Pattern Recognition 48 (2015) 984–992

Contents lists available at ScienceDirect

Pattern Recognition

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

Least squares twin multi-class classification support vector machine

Jalal A. Nasiri, Nasrollah Moghadam Charkari *, Saeed Jalili

Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Article history:
Received 6 July 2013
Received in revised form
20 July 2014
Accepted 18 September 2014
Available online 30 September 2014

Keywords:
Twin support vector machine
Least squares
Multi-class classification
Nonparallel plane
K-SVCR

ABSTRACT

Twin K-class support vector classification (Twin-KSVC) is a novel multi-class method based on twin support vector machine (TWSVM). In this paper, we formulate a least squares version of Twin-KSVC called as LST-KSVC. This formulation leads to extremely simple and fast algorithm. LST-KSVC, same as the Twin-KSVC, evaluates all the training data into a “1-versus-1-versus-rest” structure, so it generates ternary output $\{-1, 0, +1\}$. In LST-KSVC, the solution of the two modified primal problems is reduced to solving only two systems of linear equations whereas Twin-KSVC needs to solve two quadratic programming problems (QPPs) along with two systems of linear equations. Our experiments on UCI and face datasets indicate that the proposed method has comparable accuracy in classification to that of Twin-KSVC but with remarkably less computational time. Also, because of the structure “1-versus-1-versus-rest”, the classification accuracy of LST-KSVC is higher than typical multi-class method based on SVMs.

1. Introduction

Support vector machine (SVM) was originally proposed by Vapnik [1,2] for binary classification. In contrast with other machine learning approach like artificial neural network which aims at reducing empirical risk, SVM implements the structural risk minimization (SRM) that minimizes the upper bound of generation error. SVM has been successfully applied in wide spectrum of research areas like face recognition, text categorization, and biomedicine [3–9]. One of the main challenges in the classical SVM is the high computational complexity of quadratic programming problem (QPP) [10]. The computational complexity of SVM is $O(l^3)$, where l denotes as the total size of training data. This drawback restricts the application of SVM to large-scale problem domains.

Twin support vector machines (TWSVM) were proposed by Jayadeva et al. in [11] for binary classification. TWSVM generates two nonparallel hyper-planes by solving two smaller-sized QPPs such that each hyper-plane is closer to one class and as far as possible from the other. The idea of solving two smaller-sized QPPs rather than a single larger-sized QPP in SVM makes the learning of TWSVM four times faster than the conventional SVM [11]. Some extensions of TWSVM such as twin bounded support vector machines (TBSVM) [12], Robust TWSVM [13] and Projection TWSVM [14] have been proposed to achieve higher accuracy with lower computational time in comparing with SVM families. Least squares twin support vector machine (LS-TWSVM) [15] has been proposed as a way to replace the convex QPPs in TWSVM with a convex linear system by using a squared loss function instead of the hinge one. Inspired by LS-TWSVM, LS-PTWSVM has been introduced as a least squares version of projection twin support vector machine [16]. LS-TWSVM and LS-PTWSVM have extremely fast training speed since their separating hyper-planes are determined by solving a single system of linear equations.

SVM and TWSVM are suitable for binary classification problems. However, Multi-class classification problem is often occurred in real life. In the SVM and TWSVM family framework, “1-versus-rest” [17] and “1-versus-1” [18] approaches are usually resolve multi-class classification. In “1-versus-rest”, K binary SVM classifiers are constructed. Each binary SVM is trained with all of the patterns, so it easily leads to the class imbalance problem, Whereas TWSVM address this problem. The second structure, “1-versus-1”, needs to construct K(K-1)/2 binary (Twin) SVMs. Each classifier is involved with the training data of two classes. In this case, the information of the remaining samples is omitted in each binary classification. Therefore, unfavorable results may be received [19]. A new multi-class method based on “1-versus-1-versus-rest” structure called K-SVCR (support vector classification regression for K-class classification) was proposed in [19]. It constructs K(K- 1)/2 binary K-SVCR classifiers for a K-class classification. This structure provides better forecasting results. However, as all the training data are utilized in constricting the decision classification, its time complexity is higher than the former structures. Twin-KSVC based on K-SVCR and TWSVM was proposed in [20]. It takes the advantage of both TWSVM and

* Corresponding author. Tel.: +98 21 82883301.
E-mail addresses: j.nasiri@modares.ac.ir (J.A. Nasiri), charkari@modares.ac.ir (N. Moghadam Charkari), sjalili@modares.ac.ir (S. Jalili).

http://dx.doi.org/10.1016/j.patcog.2014.09.020
0031-3203/

K-SVCR. In the term of computational time, Twin-KSVC requires nearly the same run-time as "1-versus-rest" structure of TWSVM, while its runtime is far lower than K-SVCR.

In this paper, following the line of research in [11,15,20], we propose a least squares version of Twin-KSVC, called least squares twin K-class support vector classification (LST-KSVC) using the strategy of LS-TWSVM and K-SVCR. The QPPs of our LST-KSVC have only equality constraints while inequality constraints appear in the Twin-KSVC. Thus, the solution of LST-KSVC follows directly from solving two systems of linear equation as opposed to solving two QPPs and two systems of linear equation in Twin-KSVC. It takes both the advantages of LS-TWSVM in time complexity and K-SVCR in higher multi-class classification accuracy based on "1-versus-1-versus-rest" structure. The experimental results on benchmark datasets show that the proposed LST-KSVC has comparable classification accuracy to that of the Twin-KSVC but with remarkably less computational time. In addition, the proposed algorithm can properly cope with large dataset without any external optimizers.

This paper is organized as follows. Section 2 briefly dwells on the TWSVM, K-SVCR and Twin-KSVC. LST-KSVC is formulated and described in Section 3, which includes linear, nonlinear cases and classification decision rule. Section 4 provides some interesting experimental results on datasets to investigate our proposed multi-class algorithm and concluding remarks are given in Section 5.

2. Preliminaries

In this section, we give a brief description of TWSVM and Multi-Class SVM based on "1-versus-1-versus-rest" structure for classification purposes.

2.1. TWSVM

TWSVM is a binary classifier that performs classification by the use of two non-parallel hyperplanes unlike SVM which used a single hyperplane [11]. Let us consider dataset D which d^+ is training set with label $+1$ and d^- is training set with label -1 in the m-dimensional real space R^m. Let matrix $A \in R^{d^+ \times m}$ represent the training data belong $+1$ and matrix $B \in R^{d^- \times m}$ represents the training data belong to the class -1. The linear TWSVM search for two non-parallel hyper-planes in R^m as follows:

$$x^T w_{(1)} + b_{(1)} = 0 \quad \text{and} \quad x^T w_{(2)} + b_{(2)} = 0 \tag{1}$$

Such that each hyperplane is closest to the training data of one class and farthest from the training data of another class. A new data sample is assigned to class $+1$ or -1 depends on which of the two planes is closest to it. The linear TWSVM solves two QPPs (2) and (3) with objective function corresponding to one class and constraints corresponding to the other class.

$$\begin{aligned} \min_{w_{(1)},b_{(1)}} \quad & \frac{1}{2}\| Aw_{(1)} + e_1 b_{(1)} \|^2 + c_1 e_2^T \lambda_2 \\ \text{s.t.} \quad & -(Bw_{(1)} + e_2 b_{(1)}) + \lambda_2 \ge e_2, \quad \lambda_2 \ge 0. \end{aligned} \tag{2}$$

and

$$\begin{aligned} \min_{w_{(2)},b_{(2)}} \quad & \frac{1}{2}\| Bw_{(2)} + e_2 b_{(2)} \|^2 + c_2 e_1^T \lambda_1 \\ \text{s.t.} \quad & (Aw_{(2)} + e_2 b_{(2)}) + \lambda_1 \ge e_1, \lambda_1 \ge 0. \end{aligned} \tag{3}$$

where c_1, $c_2 > 0$ are penalty parameters, e_1 and e_2 are vectors of ones of appropriate dimensions and λ_1 and λ_2 are vectors of slack variables respectively. Let $P = [B\ e_2]$ and $Q = [A\ e_1]$. The Wolf dual problems of (2) and (3) have been shown to be

$$\begin{aligned} \max_{\alpha} \quad & e_2^T \alpha - \frac{1}{2}\alpha^T P \left(Q^T Q\right)^{-1} P^T \alpha \\ \text{s.t.} \quad & 0 \le \alpha \le c_1 e_2, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \max_{\alpha} \quad & e_1^T \beta - \frac{1}{2}\beta^T Q \left(P^T P\right)^{-1} Q^T \beta \\ \text{s.t.} \quad & 0 \le \beta \le c_2 e_1, \end{aligned} \tag{5}$$

where Lagrangian multipliers are $\alpha \in R^{m_2}$ and $\beta \in R^{m_1}$. In order to deal with the case when $P^T P$ or $Q^T Q$ becomes singular and to avoid the possible ill-conditioning of $P^T P$ and $Q^T Q$, TWSVM introduces a term εI $(\varepsilon > 0)$ where I is an identity matrix of appropriate dimensions. The non-parallel hyperplanes (1) can be obtained from the solutions α and β of (4) and (5) by

$$z_1 = -\left(Q^T Q + \varepsilon I\right)^{-1} P^T \alpha \quad \text{and} \quad z_2 = -\left(P^T P + \varepsilon I\right)^{-1} Q^T \beta, \tag{6}$$

where $z_{(i)} = \left[w_{(i)}^T\ b_{(i)}\right]^T$, $(i = 1, 2)$.

A new point $x \in R^m$ is assigned to class i $(i = +1, -1)$, depending on which of the two hyperplanes in (1) is closer to, i.e.

$$Class(i) = \arg\min_{i=1,2} \frac{|x^T w_{(i)} + b_{(i)}|}{\| w_{(i)} \|} \tag{7}$$

where $|.|$ is the absolute value.

TWSVM was also extended in [11] to handle nonlinear kernels by considering two non-parallel kernel generated surfaces.

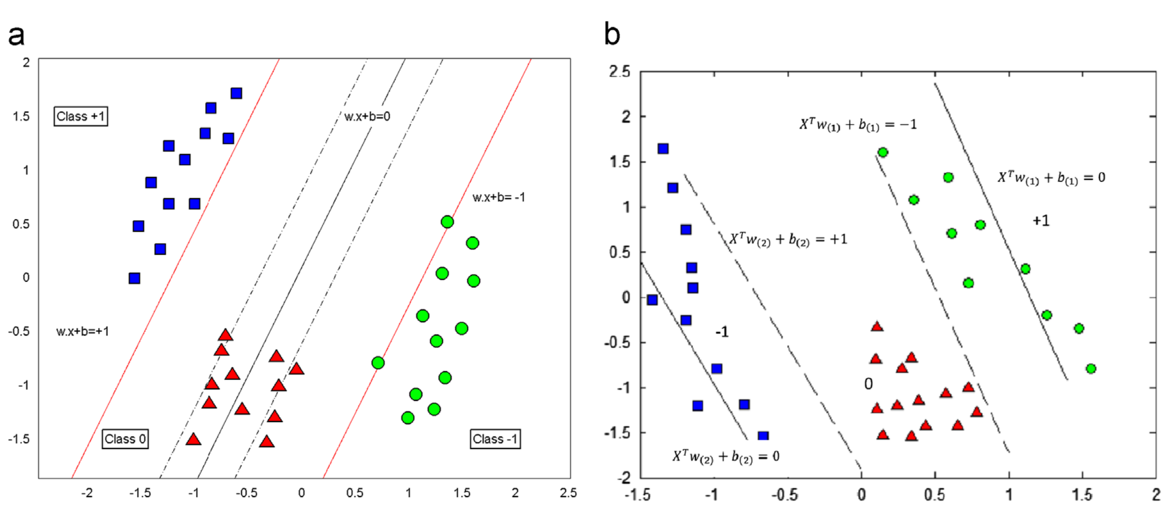

Fig. 1. Illustration of multi-class SVM and TWSVM with ternary output $\{-1, 0, +1\}$: (a) K-SVCR and (b) Twin-KSVC.