Pattern Recognition 46 (2013) 2267–2277

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A GA-based model selection for smooth twin parametric-margin support vector machine

Zhen Wang [a,*], Yuan-Hai Shao [b], Tie-Ru Wu [a]

[a] Mathematics College of Jilin University, Changchun 130012, PR China
[b] Zhijiang College, Zhejiang University of Technology, Hangzhou 310024, PR China

ARTICLE INFO

Article history:
Received 31 January 2012
Received in revised form
17 September 2012
Accepted 14 January 2013
Available online 29 January 2013

Keywords:
Pattern classification
Support vector machine
Twin support vector machine
Smoothing techniques
Genetic algorithm

ABSTRACT

The recently proposed twin parametric-margin support vector machine, denoted by TPMSVM, gains good generalization and is suitable for many noise cases. However, in the TPMSVM, it solves two dual quadratic programming problems (QPPs). Moreover, compared with support vector machine (SVM), TPMSVM has at least four regularization parameters that need regulating, which affects its practical applications. In this paper, we increase the efficiency of TPMSVM from two aspects. First, by introducing a quadratic function, we directly optimize a pair of QPPs of TPMSVM in the primal space, called STPMSVM for short. Compared with solving two dual QPPs in the TPMSVM, STPMSVM can obviously improve the training speed without loss of generalization. Second, a genetic algorithm GA-based model selection for STPMSVM in the primal space is suggested. The GA-based STPMSVM can not only select the parameters efficiently, but also provide discriminative feature selection. Computational results on several synthetic as well as benchmark datasets confirm the great improvements on the training process of our GA-based STPMSVM.

1. Introduction

Support vector machine (SVM), being computationally powerful tools for supervised learning [1–3], has already outperformed most other systems in a wide variety of applications [4–7]. For the standard SVM, its primal problem can be understood in the following way: constructs two parallel supporting hyperplanes such that, on the one hand, the band between the two parallel hyperplanes separates the two classes (the positive and negative data points) well; on the other hand, the width of the band is maximized, leading to a quadratic programming problem (QPP). The final separating hyperlane is selected to be the "mid-one" between the two supporting hyperplanes after solving the QPP.

Different from SVM with two parallel hyperplanes, some nonparallel hyperplane classifiers such as generalized eigenvalue proximal support vector machine (GEPSVM), twin support vector machine (TWSVM) and twin parametric margin support vector machine (TPMSVM) were proposed in [8–10], respectively. Similar to GEPSVM, TWSVM [9,11] seeks two nonparallel proximal hyperplanes such that each hyperplane is closest to one of two classes and has a certain distance far from the other. Further, TPMSVM [10] was proposed in the spirit of TWSVM. Different from the TWSVM, TPMSVM seeks two nonparallel parametric-margin hyperplanes such that each hyperplane is as far as possible away from one side of one of two classes. A fundamental difference between TPMSVM and SVM is that, TPMSVM solves two smaller size QPPs, whereas SVM solves a large one. Therefore, TPMSVM works faster than SVM. Experimental results in [10] showed the effectiveness of TPMSVM over both standard SVM and TWSVM on several public available datasets. Thus, the methods of constructing the nonparallel hyperplanes have been studied extensively [12–17].

Compared with SVM, TPMSVM [10] owns better generalization and is suitable for many cases for it obtains a pair of nonparallel parametric-margin hyperplanes. The training process of TPMSVM involves the solution of two QPPs. Its computational complexity still approximately equal to $1/4\,O(m^3)$, where m is the total number of training data points. Though computing results show that TPMSVM is faster than SVM by solving dual QPPs using standard QP solvers, there exist several much faster decomposition methods to solve SVM (e.g., SMO [18,19]). Hence the authors have left speeding up TPMSVM as a subject of future work. Further, there are at least four regularization parameters that need regulating in the TPMSVM, a serious problem is the parameters selection in the TPMSVM. These drawbacks restrict the application of TPMSVM to the large scaled problems.

In this paper, we make some improvements on TPMSVM and propose a primal version, named smooth twin parametric-margin support vector machine (STPMSVM). First, following [12,20], we

* Corresponding author. Tel./fax: +86 571 87313551.
E-mail addresses: wangz11@mails.jlu.edu.cn, wangzhen1882@126.com (Z. Wang).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.01.023

solve two QPPs of TPMSVM in the primal space by converting them into two unconstrained minimization problems (UMPs), rather than two QPPs solved in the dual space in [10]. Smoothing techniques are used to make the objective function of UMPs twice differentiable, which makes the UMPs can be solved using fast Newton method with Armijo stepsize [20]. Second, we employ the genetic algorithm (GA) to our STPMSVM, our GA-based STPMSVM can not only select parameters efficiently but also provide discriminative feature selection. Computational comparisons of our STPMSVM against several state of the art binary classifiers, in terms of classification accuracy, computing time and model selection, have been made on benchmark datasets for both linear and nonlinear kernel.

The primary objective of this paper is to improve training procedure of TPMSVM such that it can be applied to large scaled problems without any loss of the generalization ability. In what follows, we give the reasons for choosing smoothing technique and GA to improve TPMSVM, and summarize favorable and attractive characteristics of the proposed as follows.

(i) While the computational complexity of primal QPP is the same as its dual problem, solving primal QPP is believed to be advantageous than solving dual QPP for large scaled problems [21]. This is because when the number of data points is very large, it becomes intractable to compute exact solution of the QPP and one has to look for approximate solutions. The approximate dual solution would not be a good approximate primal solution which we are ultimately interested in. Many decomposition methods, like SOR [22], SMO [18] and SVMlight [23], have been proposed for solving dual QPP but, smoothing techniques for solving primal QPP have already been successfully applied to SVM [20] and TWSVM [12], and have been shown to be faster than decomposition methods like SOR, SMO and SVMlight. In addition, the same as TWSVM, TPMSVM loses the sparsity [10], directly optimizing a small subset of the variables in the dual space during the iteration procedure (such as SMO) may be unsatisfactory. However, smoothing techniques enable us to solve primal QPP in spite of the sparsity. Hence, they are more suitable for large scaled problems.

(ii) The genetic algorithm (GA) [24,25] has been successfully presented for SVM model selection [26,27], both for parameter selection and feature selection. With GA, SVM has achieved better performance in terms of both computing time and generalization than the grid and some other techniques [26]. It has been shown that GA is more suitable for the case with more parameters because of its reduced memory usage and fast computing time [26,27]. Different from SVM and TWSVM, TPMSVM has at least four parameters to choose properly, so does STPMSVM. Without any loss of generalization, STPMSVM is faster than the TPMSVM but not enough in the model selection, so we introduce GA in the parameter selection to save training time. Moreover, in [26,27], it has been shown GA can provide discriminative feature selection. Thus, employing GA to our STPMSVM will obtain the benefit of selecting the parameters and features through one procedure.

(iii) Reduced kernel technique [28,29] has been proposed for solving Smooth SVM (SSVM) [20] and Smooth TWSVM (STWSVM) [12] with nonlinear kernel, which use a rectangular kernel $K(A,\overline{A}^{\top})$ instead of the primal kernel $K(A,A^{\top})$, where $\overline{A}$ is a randomly selected subset from the original dataset A that is typically 10% or less. With rectangular kernel, RSVM [28] has achieved better performance in terms of both computing time and generalization than using the complete square kernel $K(A,A^{\top})$. It has been shown that RSVM is more suitable for solving very large scaled problems with nonlinear kernel because of its reduced memory usage and fast computing time. In later sections, we will show that reduced kernel technique for nonlinear STPMSVM is a natural extension. Experiments show that the reduced nonlinear STPMSVM can deal with 100 K data points with 32 features in 1.8 s, and enjoy more computational advantages than TPMSVM.

This paper is organized as follows. In Section 2, a briefly review of SVM, TWSVM and TPMSVM is given. Our linear and nonlinear STPMSVM is formulated in Section 3. In Section 4, our GA-based strategy for model selection of STPMSVM is arranged. Experimental results are described in Section 5, and concluding remarks are given in Section 6.

2. Preliminaries

Consider a binary classification problem in the n-dimensional real space R^n. The set of training data points is represented by $T=\{(x_i,y_i)|i=1,2,\ldots,m\}$, where $x_i\in R^n$ is input and $y_i\in\{+1,-1\}$ is corresponding output. We further organize the m_1 inputs of Class $+1$ by matrix $X_1\in R^{m_1\times n}$ and the m_2 inputs of Class -1 by matrix $X_2\in R^{m_2\times n}$, correspondingly, the m_1 outputs of Class $+1$ by vector $Y_1\in R^{m_1}$ and the m_2 outputs of Class -1 by vector $Y_2\in R^{m_2}$. Below, we give a brief outline of the SVM, TWSVM and TPMSVM.

2.1. v-SVM

Linear v-support vector machine (v-SVM) [30], one formulation of standard SVM, searches for a separating hyperplane

$$f(x)=w^{\top}x+b=0, \tag{1}$$

where $w\in R^n$ and $b\in R$. To measure the empirical risk, the soft margin loss function $\sum_{i=1}^{m}\max(0,\rho-y_i(w^{\top}x_i+b))$ is used. By introducing the regularization term $\frac{1}{2}\|w\|^2$ and the slack vector $\eta=(\eta_1,\ldots,\eta_m)^{\top}$, the primal problem of v-SVM can be expressed as

$$\begin{aligned}&\min_{w,b,\eta,\rho}\ \frac{1}{2}\|w\|^2-v\rho+\frac{1}{m}\sum_{i=1}^{m}\eta_i,\\ &\text{s.t. } y_i(w^{\top}x_i+b)\geq\rho-\eta_i,\quad \eta_i\geq 0,\ \rho\geq 0,\ i=1,\ldots,m,\end{aligned} \tag{2}$$

where $v\in(0,1)$ is a parameter with some quantitative meanings [30]. To be more precise, v is an upper bound on the fraction of margin errors and a lower bound on the fraction of support vectors. In addition, with probability 1, asymptotically, v equals to both fractions. Note that the minimization of the regularization term $\frac{1}{2}\|w\|^2$ is equivalent to the maximization of the margin between two parallel supporting hyperplanes $w^{\top}x+b=\pm\rho$. And the structural risk minimization principle is implemented in the v-SVM.

2.2. TWSVM

The linear TWSVM [9,11] seeks two nonparallel hyperplanes in R^n which can be expressed as

$$f_1(x)=w_1^{\top}x+b_1=0\quad\text{and}\quad f_2(x)=w_2^{\top}x+b_2=0, \tag{3}$$

such that each hyperplane is the closest to the data points of one class and has a certain distance far from the data points of the other class. A new data point is assigned to Class $+1$ or -1 depending upon the distances to the two nonparallel hyperplanes. To find the hyperplanes, it is required to get the solutions