DOI: 10.21276/sjet.2017.5.2.4

Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2017; 5(2):54-57 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Research Article

A Linear Programming Model of Multi-class Support Vector Machine

Sa Qi, Yuan Yuping *, Tong Ying, Kan Tao, Zhang Jing
College of Sciences, Heilongjiang Bayi Agricultural University, Daqing163319, China

***Corresponding author**
Yuan Yuping
Email: bynd_yyps@sina.com

Abstract: The structure of K-SVCR algorithm is ‘one-against-one-against-rest’. Its advantage is in the process of each decomposition , make all training points of information have been fully taken advantage of. To a certain extent, can prevent the classification error by incomplete information . But this algorithm constructed a quadratic programming which restricted this algorithm’s speed and the range of applications. So, this paper constructed a linear programming model based on K-SVCR , and then adopt the effective algorithm of Predictor-corrector Method of Mehrotra to solve the linear programming. Preliminary numerical experiments on benchmark datasets show that the algorithm has good performance on both accuracy and training speed than K-class Support Vector Classification-Regression.
Keywords: Support vector machine ; Multi-class classification ; Primal dual interior point method.

INTRODUCTION

Support Vector Machine (SVM), motivated by statistical learning theory [1], is a new tool by means of optimization method for solving the problem of machine learning. Due to its unique global solutions and the accuracy, is a very effective method for classification and regression question. It has been widely applied to many aspects, such as speech recognition, font recognition, face recognition, etc., and it has been a lot of achievements.

SVM transforms machine learning to solve an optimization problem, and to solve a convex quadratic programming problem by the optimization theory. Solving the quadratic programming is an iterative process, the difficulty is that the Hessian matrix of denseness. When large amounts of sample data, will face a dimension disaster or cannot be calculated due to occupy larger computer memory. Consequently, now a lot of research is basically aimed at how to reduce the amount of calculation and complexity for solving quadratic programming problems. Such as decomposition of the Sequential Minimal Optimization （SMO） [2], and Least Squares Support Vector Machine (LSSVM) [3], Smooth Support Vector Machine(SSVM) [4], Linear Programming Support Vector Machine(LPSVM) [5] etc.

In this paper, we first propose a new linear model based on the K-SVCR method [6], then take infeasible primal dual interior point method to solve it. Due to the linear programming problems is a mature discipline on the theoretical and calculation method, on the computing speed much faster than the quadratic programming, the experimental results also show that this algorithm on the velocity ratio based on quadratic programming of K-SVCR has good performance.

INFEASIBLE PRIMAL DUAL INTERIOR POINT METHOD

Infeasible primal dual interior point method is based on the logarithmic barrier function, adopts the primal dual interior point algorithm structure, produced a series of points does not limit points in the feasible region. Each iteration either constrain equations of the original problem or the dual problem of constrain equations was established, but the original problem and the dual problem of variable must satisfy of positivity. The algorithm is the key to control the search point in the interior of the feasible region, and set up a barrier on the boundary of the feasible region. When iteration points close the feasible region boundary, the objective function value increases quickly, and proper control in the iteration step length, make the iteration point always stay inside the feasible region. With fewer obstacle factors, the role of the barrier function will reduce gradually, finally will converge to the optimal solution of original problem.

Because in the infeasible primal-dual interior point method solving the Newton equation is the main computation of interior-point methods, include the factor of coefficient matrix, the former generation and generation. The amount of calculation of the factor higher than the former generation and generation. Mehrotra predictor-corrector method is to Newton direction is divided into two parts: direction of part is predicted to cut the infeasible of primal dual and dynamic estimation parameters, the other part is the correct direction to maintain the current iteration point away from the boundary of the feasible region [7-9]. Therefore Mehrotra predictor-corrector method in each iteration step, needs to solve two same size and same sparse equations. Although increased each iteration calculation, need only one factor, can reduce the total number of iterations and computation time significantly, greatly improve the performance of the primal-dual interior point method.

Variables in the original problem or the dual problem, the moving direction are affected by the obstacle factors. Selection of obstacle factors directly affect the convergence of the algorithm. Mehrotra predictor-corrector method can effectively estimate obstacle factors, well coordinate the relationship between the optimality and feasibility of solution, and the convergence of the algorithm is improved.

THE LINEAR PROGRAMMING MODEL OF K-SVCR

Definition (multi-classification problems)

Let the training set:

$$T=\{(\boldsymbol{x}_1,y_1),\cdots,(\boldsymbol{x}_l,y_l)\}\in\{X\times Y\}^l, \tag{1}$$

where $\boldsymbol{x}_i\in X\subset R^n$, $y_i\in Y=\{\theta_1,\ \theta_2,\cdots,\ \theta_K\}$, $i=1,\cdots,\ l$,we wish to construct a from the input space $X\subset R^n$ on a decision function $f(\boldsymbol{x})$, in order to use decision function $f(\boldsymbol{x})$ concluded that any patterns $\boldsymbol{x}$ correspond y .Therefore, multi-classification problems is to find rules that put the points on the R^n into K parts.

Currently, there are roughly two types of approaches for multi-class classification. One is the “decomposition-reconstruction” architecture approach which consists of the ‘one-against-all’ method [10], the other is the “all-together” approach [11-12] which solves multi-class classification problems by considering all patterns from all classes in one optimization formulation. We wish to construct a decision function $f(x)$ which separates the two classes θ_j and θ_k as well as the remaining classes by ‘one-against-one-rest’ structure. Without loss of generality, let patterns $\boldsymbol{x}_i,\ i=1,\cdots,\ l_1$, and $\boldsymbol{x}_i,\ i=l_1+1,\cdots,\ l_1+l_2$ belong to the classes θ_j and θ_k which will be labeled $+1$ and -1, respectively, and the remaining patterns belong to the other classes which will be labeled 0 . Specifically, we wish to find a function $f(x)$ such that:

$$f(\boldsymbol{x}_i)=\begin{cases}+1, & i=1,\cdots,\ l_1,\\ -1, & i=l_1+1,\cdots,\ l_1+l_2,\\ 0, & i=l_1+l_2+1,\cdots,\ l.\end{cases} \tag{2}$$

Formulation

The recently proposed K-SVCR method in ref.[6], original problem as follows:

$$\min_{(w,b,\xi,\eta,\eta^*)\in R^{n+1+l_{12}+l_3+l_3}} \frac{1}{2}\|w\|^2+\delta_1 e\xi+\delta_2(e\eta+e\eta^*) \tag{3}$$

$$s.t.\ \ D(Aw+eb)\ge e-\xi\ , \tag{4}$$

$$Bw+eb\le\varepsilon e+\eta\ , \tag{5}$$

$$-Bw-eb\le\varepsilon e+\eta^*. \tag{6}$$

$$\xi,\ \ \eta,\ \ \eta^*\ge 0 \tag{7}$$

where $\delta_1,\delta_2>0$ and $\varepsilon\in[0,1)$ chosen a priori, $A\in R^{l_{12}\times n}$ with the *i*th row corresponding to each pattern in θ_j and θ_k , D is a diagonal matrix with plus ones and minus ones according to the class labels of $A_{i\cdot}$. $B\in R^{l_3\times n}$ with $B_{i\cdot}$ being the *i*th row corresponding to each pattern in the remaining classes. $\boldsymbol{e}$ is the vector of ones.

$\left(\alpha, \ \alpha^*, \ \xi, \ \eta, \ \eta^*, \ b\right)$ is variable of the formulation (3)-(7) in the dual problem.

where $\quad \alpha = \left(\alpha_1, \ \alpha_2, \cdots \alpha_l\right)^T, \ \alpha^* = \left(\alpha_1^*, \ \alpha_2^*, \cdots \alpha_l^*\right)^T$

$$\left(\alpha_i, \ \alpha_i^* \geq 0, \ i=1,\cdots, \ l\right) \tag{8}$$

$$w = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i^*\right) x_i = C(\alpha - \alpha^*) \tag{9}$$

$$\left\| \alpha^{(*)} \right\| = \sum_{i=1}^{l} \left(\alpha_i + \alpha_i^*\right) = e(\alpha + \alpha^*) \tag{10}$$

The formulation (3)-(7) will the following transform:

(i) Use the program (10) instead of the $\|w\|^2$ of the program (3);

(ii) Put the program (9) into the formulation (4)-(6);

(iii) Adding constraint condition the program (8)

The formulation (3) - (7) is equivalent to the following formulation.

$$\min_{\alpha,\alpha^*,\xi,\eta,\eta^*,b} \quad e(\alpha+\alpha^*) + \delta_1 e\xi + \delta_2 e(\eta+\eta^*) \tag{11}$$

$$s.t. \quad DAK\left(C,(\alpha-\alpha^*)^T\right) + Deb \geq e - \xi, \tag{12}$$

$$BK\left(C,(\alpha-\alpha^*)^T\right) + eb \leq \varepsilon e + \eta, \tag{13}$$

$$-B\left(C,(\alpha-\alpha^*)^T\right) - eb \leq \varepsilon e + \eta^*, \tag{14}$$

$$\alpha, \ \alpha^*, \ \xi, \ \eta, \ \eta^* \geq 0. \tag{15}$$

where $\boldsymbol{e}$ is the vector of ones. $\delta_1, \delta_2 > 0$ and $\varepsilon \in [0,1)$ chosen a priori, C is a matrix of all training points. $K(\cdot,\cdot)$ is kernel function. The optimal solution of the formulation (11)-(15) is $\left(\tilde{\alpha}, \ \tilde{\alpha}^*, \ \tilde{\xi}, \ \tilde{\eta}, \ \tilde{\eta}^*, \ \tilde{b}\right)$, the hyperplane decision function can be written as

$$f_{\theta_{j,k}}(\boldsymbol{x}) = \begin{cases} +1, & \text{if } \boldsymbol{x}^{\mathrm{T}} K\left(C,(\tilde{\alpha}-\tilde{\alpha}^*)^T\right) + \tilde{b} \geq \varepsilon \\ -1, & \text{if } \boldsymbol{x}^{\mathrm{T}} K\left(C,(\tilde{\alpha}-\tilde{\alpha}^*)^T\right) + \tilde{b} \leq -\varepsilon \\ 0, & others \end{cases} \tag{16}$$

In the K-class problem, for each pair, say θ_j and θ_k, we have a classifier $f_{j,k}(\cdot)$ gotten by (16) to separate them as well as the other classes. So we have $\mathrm{K}(\mathrm{K}-1)/2$ decision function in total. For an example $\boldsymbol{x}_p$, we get $\mathrm{K}(\mathrm{K}-1)/2$ outputs. We translate the outputs as follows: when $f_{j,k}(\boldsymbol{x}_P) = +1$, a positive vote is added on θ_j, and no votes are added on the other classes; when $f_{j,k}(\boldsymbol{x}_P) = -1$, a positive vote is added on θ_k, and no votes are added on the other classes; when $f_{j,k}(\boldsymbol{x}_P) = 0$, a negative vote is added on both θ_j and θ_k, and no votes are added on the other classes; After we translate all of the $\mathrm{K}(\mathrm{K}-1)/2$ outputs, we will get the total votes of each class by adding the positive and negative votes on the class. Finally, x_p will be assigned to the class that gets the most votes.

Algorithm 3.1 (The infeasible primal-dual interior point algorithm for program (11)-(15))

Step 0: The training set $T = \{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_l, y_l)\} \in \{X \times Y\}^l$, $\boldsymbol{x}_i \in X = R^n$, $y_i \in Y = \{\theta_1, \ \theta_2, \cdots, \ \theta_K\}, i = 1, \cdots, \ l$;

Step 1: For each pair, say θ_j and θ_k, chosen the appropriate parameter values of

$\delta_1, \ \delta_2, \ \varepsilon\left(\delta_1, \ \delta_2 \geq 0, \ \varepsilon > 0\right)$ and chosen the proper kernel function $K\left(\mathbf{x}, \ \mathbf{x}'\right)$;

Step 2: Use the infeasible primal-dual interior point algorithm to calculate the program (11)-(15), to get the optimal solution $\left(\tilde{\alpha}, \ \tilde{\alpha}^*, \ \tilde{\xi}, \ \tilde{\eta}, \ \tilde{\eta}^*, \ \tilde{b}\right)$;

Step 3: Construct the decision function (16) ;

Step 4: For an new example x_p, so we have $\mathrm{K(K}-1)/2$ decision function in total, x_p will be assigned to the class that gets the most votes .

PRELIMINARY NUMERICAL RESULTS

We select Iris , Wine and Glass data sets, select the kernel function $K\left(x, \ x'\right) = \left(\left(x \cdot x'\right) + 1\right)^d, d = 1$, and $K\left(x, \ x'\right) = \exp\left(-\|x - x'\|^2 / 2\sigma^2\right), \ \sigma = 2$.Comparisons between the algorithm 3.1 results and $\mathrm{K-SVCR}$ method results, in table 3.1.

Table 1: The percentage of error and the training time by K-SVCR and Algorithm 3.1

Data Set	#pts	#atr	#class	K-SVCR		Algorithm 3.1	
				the percentage of error (%)	train test time (s)	The percentage of error (%)	train test time (s)
Iris	150	3	4	[1.93, 3.0]	154.23	[0.67,0.67]	5.065
Wine	178	3	13	[2.29, 4.29]	178.11	[1.76,1.76]	8.829
Glass	214	6	9	[30.47, 36.35]	2577.20	[20.95,27.62]	169.861

[·, ·] The first data indicates the testing precision while test points be categoried as classification error, the second data indicates the testing precision while test point be categoried as classification correct.

CONCLUSION

We have first describes the theory involved in infeasible primal-dual interior point algorithm. The algorithm is one of the most effective algorithm to solve the problem of large scale linear programming. With the algorithm to solve the linear programming of K-SVCR, the results of numerical experiments show that the algorithm achieve good performance in both speed and accuracy.

This work was supported by the project of College students's innovative entrepreneurial training in heilongjiang province (201610223003). This work was supported by the project of Daqing's science and technology instruction (zd-2016-078); This work was supported by the project of natural fund in heilongjiang province, 2016.

REFERENCES

1. Vapnik V. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.
2. Platt J. Sequential minimal optimization: a fast algorithm for training support vector machines. Scholkopf B, Burges C J C, Smola A J. Advances in Kernel Methods-Support Vector Learning. Cambridge: MIT Press,1999;185-208.
3. Suykens JAK, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett.1999;9:293-300.
4. Lee Yuh-jye, Mangasarian OL. A smooth support vector machine. Comput Optim and Appl. 2001;20(1):5-22.
5. Zhou W, Zhang L, Jiao L. Linear programming support vector machine. Pattern Recog. 2002;35:2927-2936.
6. Angulo C, Parra X, Catala A. K-SVCR A support vector machine for multi-class classification. Neurocomputing. 2003;55:57-77.
7. Hu QW, Wei YM. Linear programming and its application. Science press, Bei jing, 2004.
8. Lusting IJ, Marsten RE, Shanno DF. Interior point method for linear programming: computational state of the art .ORSA Journal on Computing.1994;(6):1-14.
9. Gonzaga CC. Path following methods for linear programming. SIAM Review. 1992;34(2):167-227.
10. Bottou L, Cortes C, Denker JS, Drucker H, Guyon I, Jacker LD, Le Cun Y, Muller UA, Sackinger E, Simard P, Vapnik V. Comparison of classifier methods: a case study in handwriting, in Proceeding of the International Conference on Pattern Recognition (ed, IAPR), IEEE Computer Society Press. 1994;77-82.
11. Grammer K, Singer Y. On the algorithmic implementation of multiclass kernel-basd vector machines. Journal of Machine Learning Research. 2002;2:265-292.
12. Zhong P, Fukushima M. A new multi-class support vector algorithm. Optimization Methods & Software. 2006;21:359-372.