Misclassification Minimization Based on Multiple Criteria Linear Programming

Bo Wang*†, Yong Shi*†, Wayne Wei Huang ‡ and Guanfeng Liu§
*Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences
Beijing, China 100190
†The Key Research Lab on Big Data Mining and Knowledge Management, Chinese Academy of Sciences
Beijing, China 100190
Email: wangbo@ucas.ac.cn, yshi@ucas.ac.cn
‡School of Management, Xi'an Jiaotong University
Xi'an, China 710049
Email: whuang@mail.xjtu.edu.cn
§School of Computer Science and Technology, Soochow University
Suzhou, China 215006

***Abstract*—Misclassification minimization is an important and interesting topic in classification problem. Obviously, exploring the solution for this topic will benefit to many real life problems, such as credit card clients classification. This paper focuses on misclassification minimization based on multiple criteria linear programming (MCLP), proposing two different schemes to minimize the number of misclassified points in original MCLP. Especially, the complementarity is used to construct the first scheme and linear approximation technique is applied to solve it. Furthermore, successive linearization algorithm (SLA) is employed to achieve minimization the second scheme. Finally, numerical experiment tests the effect of this idea.**

Keywords*—*Misclassification minimization, multiple criteria linear programming, linear approximation, successive linearization algorithm.

I. Introduction

Recently, optimization based data mining methods has been greatly developed for various kinds of real life applications. Especially in many financial problems, such as credit card clients classification problem, optimization based methods show a distinct superiority in result explanation and suspicious client detection [1], [2]. There are many optimization based classification methods, including support vector machine (SVM). Especially, knowledge based SVM has obtained a good result in complicated classification problems [3], [4]. Among various optimization based methods, multiple criteria linear programming (MCLP), which is firstly posed in [5], is widely used in many fields, for example, credit scoring [9]. Besides, in [5], multiple criteria and multiple constraint levels linear programming (MC^2LP) has been also posed, which can deal with multiple resource levels. By using this model, [7] developed a MC^2LP based classification algorithm for credit scoring. However, this kind of algorithm was domain knowledge needed. Particularly, an interval corresponding to multiple constraint levels should be determined in advance. In addition, as it was discussed in [2], the interval was not full utilized, which led to poor performance. At the same time, [2] proposed a new MC^2LP based model. Taking advantage of multiple constraint levels, this new model could control and correct two types of classification error to a certain extent.

In a typical card clients classification problem, how to control the misclassified cases for both classes is a key issue. After defining "Bad" client according to his/her bankrupt record or credit card fraud history, on one hand, the credit card company wants to avoid to issue credit card to the latent "Bad" ones, which will definitely cause huge loss. On the other hand, in order to expand the business, the credit card company is always prone to issue as many credit card as possible [2]. As a result, risk control is needed. The MC^2LP based classification model introduced in [2] was a brand new model that could deal with two types of error differently and flexibly, according to actual demand. In other words, it offered a systemical way to control two types of error, i.e. to reduce a certain type of error as well as keeping the other type of error above a certain level. This statistical idea realized a reasonable tradeoff between two types of error.

Nevertheless, based on the consideration of computation, it is easy to find out that MC^2LP based method has a higher computation complexity than MCLP based method, about twice in complexity. Besides, the determination of constraint interval needs special strategy, which result in inconvenient of related algorithm implementation. Additionally, no previous work focuses on error control in MCLP based method. This paper attempts to emerge an MCLP based error controlling method in classification. In this way, computation complexity and classification errors can be reduced simultaneously.

Particularly, this paper will give two schemes on this topic and implement related error reduction algorithm. The first scheme takes advantage of the complementarity of original MCLP under some certain tradeoff choosing between two objective function [2]. In this way, a nonlinear objective function will be attained. A linear approximation method will be applied to solve this problem. The second scheme will add the number of misclassification points into objective directly. Then, based on some smoothing technique, we will show the equality between different models. At last, a local linearization algorithm will be employed to obtain the solution in finite steps.

The remainder of this paper is organized as follows. In sec-

978-1-4799-4274-9/14 $31.00 © 2014 IEEE
DOI 10.1109/ICDMW.2014.10

tion II, background of the problem will be stated thoroughly, including MCLP, complementarity lemma and misclassification minimization problem. Then, two schemes of improved MCLP based misclassification minimization models will be introduced and discussed in section III and IV. In order to show the correctness of promoted models, we will give the data experiment result in section V. Finally, in section VI, we will conclude the result and refer future work.

II. Problem Description and Fundamental Lemma

In this section, a typical description for classification problem is given. Besides, we will offer a fundamental lemma of original MCLP model, which is also the foundation of the following revised MCLP based classification method. Also, the concept of misclassification minimization problem will be introduced and the corresponding optimization problem will be proposed and discussed.

A. Problem Description

In a binary classification problem, let's assume the data set is $\boldsymbol{A}$, which is an $l \times n$ matrix. In other words, there are l records, n attributes for every single one. At the same time, every element of $l \times 1$ vector $\boldsymbol{y}$ indicates the class (+1 or -1), which the corresponding record belongs to.

Next, a matrix representation of original MCLP classification model [8], [9] is shown as follows [2]:

$$\begin{aligned} \min_{\boldsymbol{X},\boldsymbol{\alpha},\boldsymbol{\beta}} \quad & (\lambda_1, \lambda_2) \begin{pmatrix} \boldsymbol{e}^T & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{e}^T \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} \\ s.t. \quad & \begin{pmatrix} \boldsymbol{A} & \boldsymbol{Y} & -\boldsymbol{Y} \end{pmatrix} \begin{pmatrix} \boldsymbol{X} \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = b\boldsymbol{e}, \\ & (\boldsymbol{\alpha}, \boldsymbol{\beta})^T \geq \boldsymbol{0}. \end{aligned} \tag{1}$$

Here, $\boldsymbol{e}$ is a column vector with 1 as all the elements, $\boldsymbol{X}$ is the weight vector of the classification hyperplane and is unrestricted. b is unrestraint as a cutoff. The $l \times l$ diagonal matrix $\boldsymbol{Y} = \{y_{ii}\}$ is defined by $y_{ii} = y_i, y_i \in \boldsymbol{y}$. Then, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ represent internal and external deviations respectively.

Solving the above programming problem, we obtain the normal vector $\boldsymbol{X}$ of decision hyperplane. Then, the decision hyperplane can be expressed as $\boldsymbol{a}^T\boldsymbol{X} = 1$.

B. Fundamental Lemma

Using the matrix representation above, the complementarity lemma of original MCLP can be stated as follows [1]:

Lemma II.1. *Let's assume that $\lambda_1 > \lambda_2$ and $(\boldsymbol{X}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ is a solution of the original MCLP model above. Then, ${\boldsymbol{\alpha}^*}^T\boldsymbol{\beta}^* = 0$ holds.*

This lemma guarantees the correctness of the intuitive relationship between internal and external deviations. That is to say, when the tradeoff vector between multiple criteria satisfies certain condition, i.e. $\lambda_1 > \lambda_2$, complementarity of two different kinds of deviations holds.

Besides, in a view of space analytic geometry, ${\boldsymbol{\alpha}^*}^T\boldsymbol{\beta}^* = 0$ means the orthogonality between external deviation vector and internal deviation vector.

C. Misclassification Minimization

In this part, we simplify the scalar product between vector $\boldsymbol{x}$ and vector $\boldsymbol{y}$ as $\boldsymbol{xy}$. Here, we introduce the concept of misclassification minimization problem in optimization based classification. Firstly, [13] gave an universal definition of misclassification minimization problem, which can be expressed below.

Definition II.1 (Misclassification Minimization Problem)**.** Given two finite point sets $\boldsymbol{A}$ and $\boldsymbol{B}$ in the n-dimensional real space $\mathbb{R}^n$, construct a plane that minimizes the number of points of $\boldsymbol{A}$ falling in one of the closed halfspaces determined by the plane and the number of points of $\boldsymbol{B}$ falling in the other closed halfspace.

Remark II.1. This definition is equal to maximize the number of points of $\boldsymbol{A}$ and $\boldsymbol{B}$ in different closed halfspace determined by the plane, which is to guarantee the maximum number of well-classified points.

Then, in order to measure the number of misclassification points, we introduce a mapping as follows:

$$(\cdot)_* : \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_l)^T \rightarrow \boldsymbol{\alpha_*} = (\alpha_{*1}, \alpha_{*2}, \cdots, \alpha_{*l})^T, \tag{2}$$

where

$$\alpha_{*i} = \begin{cases} 1, & \text{when } \alpha_i > 0; \\ 0, & \text{else.} \end{cases} \tag{3}$$

Case 1 (Linear Separable)**.** *Suppose that $\boldsymbol{wx} = \gamma$ is the plane obtained by some classification algorithm, and $\boldsymbol{A}$, $\boldsymbol{B}$ are two data matrices representing two groups of data points. As a result, misclassification minimization problem can be presented as follows:*

$$\boldsymbol{e}(-\boldsymbol{Aw} + \boldsymbol{e}\gamma + \boldsymbol{e})_* + \boldsymbol{e}(\boldsymbol{Bw} - \boldsymbol{e}\gamma - \boldsymbol{e})_* = 0. \tag{4}$$

Case 2 (Linear Inseparable)**.** *In this case, the goal function of misclassification minimization problem can be rewritten as follows:*

$$\min_{\boldsymbol{w},\gamma} \boldsymbol{e}(-\boldsymbol{Aw} + \boldsymbol{e}\gamma + \boldsymbol{e})_* + \boldsymbol{e}(\boldsymbol{Bw} - \boldsymbol{e}\gamma - \boldsymbol{e})_*. \tag{5}$$

For universality, we can state this kind of problem in (6).

$$\begin{aligned} \min_{\boldsymbol{w},\gamma,\boldsymbol{y},\boldsymbol{z}} \quad & \boldsymbol{ey_*} + \boldsymbol{ez_*} \\ s.t. \quad & \boldsymbol{y} \geqq -\boldsymbol{Aw} + \boldsymbol{e}\gamma + \boldsymbol{e}, \boldsymbol{y} \geqq 0, \\ & \boldsymbol{z} \geqq \boldsymbol{Bw} - \boldsymbol{e}\gamma - \boldsymbol{e}, \boldsymbol{z} \geqq 0. \end{aligned} \tag{6}$$

Furthermore, [13] proved that the solution of (6) achieved the optimization of (5). Also, it guaranteed the existence of the solution of (6).

III. The First Scheme

In the previous work on original MCLP based classification models, the number of the misclassified points were not minimized directly. A similar measure, the sum of external deviations, were minimized, instead of focusing on the accuracy. Based on this situation, objective function, which represents the accuracy, can be add into the model, in order to satisfy the demand of error control.

Let's assume that there is no point locating on the classification hyperplane exactly, which leads to the definition of the vector below:

$$\boldsymbol{\theta} = \frac{1}{\boldsymbol{\alpha} + e^{-\boldsymbol{\beta}}} = (\frac{1}{\alpha_1 + e^{-\beta_1}} + \frac{1}{\alpha_2 + e^{-\beta_2}} + \cdots + \frac{1}{\alpha_l + e^{-\beta_l}})^T. \tag{7}$$

Thus, when the deviations meet the complementarity property, the inner product satisfies $\boldsymbol{\theta}^T\boldsymbol{\alpha} = N - \sum \frac{1}{\alpha_i+1}$. Here, N represents the number of misclassified points, which is exactly what should be reduced. Therefore, a classification model depending on MCLP can be resulted as follows:

$$\begin{aligned} &\min \quad \boldsymbol{e}^T\boldsymbol{\alpha} \\ &\min \quad \boldsymbol{\theta}^T\boldsymbol{\alpha} \\ &\max \quad \boldsymbol{e}^T\boldsymbol{\beta} \\ s.t. \quad &\boldsymbol{A_i X} = b + \alpha_i - \beta_i, \quad \boldsymbol{A_i} \in Bad, \\ &\boldsymbol{A_i X} = b - \alpha_i + \beta_i, \quad \boldsymbol{A_i} \in Good, \\ &\alpha_i, \beta_i \geq 0, i = 1, 2, \cdots, l. \end{aligned} \tag{8}$$

Here, $\boldsymbol{X}$ is the weight vector of the classification hyperplane and is unrestricted. Moreover, b is given in advance. Note that because there is one more criterion in the objective functions, comparing with the original MCLP model, Lemma II.1 cannot be applied to show the complementarity of this model directly. As a result, a stronger lemma is employed to discuss this problem [9].

Lemma III.1. *Let $\boldsymbol{y^+}$ and $\boldsymbol{y^-}$ be vectors of $\mathbb{R}^p$. Then consider the following problem:*

$$\begin{aligned} (GP') \qquad &min \;\; g(\boldsymbol{y^+}, \boldsymbol{y^-}) \\ s.t. \quad &f(\boldsymbol{x}) + \boldsymbol{y^+} - \boldsymbol{y^-} = \boldsymbol{\overline{f}}, \\ &\boldsymbol{y^+}, \boldsymbol{y^-} \geq \boldsymbol{0}, \\ &\boldsymbol{x} \in X. \end{aligned} \tag{9}$$

Suppose that the function g is monotonically increasing with respect to elements of $\boldsymbol{y^+}$ and $\boldsymbol{y^-}$ and strictly monotonically increasing with respect to at least either y_i^+ or y_i^- for each i $(1 \leq i \leq p)$. Then, the solution $\hat{\boldsymbol{y^+}}$ and $\hat{\boldsymbol{y^-}}$ to the preceding problem satisfy

$$\hat{y^+}_i \hat{y^-}_i = 0, i = 1, 2, \cdots, p.$$

According to Lemma III.1, let's assume the tradeoff vector of the multiple criteria is $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$. Then, we can conclude the result in the following theorem [2].

Theorem III.1. ***(The Complementarity Theorem)*** *Let's assume that $\lambda_1 > \lambda_3$ and $(\boldsymbol{X^*}, \boldsymbol{\alpha^*}, \boldsymbol{\beta^*})$ is a solution of model (8). Then, $\boldsymbol{\alpha^*}^T \cdot \boldsymbol{\beta^*} = 0$.*

Remark III.1. When the tradeoff vector satisfies $\lambda_1 > \lambda_3$, the objective function

$$g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \lambda_1 \boldsymbol{e}^T\boldsymbol{\alpha} - \lambda_3 \boldsymbol{e}^T\boldsymbol{\beta} + \lambda_2 \boldsymbol{\theta}^T\boldsymbol{\alpha}$$

is a strictly monotonically increasing function with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. On the basis of Lemma III.1, the complementarity property holds.

In this case, (8) can be transformed into the following system by keeping the compromise parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ satisfying $\lambda_1 > \lambda_3$.

$$\begin{aligned} g(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \lambda_1 &\sum_{i=1}^{l} \alpha_i - \lambda_3 \sum_{i=1}^{l} \beta_i + \lambda_2 \sum_{i=1}^{l} \frac{\alpha_i}{\alpha_i+1} \\ s.t. \quad &\boldsymbol{A_i X} = b + \alpha_i - \beta_i, \quad \boldsymbol{A_i} \in Bad, \\ &\boldsymbol{A_i X} = b - \alpha_i + \beta_i, \quad \boldsymbol{A_i} \in Good, \\ &\alpha_i, \beta_i \geq 0, i = 1, 2, \cdots, l. \end{aligned} \tag{10}$$

However, there is only one challenge in solving this problem, which is the second objective function in (8) is nonlinear. Linear approximation method can be used to overcome this problem [10], [11]. Here, we introduce a typical method to handle this kind of nonlinear programming problem.

Firstly, for a real life classification problem, by using original MCLP model, the external deviation vector $\boldsymbol{\alpha}$ can be obtained. Consequently, $\boldsymbol{\alpha}$ offers an estimate for the "grey zone", where misclassified points locate. Secondly, in (10), the continuity of $g(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is obvious. According to the analysis above, we can assume that the external deviation of each point in model (10) is no more than twice that of original MCLP model. In this case, let $\boldsymbol{\alpha}$ be the external deviation vector corresponding to original MCLP model, then $\bar{\boldsymbol{\alpha}}$ is the one of model (10). we have:

$$0 \leq \bar{\alpha_i} \leq 2\alpha_i, 0 \leq i \leq l.$$

Next, divide the interval $[0, 2\alpha_i]$ into 10 equal lengths, denoted by $\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{i10}$.

Now, we can transform model (10) into the following model.

$$\begin{aligned} \min_{\boldsymbol{X}, \boldsymbol{\beta}, \boldsymbol{h}} h(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \lambda_1 \sum_{i=1}^{l} \sum_{j=1}^{10} h_{ij}\alpha_{ij} - \lambda_3 \sum_{i=1}^{l} \beta_i \\ &+ \lambda_2 \sum_{i=1}^{l} \sum_{j=1}^{10} h_{ij} \frac{\alpha_{ij}}{\alpha_{ij}+1} \\ s.t. \quad &\boldsymbol{A_i X} = b + \sum_{j=1}^{10} h_{ij}\alpha_{ij} - \beta_i, \quad \boldsymbol{A_i} \in Bad, \\ &\boldsymbol{A_i X} = b - \sum_{j=1}^{10} h_{ij}\alpha_{ij} + \beta_i, \quad \boldsymbol{A_i} \in Good, \\ &\sum_{j=1}^{10} h_{ij} = 1, i = 1, 2, \cdots, l. \\ &\beta_i \geq 0, i = 1, 2, \cdots, l. \end{aligned} \tag{11}$$

In this way, we successfully transform nonlinear programming problem (10) into linear programming problem (11), which can be solved by simplex method.

IV. The Second Scheme

The second criterion of the first scheme is a non-linear function of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In addition, it can estimate the number of misclassification points only when $\lambda_1 > \lambda_3$. Thus, an MCLP based classification model with fewer beforehand conditions needs to be built. Consequently, $\boldsymbol{e}^T\boldsymbol{\alpha_*}$ is equal to the number of misclassified points. Then, we add it into the objective functions. Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ be the compromising parameter.

The former MCLP model can be developed into the following model:

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{X},b} \quad & \lambda_1 \boldsymbol{e}^T\boldsymbol{\alpha} - \lambda_2 \boldsymbol{e}^T\boldsymbol{\beta} + \lambda_3 \boldsymbol{e}^T\boldsymbol{\alpha_*} \\ s.t. \quad & \boldsymbol{A_i X} = b + \alpha_i - \beta_i, \quad \boldsymbol{A_i} \in Bad, \\ & \boldsymbol{A_i X} = b - \alpha_i + \beta_i, \quad \boldsymbol{A_i} \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, ..., l.\end{aligned} \tag{12}$$

Here, $|\boldsymbol{\alpha}| = (|\alpha_1|, |\alpha_2|, \cdots, |\alpha_l|)^T$.

A. Smoothing Technique

Now, let's discuss the smoothing of the above mapping. [12] researched this issue. Their result to the second scheme can be applied as follows. First, a parameter ε is introduced. It is clear that when ε is a big enough positive number, the following equation holds:

$$\alpha_{*i} \approx t(\alpha_i, \varepsilon) = 1 - exp(-\varepsilon\alpha_i), i = 1, 2, \cdots, l, \forall \alpha_i \in [0, \infty). \tag{13}$$

As a result, by fixing the compromising parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$, the model above can be converted to an error correcting classification model as follows:

$$\begin{aligned}\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{X},b} \quad & \lambda_1 \boldsymbol{e}^T\boldsymbol{\alpha} - \lambda_2 \boldsymbol{e}^T\boldsymbol{\beta} + \lambda_3 \boldsymbol{e}^T(\boldsymbol{e} - exp(-\varepsilon\boldsymbol{\alpha})) \\ s.t. \quad & \boldsymbol{A_i X} = b + \alpha_i - \beta_i, \quad \boldsymbol{A_i} \in Bad, \\ & \boldsymbol{A_i X} = b - \alpha_i + \beta_i, \quad \boldsymbol{A_i} \in Good, \\ & \alpha_i, \beta_i \geqslant 0, i = 1, 2, \cdots, l.\end{aligned} \tag{14}$$

Firstly, the existence of solution of concave problem (14) can be guaranteed by Corollary 32.3.3 in [16]. Secondly, the following theorem shows that: for some certain sequence ε_i, nondecreasing and converging to positive infinity, there is at least one vertex, which solves (12) and the sequencing models of (14) simultaneously [12].

Theorem IV.1. *Let $f : \mathbb{R}^k \to \mathbb{R}$ be bounded below on the polyhedral set S that contains no straight lines going to infinity in both directions, let f be concave on $\mathbb{R}^k$. Then for a sufficiently large positive but finite value ε_0 of ε the smooth problem (14) has a vertex solution that also solves the original non-smooth problem (12).*

Remark IV.1. Here, the key point is the number of vertices is assumed to be finite, which leads to repeatedly solving this problem on some vertices. That is to say, we can choose at least one vertex, which is the solution corresponding to a set of parameters, namely $(\varepsilon_1, \varepsilon_2, ...) \uparrow \infty$, i.e. monotonically increasing to positive infinity.

According to the discussion above, for a finite and big enough value of ε the smooth problem (14) generates an exact solution of the non-smooth problem (12). Then, note that problem (14) is a minimization of a concave objective function over a polyhedral set, which is difficult to find a global solution. However, [13], [14] posed a fast successive linearization algorithm (SLA) which terminated finitely at a stationary point which satisfied the minimum principle necessary optimality condition for problem (14). As a result, we can employ this method and lead to obtain a sparse $\boldsymbol{\alpha}$ with good generalization properties, especially in error correcting.

B. Successive Linearization Algorithm (SLA)

In this part, SLA, which is essentially a Frank-Wolfe algorithm [17] without a stepsize, will be introduced. This algorithm can solve a typical concave problem in finite steps. In other words, the iteration stops finitely. Particularly, a concave programming on a polyhedron can be stated in following system.

$$\min_{\boldsymbol{x}} \{f(\boldsymbol{x}) | \boldsymbol{Ax} \leqq \boldsymbol{b}, \boldsymbol{x} \geqq 0\}. \tag{15}$$

Here, $f(\boldsymbol{x})$ is a differentiable concave function, bounded below on the nonempty polyhedral feasible region of (15).

To solve the system above, linear searching is applied. Then, SLA can be described below.

Start with a random $\boldsymbol{x^0} \in \mathbb{R}^n$. Having $\boldsymbol{x^i}$ determine $\boldsymbol{x^{i+1}}$ as follows:

$$\boldsymbol{x^{i+1}} \in arg\ vertex \min_{\boldsymbol{x} \in X} \nabla f(\boldsymbol{x^i})(\boldsymbol{x} - \boldsymbol{x^i}),$$
$$X = \{\boldsymbol{Ax} \leqq \boldsymbol{b}, \boldsymbol{x} \geqq \boldsymbol{0}\}.$$

Remark IV.2. There are two remarks need to be stated:

(1) The sign $arg\ vertex \min_{\boldsymbol{x} \in X} f(\boldsymbol{x})$ denotes the set of vertex minimizers of $f(\boldsymbol{x})$ on the polyhedral set X.

(2) SLA can stop in finite steps of iteration [13], [14]. In other words, after finite steps, SLA can find a point $\boldsymbol{x^l}$, which satisfies the minimum principle necessary optimality conditions:

$$\nabla f(\boldsymbol{x^l})(\boldsymbol{x} - \boldsymbol{x^l}) \geqslant 0, \boldsymbol{x} \in X.$$

V. Experiment

In this section, MCLP based misclassification minimization model will be solved by SLA on publicly UCI databases [18]. Here, three data sets are chosen to test our algorithm, including balance and unbalance data sets. Particularly, we will compare original MCLP and misclassification minimization model based on MCLP (MMMCLP) in the number of misclassified points. Here, $\varepsilon = 10$ is fixed for (14). Besides, $\boldsymbol{\lambda} = (0.5, 0.3, 0.2)$. In TABLE I, each data set, 10 fold cross validation is applied and the results are all the average values. In addition, average iteration times are displayed.

TABLE I. A Comparison of SLA for the Smooth Misclassification Minization Based on MCLP to Original MCLP

Data Set	Accuracy Number of Iteration	
	Original MCLP	MMMCLP
Iris	90%	**93.5%** 5.8
Vertebral Column	86.8%	**89.1%** 6.5
German Credit Card	87.5%	**90.2%** 4.2

As we can see in TABLE I, by appropriate parameter choosing, the smooth of misclassification minimization based on MCLP has higher accuracies in all three data sets. The number of iteration is finite and within 10, which means SLA has a good efficiency in solving concave function minimization on a polyhedron.

VI. Conclusion

This paper successfully constructed two different schemes on misclassification minimization in MCLP based classification. For the first scheme, complementarity of original MCLP was employed to build a nonlinear programming problem. An item, which estimated the number of misclassified points, was included in the objective function. Then, linear approximation for nonlinear programming was applied in order to find the solution by simplex method. For the second scheme, taking advantage of a non-smooth mapping, the number of misclassified points was also included in the objective function. Then, based on smoothing of indicative function, an equivalent problem with minimizing differentiable concave goal function on a polyhedron was built. Successive linearization algorithm was applied to this new programming and the effect was notable, according to the numerical experiment result.

The models in this paper are solved without parameter selection. In the further work, a parameter selection strategy will be studied and more details of application will be discussed.

Acknowledgment

This work has been partially supported by the following Grants: Key Project (no. 71331005), Major International Joint Research Project (no. 71110107026) and Youth Fund (no. 71201143) from the National Natural Science Foundation of China.

References

[1] B. Wang and Y. Shi, *Error correction method in classification by using multiple-criteria and multiple-constraint levels linear programming*. International Journal of Computers, Communications and Control, 7, 976-989, 2012.

[2] B. Wang and Y. Wang, *A Multiple-Criteria and Multiple-Constraint Levels Linear Programming Based Error Correction Classification Model*. Procedia Computer Science, 17, 1073-1082, 2013.

[3] Z. Qi, Y. Tian and Y. Shi, *Structural Twin Support Vector Machine for Classification*. Knowledge-Based Systems, 43, 74-81, 2013.

[4] Z. Qi, Y. Tian and Y. Shi, *Twin Support Vector Machine with Universum Data*. Neural Networks, 36C, 112-119, 2012.

[5] Y. Shi, *Multiple Criteria and Multiple Constraint Levels Linear Programming: Concepts, Techniques and Applications*. Singapore: World Scientific, 2001.

[6] Y. Shi, Y. J. Tian, G. Kou, Y. Peng and J. P. Li, *Optimization Based Data Mining: Theory and Applications. Advanced Information and Knowledge Processing*. London: Springer-Verlag, 2011.

[7] J. He, Y. Zhang, Y. Shi and G. Huang, *Domain-driven classification based on multiple criteria and multiple constraint-level programming for intelligent credit scoring*. IEEE Transactions on Knowledge and Data Engineering, 22, 826-838, 2010.

[8] Y. Shi, *Multiple criteria optimization-based data mining methods and applications: a systematic survey*. Knowledge Information System, 24(3), 369-391, 2010.

[9] Y. Sawaragi, H. Nakayama and T. Tanino, *Theory of Multiobjective Optimization*. New York: Academic Press, 1994.

[10] C. Chis and F. Cret, *Solving Nonlinear Programming Problems by Linear Approximations*. Scientifical Researches. Agroalimentary Processes and Technologies, 11(2), 485-490, 2005.

[11] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear programming: theory and algorithms*. New Jersey: John Wiley and Sons, 2013.

[12] P. S. Bradley, O. L. Mangasarian and J. B. Rosen, *Parsimonious Least Norm Approximation*. Computational Optimization and Applications, 11, 5-21, 1998.

[13] O. L. Mangasarian, *Machine learning via polyhedral concave minimization*. Applied Mathematics and Parallel Computing - Festschrift for Klaus Ritter. Physica - Verlag A Springer - Verlag Company, Heidelberg, 1996.

[14] O. L. Mangasarian, *Solution of general linear complementarity problems via nondifferentiable concave minimization*. Acta Mathematica Vietnamica, 22, 199-205, 1997.

[15] O. L. Mangasarian, *Misclassification Minimization*. Journal of Global Optimization, 5(4), 309-323, 1994.

[16] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.

[17] M. Frank and P. Wolfe, *An algorithm for quadratic programming*. Naval Research Logistics Quarterly, 3, 95-110, 1956.

[18] P. M. Murphy and D. W. Aha, *UCI repository of machine learning databases*. Department of Information and Computer Science, University of California, Irvine, California, http://www.ics.uci.edu/AI/ML/MLDBRepository.html., 1992.