Kernel Based Regularized Multiple Criteria Linear Programming Model

Yuehua Zhang^{1,*}, Peng Zhang¹, and Yong Shi^{1,2}

¹ CAS Research Center on Fictitious Economy and Data Sciences,
Beijing 100080, China

² College of Information Science & Technology,
University of Nebraska at Omaha,
Omaha, NE 68182, USA

{zhangyuehua07, zhangpeng04}@mails.gucas.ac.cn
yshi@omaha.edu

Abstract. Although Regularized Multiple Criteria Linear Programming (RMCLP) model has shown its effectiveness in classification problems, its inherent drawback of linear formulation limits itself into only solving linear classification problems. To extend RMCLP into solving non-linear problems, in this paper, we propose a kernel based RMCLP model by using a form $w = \sum_{i=1}^N \beta_i \phi(x_i)$ to replace the original weight w in RMCLP model. Empirical studies on synthetic and real-life datasets demonstrate that our new model is capable to classify non-linear datasets. Moreover, comparisons to SVM and MCQP also exhibit the fact that our new model is superior to other non-linear models in classification problems.

Keywords: MCLP, RMCLP, classification, kernel function.

1 Introduction

Data mining is defined as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data"[1]. Traditionally, data mining uses machine learning, statistical and visualization techniques to discover and present knowledge in a form which is easily comprehensible to humans. From the aspect of methodology, data mining can be performed through association, classification, clustering, prediction, sequential patterns, and similar time sequences. [2]. Classification is one of the most important parts of data mining. Classification generally includes three steps: the first step is to build a model by using a given data which is predetermined. The next step is to test the model. The last step is to use the model to predict unlabeled data if the model accuracy is high enough.

Recent years have witnessed a large body of research work on mining useful knowledge by multiple criteria mathematical programming method where various of classification models have been proposed and received great success in business intelligence [1]. All these models are created mainly by adapting the objective functions of the original multiple criteria linear programming (MCLP) model [4] to improve

* Corresponding author.

MCLP's accuracy and stability, and lots of research works have exhibited their powerful ability to classify different kinds of real-life data. Among all these models, the most recent Regularized Multiple Criteria Linear Programming (RMCLP) model [3] which is created by adding two regularized objective functions into the original MCLP model, has been theoretically demonstrated that it is stable in finding the global optimal solution. The experiment results show that RMCLP model is an effective classification model.

However, RMCLP is inadequate to classify non-linear data sets. To overcome this shortage, in this paper, we add kernel function into the original RMCLP model to enable it to solve non-linear classification problems.

The remained of this paper is organized as follows. In section 2, we introduce the kernel algorithm. In section 3, we give a short review of MCLP and RMCLP model. In section 4, we formulate our new kernel based RMCLP model. In Section 5, we perform experiments on a synthetic dataset. In Section 6, we use two real-life data sets to compare our model with two other non-linear models: MCQP and SVM. In the last section, we conclude our paper with discussions.

2 Kernel Algorithms

Linear classification problem is the most basic situation in the entire classification problems. To classify the non-linear problems, we need to construct non-linear mapping functions. For example, SVM uses a non-linear map ϕ to map a input vector x to a higher dimensional space Z (which is also called as the feature space), then it constructs Optimal Hyper plane in the feature space Z . Assuming the non-linear map is $\phi: R^d \rightarrow Z$, we transform the vector x in the original space to a new vector $z = \phi(x)$ in the feature space Z . Considering the input vector is $x = (x^1, x^2, \dots, x^d)$ with weights $w_i = \alpha_i y_i$, we can get the SVM discrimination function as follows:

$$y = \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right). \quad (1)$$

The parameter $K(x_i, x)$ is the non-linear transformation of the i -th support vector. Figure 1 shows the framework of SVM in the sense of neural networks:

Fig. 1. The framework of SVM

In the discrimination function:

$$\begin{aligned}
 y &= \text{sgn}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \\
 &= \text{sgn}\left(\sum_{i=1}^N \beta_i \cdot \phi(x_i) \cdot \phi(x) + b\right) \\
 &= \text{sgn}(w \cdot \phi(x) + b)
 \end{aligned} \tag{2}$$

Then we set $w = \sum_{i=1}^N \beta_i \phi(x_i)$ where ϕ denotes a higher dimensional feature map as-

sociated with the nonlinear kernel and β_i denotes the variables. By the Representer theorem (Scholkopf and Smola, 2002), we indeed know that the optimal solution has the following form:[6]

$$K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle \tag{3}$$

The most common kernel functions are listed as follows [8]:

(1) Radial Basis Kernel Function:

$$K(x_i, x) = \exp(-c \|x - x_i\|^2 / \sigma^2), \tag{4}$$

where c is a constant, σ is variance;

(2) Polynomial Kernel Function:

$$K(x_i, x) = ((x \cdot x_i) + c)^d, \tag{5}$$

where d is a positive real number, $c \geq 0$;

(3) Sigmoid Kernel Function:

$$K(x_i, x) = \tanh(b(x \cdot x_i) + c), \tag{6}$$

Where $b > 0$, $c < 0$.

3 MCLP and RMCLP Models

3.1 The Formulation of MCLP Model

Suppose each evaluated target x_i is described by n attributes (or variables). Consider l targets where data observation of the i -th target is $x_i = (x_{i1}, \dots, x_{in})^T$, for $i = 1 \dots l$. In linear discriminate analysis, the purpose is to determine the optimal coefficients (or weights) for the attributes, denoted by $w = (w_1, \dots, w_n)^T$ and a boundary value (scalar) b to separate two predetermined classes: G (Good) and B (Bad); that is

Fig. 2. Overlapping of two-class linear discriminate analysis

To measure the separation of Good and Bad, α_i is defined to be the overlapping of the two-class boundary for case x_i . Then, α is denoted as the max overlapping of the two-class boundary for all cases x_i ($\alpha_i < \alpha$). β_i is also defined to be the distance of case x_i from its boundary, while β is the minimum distance of all cases x_i to the boundary ($\beta_i < \beta$). To find the compromise solution of MMD (maximize the minimum distances) and MSD (minimizing the sum of the deviations) for data separation, we want to minimize the sum of α_i and maximize the sum of β_i simultaneously, as follows:

$$\begin{aligned}
 & \text{Minimize } C \sum_{i=1}^l \alpha_i - \sum_{i=1}^l \beta_i \\
 & \text{s.t. :} \\
 & x_{i1}w_1 + \dots + x_{in}w_n = b + \alpha_i + \beta_i, \text{ for } x_i \in B, \\
 & x_{i1}w_1 + \dots + x_{in}w_n = b + \alpha_i + \beta_i, \text{ for } x_i \in G, \\
 & x_{i1}w_1 + \dots + x_{in}w_n = b + \alpha_i + \beta_i, \text{ for } x_i \in G, \\
 & \alpha_i \geq 0, \quad i = 1, \dots, l, \\
 & \beta_i \geq 0, \quad i = 1, \dots, l, \\
 & w_i \in R^n.
 \end{aligned} \tag{7}$$

As we discussed above, the inherent drawback of linear formulation makes MCLP can only solve linear classification problems.

3.2 RMCLP Model

Lots of empirical studies have shown that MCLP is a powerful tool for classification. However, there is no theoretical work on whether MCLP always can find an optimal

solution under different kinds of training samples. To go over this difficulty, recently, Shi et.al [5] proposed a RMCLP model by adding two regularized items $\frac{1}{2}x^T Hx$ and $\frac{1}{2}\alpha^T Q\alpha$ on MCLP as follows:

$$\begin{aligned}
 & \text{Minimize } \frac{1}{2}w^T Hw + \frac{1}{2}\alpha^T Q\alpha + d^T \alpha - c^T \beta \\
 & \text{s.t. :} \\
 & \quad x_{i1}w_1 + \dots + x_{in}w_n = b + \alpha_i - \beta_i, \forall x_i \in G1; \\
 & \quad x_{i1}w_1 + \dots + x_{in}w_n = b - \alpha_i + \beta_i, \forall x_i \in G2; \\
 & \quad \alpha_i, \beta_i \geq 0.
 \end{aligned} \tag{8}$$

where $H \in R^{r \times r}, Q \in R^{n \times n}$ are symmetric positive definite matrices. $d^T, c^T \in R^n$. The RMCLP model is a convex quadratic program. Theoretically studies [5] have shown that RMCLP can always find a global optimal solution.

4 Kernel Based RMCLP Model

In order to solve the non-linear classification problem, we add kernel function into the RMLCP model. As discussed above, we use $w = \sum_{i=1}^N \beta_i \phi(x_i)$ to replace the original weight w in RMCLP model. By letting $\beta_i = \lambda_i y_i$, we get $w = \sum_{i=1}^N \lambda_i y_i \phi(x_i)$ and the new RMCLP model can be formulated as follows:

$$\begin{aligned}
 & \text{Minimize } \frac{1}{2}w^T Hw + \frac{1}{2}\alpha^T Q\alpha + d^T \alpha - c^T \beta \\
 & \text{s.t. :} \\
 & \quad \lambda_1 y_1 (\phi(x_1) \cdot \phi(x_i)) + \dots + \lambda_n y_n (\phi(x_n) \cdot \phi(x_i)) = b + \alpha_i - \beta_i, \forall x_i \in G1; \\
 & \quad \lambda_1 y_1 (\phi(x_1) \cdot \phi(x_i)) + \dots + \lambda_n y_n (\phi(x_n) \cdot \phi(x_i)) = b - \alpha_i + \beta_i, \forall x_i \in G2; \\
 & \quad \alpha_i, \beta_i \geq 0. \\
 & \quad C \leq \lambda_i \leq E;
 \end{aligned} \tag{9}$$

Where $H \in R^{r \times r}, Q \in R^{n \times n}$ are symmetric positive definite matrices. $d^T, c^T \in R^n$. Besides, we add the last constrain $C \leq \lambda_i \leq E$ when we compare the model with SVM. By letting $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, we get the kernel based RMCLP model in Equ. (10).

$$\begin{aligned}
& \text{Minimize } \frac{1}{2} w^T H w + \frac{1}{2} \alpha^T Q \alpha + d^T \alpha - c^T \beta \\
& \text{s.t. :} \\
& \quad \lambda_1 y_1 K(x_1, x_i) + \dots + \lambda_n y_n K(x_n, x_i) = b + \alpha_i - \beta_i, \forall x_i \in G1; \\
& \quad \lambda_1 y_1 K(x_1, x_i) + \dots + \lambda_n y_n K(x_n, x_i) = b - \alpha_i = \beta_i, \forall x_i \in G2; \\
& \quad \alpha_i, \beta_i \geq 0. \\
& \quad C \leq \lambda_i \leq E;
\end{aligned} \tag{10}$$

5 Testing on Synthetic Dataset

In order to investigate the performance of our new model in classifying non-linear problems, we design a two-group non-linear dataset as shown in Fig. 3:

Fig. 3. Original dataset

Fig. 4. Result with original RMCLP

Fig. 5. Result with Polynomial kernel RMCLP

Fig. 6. Result with RBF kernel RMCLP

Our synthetic dataset contains 90 records, with 38 records in Group 1 (the blue points) and 52 records in Group 2 (the red points). Since the number of instances in the two groups is imbalanced, RMCLP model puts all records into Group 2 which is

bigger than Group 1. It can reach the best accuracy among all linear classifiers. Fig. 5 and Fig. 6 show the results of the kernel based RMCLP (with Polynomial kernel and RBF kernel respectively). The two groups are perfectly separated.

6 Experimental Results

In this section, we use two public UCI datasets to compare the performance of Kernel based RMCLP with two other methods: SVM and MCQP (a quadratic formulation of MCLP). Here we only use linear kernel to SVM.

Before building models, we scale each feature into the range $[0, 1]$. For the *Australian* dataset, we randomly divide it into two parts: training set with 200 records and testing set with 490 records. For *Heart* dataset, we also randomly split it into two parts: training set with 100 records and testing set with 170 records.

In every training process, parameters in Kernel based RMCLP are selected in some discrete sets in order to get the best accuracy. We select the parameter E from the set $[0.1 \ 1 \ 16 \ 128 \ 1024 \ 10000]$, C from the set $[-\infty \ 0]$, and γ from $[0.001 \ 0.01 \ 0.1 \ 1 \ 16 \ 128 \ 1024]$. [2] When getting the model for every C , E and γ , we test the performances on the training sets. Then, we fix the C , E and γ set which achieves the highest accuracies to predict the test sets and record the accuracies in Table 1 and Table 2.

Table 1. Test On Australian Dataset

Classification	Training(200 records)	Testing (490 records)
Algorithms	Training Accuracy	Testing Accuracy
MCQP	83.5%	84.49%
SVM	85%	84.55%
RMCLP	85.5%	81.84%
K-RMCLP	90.5%	87.14%

Table 2. Test On Heart Dataset

Classification	Training(100 records)	Testing (170 records)
Algorithms	Training Accuracy	Testing Accuracy
MCQP	84%	83.53%
SVM	84%	82.35%
RMCLP	79%	81.18%
K-RMCLP	90%	85.88%

In the table above, we can observe that RBF Kernel based RMCLP performances better than the other two algorithms. And the prediction accuracy improvement of our new model has a strong relationship with the linear-separable degree of the training sample. Through these experiments, we find out that the Kernel based RMCLP is a competitive method in non-linear classification.

7 Conclusions

In this paper, we add the kernel function into the Regularized Multiple Criteria Linear Programming (RMCLP) model to deal with the non-linear classification problems. By the *Representer Theorem*, we know that the optimal solution has the following form:

$$K(x_i, x) = \langle \phi(x_i), \phi(x) \rangle. \text{ After replacing } w \text{ in RMCLP model with } w = \sum_{i=1}^N \beta_i \phi(x_i),$$

we upgrade the original RMLCP model into tackling the non-linear classification problems. Experimental results on both synthetic and real-life datasets show that our model is effective in classifying non-linear datasets. In the future work, we will study the parameters' effect of the kernel functions in our new kernel based RMCLP model.

References

1. Olson, D., Shi, Y.: Introduction to Business Data Mining. McGraw-Hill/Irwin (2007)
2. Zhang, Z., Zhang, D., Tian, Y., Shi, Y.: Kernel Based Multiple Criteria Linear Program
3. Frawley, W., Piatetsky-Shapiro, G., Matheus, C.: Knowledge Discovery in Databases: An Overview. *AI Magazine*, 213–228 (Fall 1992)
4. Shi, Y.: Multiple criteria and multiple constraint levels linear programming: concepts, techniques and applications. World Scientific Pub. Co Inc., New Jersey (2001)
5. Shi, Y., Tian, Y., Chen, X., Zhang, P.: A Regularized Multiple Criteria Linear Program for Classification. In: *ICDM Workshops 2007*, pp. 253–258 (2007)
6. Chapelle, O., Sindhwani, V.: Optimization Techniques for Semi-Supervised Support Vector Machines. *Journal of Machine Learning Research* 9, 203–233 (2008)
7. Zhang, P., Tian, Y., Zhang, Z., Li, X., Shi, Y.: Supportive instances for Regularized Multiple Criteria Linear Programming Classification
8. Deng, N., Tian, Y.: New Approach in Data Mining – Support Vector Machine. Science Press, Beijing (2004)