*) \quad \text{and}$$

$$x^* = -HA^TG(Eu^* - Ev^* - \mu ee^TGE(u^* - v^*))$$

is a solution of (2). Moreover, from the KKT condition, it is easy to verify that the boundness of the solution set of (3) implies the boundness of the solution set of (2).

4 Numerical test

In this section, we will compare the performance of RMCLP with other methods: MCLP, MCQP, and SVM on four publicly available datasets from UCI Machine Learning Repository [31] and credit card dataset. Here we only use SVM with linear Kernel because the other three algorithms are linear classifiers.

For every dataset, we randomly separate it into two parts, one part is for training, and the other for testing, then apply the above four algorithms to train and test.

This process is performed ten times, every time the accuracy on training and testing are recorded, at last the average accuracy is computed and shown in Table 1-Table 4.

In every training, parameters in every algorithm are selected in some discrete set in order to get the best accuracy. For example, the parameters in RMCLP needed to be chosen are H, Q, d, c, so we choose H in a set of several special matrixes, and Q in another set of several given matrixes, d and c in the sets of several given vectors.

From Table 1 to Table 4 we can see that the performance of RMCLP is better than MCLP, MCQP, and almost the same with SVM in linear Kernel.

Table 1 Test On Australian Dataset

Classification	Training (200) + Testing (490)	
Algorithms	Training Accu.	Testing Accu.
MCLP	78.0%	75.5%
MCQP	89.0%	84.5%
RMCLP	91.0%	89.2%
SVM	91.2%	88.9%

Table 2 Test On German Dataset

Classification	Training (200) + Testing (800)	
Algorithms	Training Accu.	Testing Accu.
MCLP	72.0%	66.5%
MCQP	73.5%	71.5%
RMCLP	75.0%	72.5%
SVM	74.6%	73.1%

Table 3 Test On Heart Dataset

Classification	Training (100) + Testing (170)	
Algorithms	Training Accu.	Testing Accu.
MCLP	79.0%	77.5%
MCQP	88.0%	83.2%
RMCLP	87.0%	84.7%
SVM	89.5%	87.6%

Table 4 Test On Splice Dataset

Classification	Training (400) + Testing (600)	
Algorithms	Training Accu.	Testing Accu.
MCLP	84.3%	70.8%
MCQP	86.5%	74.7%
RMCLP	87.6%	76.2%
SVM	87.9%	76.1%

Now we test the performance of RMCLP on credit card dataset.

The 6000 credit card records used in this paper were selected from 25,000 real-life credit card records of a major US bank. Each record has 113 columns or variables to describe the cardholders' behaviors, including balance, purchases, payment cash advance and so on. With the accumulated experience functions, we eventually get 65 variables from the original 113 variables to describe the cardholders' behaviors.

Cross-validation is frequently used for estimating generalization error, model selection, experimental design evaluation, training exemplars selection, or pruning outliers [32]. There are three kinds of cross validation methods: holdout cross validation, k-fold cross validation, and leave-one-out cross validation that are widely used [33]. In this paper we chose the holdout method on credit card dataset.

The holdout method separates data into training set and testing set, taking no longer to compute. The process to select training and testing set is described as follows: first, the bankruptcy dataset (960 records) is divided into 10 intervals (each interval has approximately 100 records). Within each interval, 50 records are randomly selected. Thus the total of 500 bankruptcy records is obtained after repeating 10 times. Then, as the same way, we get 500 current records from the current dataset. Finally, the total of 500 bankruptcy records and 500 current records are combined to form a single training dataset, with the remaining 460 lost records and 4540 current records merge into a testing dataset. The following steps are designed to carry out cross-validation:

Algorithm 1.
Step 1: Generate the training set (500 bankruptcy records + 500 current records) and testing set (460 bankruptcy records + 4540 current records);
Step2: Apply the RMCLP model to compute as the best weights of all 65 variables with given values of control parameters (H, D, h, d, c);
Step3: The classification $score_i = a_i x$ has been calculated to check the performance of the classification;
Step4: If the classification result of Step 3 is unacceptable, choose different values of control parameters (H, D, h, d, c) and go back to Step 1;

We have computed 10 group dataset and the result is shown in Table 5. The columns "bankruptcy" and

”current” refer to the number of records that were correctly classified as ”bankruptcy” and ”current”, respectively. The column ”accuracy” was calculated using correctly classified records divided by the total records in that class. For instance, 87.20% accuracy of Dataset one for bankruptcy record in the training dataset was calculated using 436 divided by 500 and means that 87.20% of bankruptcy records were correctly classified.

It can be observed that for the training sample, the average accuracy of RMCLP on bankruptcy records is 86.74%, on current records is 70.08%. Out of the ten testing dataset result, the highest accuracy for the bankruptcy records is 87.83% and lowest is 83.04%, averaging to 85.44%. The highest and lowest testing accuracy's deviation reaches 2.39% and 2.40%; for current records testing accuracy reached the highest of 68.72% and the lowest of 67.22%, averaging to 67.97%, the highest and lowest prediction accuracy's deviation are both 0.75%. Through the cross-validation of ten groups, we can conclude that RMCLP model is not only accurate but also stable to classify the credit card dataset.

Table 5 Cross-validation on Credit Card Dataset

cross	Training Set (500 Lost + 500 Current)			
validation	Lost	Accuracy	Current	Accuracy
DS 1	436	87.20%	356	71.20%
DS 2	434	86.80%	352	70.40%
DS 3	438	87.60%	347	69.40%
DS 4	439	87.80%	348	69.60%
DS 5	428	85.60%	353	70.60%
DS 6	430	86.00%	361	72.20%
DS 7	437	87.40%	342	68.40%
DS 8	437	87.40%	350	70.00%
DS 9	426	85.20%	356	71.20%
DS 10	432	86.40%	339	67.80%
cross	Testing Set (460 Lost + 4540 Current)			
validation	Lost	Accuracy	Current	Accuracy
DS 1	399	86.74%	3057	67.33%
DS 2	389	84.57%	3120	68.72%
DS 3	390	84.78%	3089	68.04%
DS 4	396	86.09%	3059	67.38%
DS 5	382	83.04%	3085	67.95%
DS 6	404	87.83%	3102	68.33%
DS 7	396	86.09%	3074	67.71%
DS 8	397	86.30%	3057	67.33%
DS 9	390	84.78%	3036	66.87%
DS 10	396	86.09%	3052	67.22%

5 Ordinal Multi-group RMCLP Classification Models

In this section we will generalize a new version of RMCLP to tackle multi-group classification problem. So far, there have been two ways to deal with the multi-group problems. The first way is to construct a model which is capable to handle multi-group classification, such as the well-know Decision Tree model. The second method is the hierarchical methods, such as the One-Against-All strategy and the One-Against-One strategy.

Before giving our model, we first discuss the probability distribution of multi-group dataset . Since we can only get small training samples and we cannot know the whole distribution of the dataset before we do data mining (otherwise we need not do data mining), it is necessary to consider some hypotheses ($\mathcal{H}_1$ and $\mathcal{H}_2$) .

$\mathcal{H}_1$: The distribution of the dataset is in both linear and ordinal order, as depicted in Fig. 1.

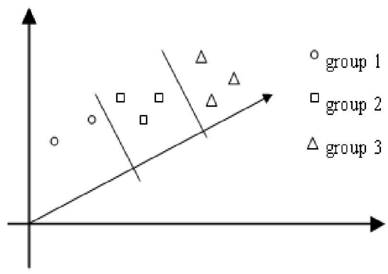

Fig. 1

$\mathcal{H}_2$: The distribution of the dataset is only in linear order, as depicted in Fig. 2.

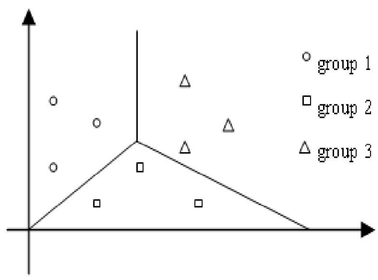

Fig. 2

5.1 Ordinal RMCLP

In case of $\mathcal{H}_1$, we can find a direction x on which all the records' projection is linear separable. As far as three-group classification problem is considered, we can find a direction x and a group of hyper planes (b_1, b_2), to any sample a_i , if $a_i x < b_1$, then a_i belongs to group 1, i.e. $a_i \in G_1$; if $b_1 \leqslant a_i x < b_2$, then $a_i \in G_2$; and if

$a_i x \geqslant b_2$, then $a_i \in G_3$. Extending this method to n group classification, we can also find a direction x and n-1 dimension vector $b = (b_1, b_2, \cdots, b_{n-1})^T \in R^{n-1}$, to make sure that to any sample a_i:

$$\begin{aligned} a_i x &< b_1, \quad && \forall a_i \in G_1, \\ b_{k-1} \leqslant a_i x &< b_k, \quad && \forall a_i \in G_k, 1 < k < n, \\ a_i x &\geqslant b_{n-1}, \quad && \forall a_i \in G_n. \end{aligned} \tag{5}$$

Now, we deduce the multi-group RMCLP classification model under condition $\mathcal{H}_1$. We first define $c_k = \frac{b_{k-1} + b_k}{2}$ as the midline in group k . Then, to the misclassified records, we define u_i^+ as the distance from c_k to $a_i x$, which equals $c_k - a_i x$, when misclassify a group $k's$ record into group $j(j < k)$, and we define u_i^- as the distance from $a_i x$ to c_k , which equals $a_i x - c_k$, when misclassify a group $k's$ record into group $j(j > k)$. Similarly, to the correct classified records, we define v_i^- when a_i is in the left side of c_k , and we define v_i^+ when a_i is in the right side of c_k. When we have an n groups training sample with size m, we have $u = (u_i^+, u_i^-) \in R^{2m}$, $v = (v_i^+, v_i^-) \in R^{2m}$, and we can build a Ordinal Regularized Multi-Criteria Linear Programming (ORMCLP) as follows:

$$\begin{aligned} \min_z \quad & \frac{1}{2} x^T H x + \frac{1}{2} u^T Q u + d^T u + c^T v, \qquad (6) \\ \text{s.t.} \quad & a_i x - u_i^- - v_i^- + v_i^+ = \frac{1}{2} b_1, \ \forall a_i \in G_1, \\ & a_i x - u_i^- + u_i^+ - v_i^- + v_i^+ = \frac{1}{2}(b_{k-1} + b_k), \\ & \qquad \forall a_i \in G_k, \ 1 < k < n, \\ & a_i x + u_i^+ - v_i^- + v_i^+ = 2 b_{n-1}, \ \forall a_i \in G_n, \\ & u_i^+, u_i^-, v_i^+, v_i^- \geqslant 0, \ i = 1, \cdots, m. \end{aligned}$$

To illustrate the proposed (6), we analyze its performance by a small synthetic dataset. As described in Table 6, we suppose there are three groups, G_1, G_2 and G_3. G_1 has two records, a_1 and a_2; G_2 has two records, a_3 and a_4; G_3 has a_5 and a_6, each record has two variables R_1 and R_2. We then suppose the separation hyper planes be $b_1 = 2$ and $b_2 = 4$, and the H and Q be the identity matrix. d and c be the vectors with all elements equal to 1.Then we can build the three-group classification problem as:

$$\begin{aligned} \min_z \quad & \frac{1}{2}\sum_i x_i^2 + \frac{1}{2}\sum_i (u_i^-)^2 + \frac{1}{2}\sum_i (u_i^+)^2 + \sum_i u_i^- \\ & + \sum_i u_i^+ + \sum_i v_i^- + \sum_i v_i^+, \\ & a_1 x_1 - u_1^- - v_1^- + v_1^+ = 1, \\ & a_2 x_2 - u_2^- - v_2^- + v_2^+ = 1, \\ & a_3 x_3 - u_3^- + u_1^+ - v_3^- + v_3^+ = 2, \\ & a_4 x_4 - u_4^- + u_2^+ - v_4^- + v_4^+ = 2, \\ & a_5 x_5 + u_3^+ - v_5^- + v_5^+ = 4, \\ & a_6 x_6 + u_4^- - v_6^- + v_6^+ = 6, \end{aligned} \tag{7}$$

We use the optimization package in Matlab 7.0 to solve this quadratic programming, and we can get $x = (1.0789, -0.3421)$ as the projection vector. Then we list each record's inner product with x as follows:

$$\begin{aligned} & G_1 : a_1 x = 0.736 < 2, \ a_2 x = 1.1315 < 2; \qquad (8) \\ & G_2 : a_3 x = 2.6051 \in [2, 4], \ a_4 x = 2.9998 \in [2, 4]; \\ & G_3 : a_5 x = 6.8681 > 4, \ a_6 x = 7.9996 > 4. \end{aligned}$$

From (8), we can see that a_1 and a_2, which belong to G_1, are in the left side of b_1, a_3 and a_4, which belong to G_2, are between the b_1 and b_2, a_5 and a_6, which belong to G_3, are in the right side of b_2. It means ORMCLP perfectly classifies this small synthetic dataset.

Table 6 Cross-validation on Credit Card Dataset

Three	Samples			
Groups	No.	R_1	R_2	$a_i x$
G_1	a_1	1	1	0.736
G_1	a_2	2	3	1.1315
G_2	a_3	4	5	2.6051
G_2	a_4	5	7	2.9998
G_3	a_5	7	2	6.8681
G_3	a_6	9	5	7.9996

5.2 Hierarchical Methods

Although ORMCLP performs well in case of $\mathcal{H}_1$, it would not work well under hypothesis $\mathcal{H}_2$. Then we can introduce another two traditional hierarchical methods: One-Against-Rest and One-Against-One for RMCLP[34,35]. With One-Against-The Rest strategy, we transform the k groups' classification problem into $k - 1$ two group's classification problems. Each time we extract one of the k groups as the first group, and combine the remained $k-1$ groups as the second group, then we can build model use two group RMCLP. In One-Against-One strategy, we build $k(k-1)/2$ models between each pairs, and then we use the winner tree to decide the final results. Since many papers [36] prove these two hierarchical methods' performance on multi-group classification, we do not need to test it on synthetic dataset as we did on ORMCLP.

5.3 Experiments on Credit Card Dataset

We now apply our new ORMCLP model to deal with real-life credit card dataset as used in Section 4. Suppose we define five classes for this dataset using a label variable: The Number of Over-limits[37]. The five classes are defined as Bankrupt charge-off accounts (THE NUMBER OF OVER-LIMITS$\geqslant$ 13), Non-bankrupt charge-off accounts (7 $\leqslant$THE NUMBER OF OVER-LIMITS$\leqslant$ 12), Delinquent accounts (3 $\leqslant$THE NUMBER OF OVER-LIMITS$\leqslant$ 6), Current accounts (1 $\leqslant$THE NUMBER OF OVER-LIMITS$\leqslant$ 2), and Outstanding accounts (no over limit). Bankrupt charge-off accounts are accounts that have been written off by credit card issuers due to reasons other than bankrupt claims. The charge-off policy may vary among authorized institutions. Delinquent accounts are accounts that have not paid the minimum balances for more than 90 days. Current accounts are accounts that have paid the minimum balances. The outstanding accounts are accounts that have not balances. In our randomly selected 6000 records, there are 72 Bankrupt charge-off accounts, 205 Non-bankrupt charge-off accounts, 454 Delinquent accounts, 575 Current accounts and 4694 outstanding accounts. These records will be used as the data source of our following experiments.

Table 7 Three Groups Training

3 Groups	ORMCLP		O-A-R		O-A-O	
(50+50+50)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	48	96.0%	47	94.0%	28	56.0%
Non-Bankrupt	44	88.0%	32	64.0%	50	100.0%
Delinquent	50	100.0%	48	96.0%	50	100.0%
	142	94.7%	127	84.7%	128	85.3%

Table 8 Three Groups Testing

3 Groups	ORMCLP		O-A-R		O-A-O	
(22+155+404)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	12	54.5%	16	72.7%	6	27.3%
Non-Bankrupt	12	7.7%	41	26.5%	55	35.5%
Delinquent	402	99.5%	313	77.5%	398	98.5%
	426	73.32%	370	63.68%	459	79.0%

Table 9 Four Groups Training

4 Groups	ORMCLP		O-A-R		O-A-O	
(50+50+50+50)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	50	100.0%	40	80.0%	49	98.0%
Non-Bankrupt	46	92.0%	26	52.0%	50	100.0%
Delinquent	47	94.0%	25	50.0%	50	100.0%
Current	50	100.0%	30	60.0%	50	100.0%
	193	96.5%	121	60.5%	199	99.5%

Table 10 Four Groups Testing

4 Groups	ORMCLP		O-A-R		O-A-O	
(22+155+404+525)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	16	72.7%	14	63.6%	14	63.6%
Non-Bankrupt	52	33.5%	29	18.7%	55	35.5%
Delinquent	38	9.4%	146	36.1%	270	66.8%
Current	525	100.0%	296	56.4%	515	98.1%
	631	57.05%	485	43.85%	854	77.22%

Table 11 Five Groups Training

5 Groups	ORMCLP		O-A-R		O-A-O	
(50+50+50+50+50)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	46	92.0%	28	56.0%	49	98.0%
Non-Bankrupt	49	98.0%	17	34.0%	50	100.0%
Delinquent	47	94.0%	19	38.0%	50	100.0%
Current	50	100.0%	28	56.0%	50	100.0%
Outstanding	50	100.0%	48	96.0%	50	100.0%
	242	96.8%	140	56.0%	249	99.6%

Table 12 Five Groups Testing

5 Groups	ORMCLP		O-A-R		O-A-O	
(22+155+404 +525+4644)	Rec.	Perc.	Rec.	Perc.	Rec.	Perc.
Bankrupt	13	59.1%	11	50.0%	14	63.6%
Non-Bankrupt	130	83.9%	21	13.5%	55	35.5%
Delinquent	273	67.6%	76	18.8%	270	66.8%
Current	161	30.7%	99	18.9%	515	98.1%
Outstanding	4644	100.0%	3226	69.5%	3964	85.4%
	5221	90.8%	3433	59.7%	4818	83.79%

In the following experiments, we test the ORMCLP, One-Against-Rest RMCLP and One-Against-One RMCLP on three groups, four groups and five groups credit card classification respectively. Table 7 and Table 8 are the training and testing results on three groups' classification. Table 9 and Table 10 are the training and testing results on four groups' classification. Table 11 and Table 12 are the training and testing results on five groups' classification. From the Table 7 to Table 12, the first column is the name of each group; the second and the third column are the correctly classified number and accuracy by ORMCLP, the fourth and the fifth column are the correctly classified number and accuracy by One-Against-All-Rest RMCLP model; the sixth and the seventh column are the correctly classified number and accuracy by One-Against-One RMCLP model; the last line is the totally correct classified records and the accuracy. For example, from the last line of Table 8, we know that ORMCLP has correctly classified 426 records, the accuracy is $426/581 = 73.32\%$. From these result tables, we can see that. For the three groups' classification, ORMCLP's training and testing accuracy are 94.7% and 73.32% respectively; One-Against-Rest RMCLP are 84.7% and 63.68% respectively; One-Against-One RMCLP are 85.3% and 79.00% respectively; to the four groups' classification, ORMCLP's training and testing accuracy are 96.5% and 57.05% respectively; One-Against-Rest RMCLP are 60.5% and 43.85% respectively; One-Against-One RMCLP are 99.5% and 77.22% respectively; For the five groups' classification, ORMCLP's training and testing accuracy are 96.8% and 90.80% respectively; One-Against-Rest RMCLP are 56.0% and 59.70% respectively; One-Against-One RMCLP are 99.6% and 83.79% respectively. That is to say, although One-Against-Rest RMCLP shows unstable and inaccurate on the testing dataset, ORMCLP and One-Against-One RMCLP are successfully separate each group. Moreover, One-Against-One RMCLP performs better than ORMCLP on three and four groups' classification, while ORMCLP is better than One-Against-One RMCLP on five groups' classification.

6 Conclusion

In this paper, a regularized multiple criteria linear program (RMCLP) has been proposed for classification problems in data mining. Comparing with the known multiple criteria linear program (MCLP) model, this model guarantees the existence of solution and is mathematically solvable. In addition to describing the mathematical structure, this paper has also conducted a series of experimental tests on comparison of MCLP, multiple criteria quadratic program (MCQP), and support vector machine (SVM) on several datasets. All results have shown that RMCLP is a competitive method in classification.

Furthermore, we have also proposed a new method to deal with ordinal multi-group classification problem: Ordinal RMCLP. Numerical test on real-life dataset has proved its efficiency.

There are some research problems still remaining to be explored. For example, is there similar solution structure for MCQP as for MCLP? What kinds of kernel functions can affect the solution of MCLP and MCQP? We shall continue working on these problems and report any significant results in the near future.

1 Vapnik V N. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer, 2000.
2 Vapnik V, Golowich S E, Smola A. Support vector method for function approximation, regression estimation, and signal processing. Advances in Neural Information Processing Systems. MIT Press, 1997, 281-287.
3 Osuna E, Freund R, Griosi F. An improved training algorithm for support vector machines. In: Neural Networks for Signal Processing, 1997, 276–285.
4 Burges C J, Scholkopf B. Improving the accuracy and speed of support vector machines. Advances in Neural Information Processing Systems. MIT Press, 1997, 375-381.
5 Zanni L, Serafini T, Zanghirati G. Parallel software for training large scale support vector machines on multiprocessor systems. Journal of Machine Learning Research, 2006, 7:1467-1492.
6 Platt J C. Fast training of support vector machines using sequential minimal optimization, Advances in kernel methods: support vector learning. Cambridge MA: MIT Press, 1999.
7 Collobert R, Svmtorch S B. Support vector machines for large-scale regression problems. Journal of Machine Learning Research, 2001, 1:143-160.
8 Ferris M, Munson T. Interior-point methods for massive support vector machines. SIAM Journal of Optimization, 2003, 3:783-804.
9 Bennett K P, Hernandez P E. The interplay of optimization and machine learning research. The Journal of Machine Learning Research, 2006, 7: 1265-1281
10 Mangasarian O L. Mathematical programming in data mining. Data Mining and Knowledge Discovery. 1997, 2(1): 183-201.
11 Paul Stephen Bradley, O.L., Mangasarian. Mathematical programming approaches to machine learning and data mining. The University of Wisconsin-Madison, 1998
12 Bradley P S, Fayyad U M, Mangasarian O L. Mathematical programming for data mining: formulations and challenges. INFORMS Journal on Computing, 1999, 11: 217-238.
13 Mangasarian O L. Generalized support vector machines. Advances in Large Margin Classifiers. Cambridge, MA: MIT Press, 2000.
14 Charnes A, Cooper W W. Management Models and Industrial Applications of Linear Programming, New York: Wiley, 1961.
15 Freed N, Glover F. Simple but powerful goal programming models for discriminant problems. Simple but Powerful Goal Programming Models for Discriminant Problems, 1981, 7: 44-60.

16 Freed N, Glover F. Evaluating alternative linear programming models to solve the two-group discriminant problem. Decision Science, 1986, 17: 151-162.

17 Olson D, Shi Y. Introduction to Business Data Mining. McGraw-Hill/Irwin, 2007.

18 Shi Y. Multiple criteria and multiple constraint levels linear programming: concepts, techniques and applications. New Jersey: World Scientific Pub Co Inc, 2001.

19 He J, Liu X, Shi Y, et al.. Classifications of credit card holder behavior by using fuzzy linear programming. International Journal of Information Technology and Decision Making, 2004, 3: 633-650.

20 Kou G, Liu X, Peng Y, et al. Multiple criteria linear programming approach to data mining: models, algorithm designs and software Development. Optimization Methods and Software, 2003, 18: 453-473.

21 Shi Y, Peng Y, Kou G, et al. Classifying credit card accounts for business intelligence and decision making: a multiple-criteria quadratic programming approach. International Journal of Information Technology and Decision Making, 2005, 4: 581-600.

22 Shi Y, Wise W, Lou M, et al. Multiple Criteria Decision Making in Credit Card Portfolio Management. Multiple Criteria Decision Making in New Millennium. 427-436, 2001.

23 Shi Y, Y Peng, Xu W, et al. Data mining via multiple criteria linear programming: applications in credit card portfolio management. International Journal of Information Technology and Decision Making, 2002, 1: 131-151.

24 Zhang J, Zhuang W, Yan N, et al. Classification of HIV-1 Mediated Neuronal Dendritic and Synaptic Damage Using Multiple Criteria Linear Programming. Neuroinformatics, 2004, 2: 303-326.

25 Shi Y, Zhang X, Wan J, et al. Prediction the distance range between antibody interface residues and antigen surface using Multiple Criteria Quadratic Programming. International Journal of Computer Mathematics, 2004, 84:690-707.

26 Peng Y, Kou G, Sabatka A, et al. Application of classification methods to individual disability income insurance fraud detection. In: ICCS 2007, Lecture Notes in Computer Science, 2007, 852-858.

27 Kou G, Peng Y, Chen Z, et al. A multiple-criteria quadratic programming approach to network intrusion detection. Chinese Academy of Sciences Symposium on Data Mining and Knowledge Management. 2004, 7: 12-14.

28 Kou G, Peng Y, Yan N, et al. Network intrusion detection by using multiple-criteria linear programming. In: International Conference on Service Systems and Service Management, 2004, 7: 19-21.

29 Kwak W, Shi Y, Eldridge S, et al. Bankruptcy prediction for Japanese firms: using multiple criteria linear programming data mining approach. International Journal of Data Mining and Business Intelligence, 2006.

30 Cottle R.W., Pang J S, Stone R E. The Linear Complementarity Problem. New York: Academic Press, 1992.

31 Murphy P M, Aha D W, UCI Repository of Machine Learning Databases. Available online at: www.ics.uci.edu_mlearnMLRepository.html, 1992.

32 Plutowski M E. Survey: Cross-Validation in Theory and in Practice. Available online at: http:www.emotivate.comCvSurvey.doc, 1996.

33 Peng Y, Kou G, Chen Z, et al. Cross-validation and Ensemble Analyses on Multiple-Criteria Linear Programming Classification for Credit Cardholder Behavior. In: ICCS 2004, Lecture Notes in Computer Science, 2004, 931-939.

34 Weston J, Watkins c. Multi-class Support Vector Machines. Technical Report CSD-TR-98-04, Royal Holloway, University of London. 1998.

35 Pontil M, Verri A. Support vector machines for 3-d object recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, 20:637-646.

36 Hsu C W, Lin C J. A comparison of methods for multi-class support vector machines. IEEE Transactions on Neural Networks, 2002, 13(2), 415-425.

37 Peng Y, Kou G, Shi Y, et al. Multiclass creditcardholers Behaviors classification methods. In ICCS 2006, Part IV, LNCS 3994, 485-492.