Neurocomputing 300 (2018) 34–43

Contents lists available at ScienceDirect

Neurocomputing

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

Twin support vector machines: A survey

Huajuan Huang[a], Xiuxi Wei[b], Yongquan Zhou[a,c,*]

[a] College of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China
[b] Information Engineering Department, Guangxi International Business Vocational College, Nanning 530007, China
[c] Key Laboratories of Guangxi High Schools Complex System and Computational Intelligence, Nanning, Guangxi 530006, China

ARTICLE INFO

Article history:
Received 12 August 2017
Revised 31 October 2017
Accepted 8 January 2018
Available online 9 March 2018

Keywords:
Support vector machine
Twin support vector machines
Non-parallel planes
Overview

ABSTRACT

Twin support vector machines (TWSVM) is a new machine learning method based on the theory of Support Vector Machine (SVM). Unlike SVM, TWSVM would generate two non-parallel planes, such that each plane is closer to one of the two classes and is as far as possible from the other. In TWSVM, a pair of smaller sized quadratic programming problems (QPPs) is solved, instead of solving a single large one in SVM, making the computational speed of TWSVM approximately 4 times faster than the standard SVM. At present, TWSVM has become one of the popular methods because of its excellent learning performance. In this paper, the research progress of TWSVM is reviewed. Firstly, it analyzes the basic theory of TWSVM, then tracking describes the research progress of TWSVM including the learning model and specific applications in recent years, finally points out the research and development prospects. This helps researchers to effectively use TWSVM as an emerging research approach, encouraging them to work further on performance improvement.

1. Introduction

Support Vector Machine (SVM) [1–7] is a computationally powerful kernel-based tool for binary data classification and regression as our known. Based on the theory of structural risk minimization principle, SVM has successfully solved the high dimensionality and local minimum problems. Therefore, compared with other machine learning methods, such as artificial neural networks [8–35], SVM has better generalization performance. So far, SVM has achieved excellent performance in many real-world predictive data mining applications such as text categorization, time series prediction, pattern recognition and image processing, etc [36–51]. Although SVM research has made a lot of remarkable achievements, it still has many deficiencies with in-depth study. For example, those problems including the relationship between statistical learning theory and other theoretical system, the processing of big data, the choice of parameters, the generalization ability of a given problem and the limitations of the applications are still not well resolved [52–56]. Especially, with the rapid development of the internet and information systems, high dimension, distributed and dynamic complex data are quickly generated. However, SVM has encountered great difficulties in dealing with these complex data.

In order to reduce the computational complexity of SVM, at present, many improved algorithms have been presented, such as chunking algorithm [57], decomposition algorithm [58] and sequential minimal optimization (SMO) [59], etc. Experimental results showed that these algorithms could improve the efficiency of SVM. However, the implementation of these algorithms is too complex. On the other hand, some deformation algorithms based on the standard SVM have been proposed in recent years. For example, in 2006, Mangasarian and Wild Edward [60] proposed a non-parallel plane classifier based on SVM, named generalized eigenvalue proximal support vector machine (GEPSVM). The essence of GEPSVM is to look for two nonparallel planes, demanding the data points of each class to be as close as possible to their own class and as far as possible from the other class. GEPSVM has good learning speed, but its classification accuracy is low. In 2007, Jayadeva and Suresh [61] proposed a new machine learning method called twin support vector machine (TWSVM) for the binary classification in the spirit of GEPSVM. TWSVM would generate two non-parallel planes, such that each plane is closer to one of the two classes and is as far as possible from the other. In TWSVM, a pair of smaller sized quadratic programming problems (QPPs) is solved, instead of solving a single large one in SVM, making the computational speed of TWSVM approximately 4 times faster than the traditional SVM. Because of its excellent performance, TWSVM has been applied to many areas such as speaker recognition, medical detection, etc [62,63]. TWSVM is a relatively new theory in the field of machine learning, which is not mature and perfect

* Corresponding author at: College of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China.
E-mail address: yongquanzhou@126.com (Y. Zhou).

https://doi.org/10.1016/j.neucom.2018.01.093
0925-2312/

[64–73,163]. Therefore, TWSVM needs further study and improvement. In this paper, we would review the latest research on TWSVM, analyze its theory and the algorithm ideas, then describe the latest progress of TWSVM in recent years and discuss the direction of future research.

This paper is organized as follows. Section 2 briefly discusses the basic theory of TWSM. In Section 3, we describe the latest progress of TWSVM. Section 4 gives summary and prospects of TWSVM.

2. Basic theory of TWSVM

Consider a binary classification problem of classifying m_1 data points belonging to class+1 and m_2 data points belonging to class −1. Then let matrix A in $R^{m_1\times n}$ represent the data points of class+1 while matrix B in $R^{m_2\times n}$ represent the data points of class −1. Two nonparallel hyper-planes of the linear TWSVM can be expressed as follows.

$$x^T w_1 + b_1 = 0 \quad \text{and} \quad x^T w_2 + b_2 = 0 \tag{1}$$

The target of TWSVM is to generate the above two nonparallel hyper-planes in the n-dimensional real space R^n, such that each plane is closer to one of the two classes and is as far as possible from the other. A new sample point is assigned to class+1 or −1 depending upon its proximity to the two nonparallel hyper-planes. The linear classifiers are obtained by solving the following optimization problems.

$$\begin{aligned}&\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\left\|Aw^{(1)}+e_1b^{(1)}\right\|^2+c_1e_2^T\xi^{(2)}\\ &s.t. \quad -\left(Bw^{(1)}+e_2b^{(1)}\right)\ge e_2-\xi^{(2)},\end{aligned} \tag{2}$$

$$\xi^{(2)}\ge 0.$$

$$\begin{aligned}&\min_{w^{(2)},b^{(2)},\xi^{(1)}} \frac{1}{2}\left\|Bw^{(2)}+e_2b^{(2)}\right\|^2+c_2e_1^T\xi^{(1)}\\ &s.t. \quad -(Aw^{(2)}+e_1b^{(2)})\ge e_1-\xi^{(1)},\\ &\xi^{(1)}\ge 0,\end{aligned} \tag{3}$$

where c_1 and c_2 are penalty parameters, $\xi^{(1)}$ and $\xi^{(2)}$are slack vectors, e_1 and e_2 are the vectors of ones of appropriate dimensions.

As with SVM, TWSVM solves the optimal solution is also the way to solve its dual problem in dual space. Introducing the Lagrangian multipliers α and β, (2) can be written as

$$\begin{aligned}L\left(w^{(1)},b^{(1)},\xi^{(2)},\alpha,\beta\right)&=\frac{1}{2}\left(Aw^{(1)}+e_1b^{(1)}\right)^T\left(Aw^{(1)}+e_1b^{(1)}\right)\\ &+c_1e_2^T\xi^{(2)}-\alpha^T\left(-\left(Bw^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}-e_2\right)-\beta^T\xi^{(2)},\end{aligned} \tag{4}$$

where $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_{m_2})^T$ and $\beta=(\beta_1,\beta_2,\cdots,\beta_{m_2})^T$ is the Lagrangian multiplier respectively. By KKT conditions, we can get

$$A^T\left(Aw^{(1)}+e_1b^{(1)}\right)+B^T\alpha=0 \tag{5}$$

$$e_1^T\left(Aw^{(1)}+e_1b^{(1)}\right)+e_2^T\alpha=0 \tag{6}$$

$$c_1e_2-\alpha-\beta=0 \tag{7}$$

$$-\left(Bw^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}\ge e_2,\ \xi^{(2)}\ge 0 \tag{8}$$

$$\alpha^T\left(-\left(Bw^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}-e_2\right)=0,\ \beta^T\xi^{(2)}=0 \tag{9}$$

$$\alpha\ge 0,\quad \beta\ge 0 \tag{10}$$

In view of $\beta\ge 0$, (7) can be express

$$0\le\alpha\le c_1 \tag{11}$$

Combining (5) and (6), we can get

$$\left[A^T \quad e_1^T\right][A \quad e_1]\left[w^{(1)} \quad b^{(1)}\right]^T+\left[B^T \quad e_2^T\right]\alpha=0 \tag{12}$$

Let us define

$$H=[A \quad e_1],\quad G=[B \quad e_2] \tag{13}$$

Let $u=[w^{(1)} \quad b^{(1)}]^T$, (12) can be express

$$H^THu+G^T\alpha=0 \tag{14}$$

In other words, there are

$$u=-\left(H^TH\right)^{-1}G^T\alpha \tag{15}$$

From the KKT condition and the formula (2), we can obtain the dual problem of (2) as follows.

$$\begin{aligned}&\max_{\alpha}\ e_2^T\alpha-\frac{1}{2}\alpha^TG(H^TH)^{-1}G^T\alpha\\ &s.t.\quad 0\le\alpha\le c_1\end{aligned} \tag{16}$$

Using the similar method, we can also get the duality of (3) as follows.

$$\begin{aligned}&\max_{\gamma}\ e_1^T\gamma-\frac{1}{2}\gamma^TP(Q^TQ)^{-1}P^T\gamma\\ &s.t.\quad 0\le\lambda\le c_2\end{aligned} \tag{17}$$

Let $P=[A \quad e_1]$, $Q=[B \quad e_2]$, $v=[w^{(2)} \quad b^{(2)}]^T$, we can get

$$v=-\left(Q^TQ\right)^{-1}P^T\gamma \tag{18}$$

Once u and v are determined, the TWSVM's two class hyper-planes can be determined.

For non-linear situations, introducing the kernel functions, the two hyper-planes of TWSVM based on kernel space can be expressed as

$$K\left(x^T,C^T\right)u_1+b_1=0,\ K\left(x^T,C^T\right)u_2+b_2=0, \tag{19}$$

where $C^T=[A,B]^T$. The nonlinear TWSVM optimization problem is expressed

$$\begin{aligned}&\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\left\|K(A,C^T)w^{(1)}+e_1b^{(1)}\right\|^2+c_1e_2^T\xi^{(2)}\\ &s.t.\quad -\left(K(B,C^T)w^{(1)}+e_2b^{(1)}\right)\ge e_2-\xi^{(2)},\\ &\xi^{(2)}\ge 0.\end{aligned} \tag{20}$$

$$\begin{aligned}&\min_{w^{(2)},b^{(2)},\xi^{(1)}} \frac{1}{2}\left\|K(B,C^T)w^{(2)}+e_2b^{(2)}\right\|^2+c_2e_1^T\xi^{(1)}\\ &s.t.\quad \left(K(A,C^T)w^{(2)}+e_1b^{(2)}\right)\ge e_1-\xi^{(1)},\\ &\xi^{(1)}\ge 0.\end{aligned} \tag{21}$$

Introducing the Lagrangian multipliers α and β, (20) can be written as

$$\begin{aligned}&L\left(w^{(1)},b^{(1)},\xi^{(2)},\alpha,\beta\right)\\ &=\frac{1}{2}\left(K(A^T,C^T)w^{(1)}+e_1b^{(1)}\right)^T\left(K(A^T,C^T)w^{(1)}+e_1b^{(1)}\right)\\ &+c_1e_2^T\xi^{(2)}-\alpha^T\left(-\left(K(B^T,C^T)w^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}-e_2\right)\\ &-\beta^T\xi^{(2)},\end{aligned} \tag{22}$$

where $\alpha=(\alpha_1,\alpha_2,\cdots,\alpha_{m_2})^T$ and $\beta=(\beta_1,\beta_2,\cdots,\beta_{m_2})^T$ is the Lagrangian multiplier respectively. By KKT conditions, we can get

$$K\left(A^T,C^T\right)^T\left(K(A^T,C^T)w^{(1)}+e_1b^{(1)}\right)+K\left(B^T,C^T\right)^T\alpha=0 \tag{23}$$

$$e_1^T\left(K\left(A^T,C^T\right)w^{(1)}+e_1b^{(1)}\right)+e_2^T\alpha=0 \tag{24}$$

$$c_1e_2-\alpha-\beta=0 \tag{25}$$

$$-\left(K\left(B^T,C^T\right)w^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}\geq e_2,\ \xi^{(2)}\geq 0 \tag{26}$$

$$\alpha^T\left(-\left(K(B^T,C^T)w^{(1)}+e_2b^{(1)}\right)+\xi^{(2)}-e_2\right)=0,\ \beta^T\xi^{(2)}=0 \tag{27}$$

$$\alpha\geq 0,\quad \beta\geq 0 \tag{28}$$

Combining (23) and (24), we can get

$$\left[K\left(A,C^T\right)^T\ \ e_1^T\right]\left[K\left(A,C^T\right)\ \ e_1\right]\left[w^{(1)}\ \ b^{(1)}\right]^T+[K(B,C^T)^T\ \ e_2^T]\alpha=0 \tag{29}$$

Let $S=[K(A,\ C^T)\ \ e_1]$, $R=[K(B,\ C^T)\ \ e_2]$, $z=[w^{(1)}\ \ b^{(1)}]^T$, (29) can be expressed

$$S^TSz+R^T\alpha=0 \tag{30}$$

$$z=-\left(S^TS\right)^{-1}R^T\alpha \tag{31}$$

From the KKT condition and the formula (20), we can obtain the dual problem of (20) as follows.

$$\max_{\alpha}\ e_2^T\alpha-\frac{1}{2}\alpha^TR(S^TS)^{-1}R^T\alpha \quad s.t.\ \ 0\leq\alpha\leq c_1 \tag{32}$$

Using the similar method, we can also get the duality of (21) as follows.

$$\max_{\gamma}\ e_1^T\gamma-\frac{1}{2}\gamma^TL(N^TN)^{-1}L^T\gamma \quad s.t.\ \ 0\leq\gamma\leq c_2, \tag{33}$$

where $L=[K(A,C^T)\ \ e_1]$, $N=[K(B,C^T)\ \ e_2]$.

Let $z_2=[w^{(2)}\ \ b^{(2)}]^T$, we can get

$$z_2=\left(N^TN\right)^{-1}L^T\gamma \tag{34}$$

Once z and z_2 are determined, the two hyper-surfaces of TWSVM with kernel classification are determined.

3. Theory research progress on TWSVM

As a new theory in the field of machine learning, TWSVM has become a hot research topic due to its obvious advantages. However, TWSVM is still insufficient, which still needs to be more improved. Up to now, many researchers have done a lot of improved research on TWSVM mainly from the perspective of improving the training speed, sparse performance and generalization performance of TWSVM. We would review the latest theory research progress on TWSVM as follows.

3.1. Research on least squares TWSVM

In order to improve the performance of TWSVM, Arun Kumar [74] formulated a least squares version of TWSVM called LSTWSVM. This formulation lead to extremely simple and fast algorithm for generating binary classifiers based on two non-parallel hyper-planes. In LSTWSVM, authors attempted to solve two modified primal problems of TWSVM, instead of two dual problems usually solved. The optimization problem of linear LSTWSVM is formulated as [79]:

$$\min_{w^{(1)},b^{(1)},\xi^{(2)}}\ \frac{1}{2}\left\|Aw^{(1)}+e_1b^{(1)}\right\|^2+\frac{c_1}{2}e_2^T\xi^{(2)} \quad s.t.\ \ -\left(Bw^{(1)}+e_2b^{(1)}\right)=e_2-\xi^{(2)}, \tag{35}$$

$$\min_{w^{(2)},b^{(2)},\xi^{(1)}}\ \frac{1}{2}\left\|Bw^{(2)}+e_2b^{(2)}\right\|^2+\frac{c_2}{2}e_1^T\xi^{(1)} \quad s.t.\ \ -(Aw^{(2)}+e_1b^{(2)})=e_1-\xi^{(1)}. \tag{36}$$

Eqs. (35) and (36) reduce to

$$\begin{bmatrix}w^{(1)}\\ b^{(1)}\end{bmatrix}=-\left(B^TB+\frac{1}{c_1}A^TA\right)^{-1}B^Te \tag{37}$$

and

$$\begin{bmatrix}w^{(2)}\\ b^{(2)}\end{bmatrix}=\left(A^TA+\frac{1}{c_2}B^TB\right)^{-1}A^Te \tag{38}$$

where $A=[X_1,e_1]$ and $B=[X_2,e_2]$.

LSTWSVM also uses kernel function to classify data samples which are non-linearly separable.

Computational results in [74] demonstrated the effectiveness of the proposed method. Unlike TWSVM, LSTWSVM is based on the regularization theory. This improvement makes the LSTWSVM solution greatly simplified, but also led to the LSTWSVM lost the robustness and anti-noise ability. To solve this problem, Chen and Ji [75] proposed a weighted version of least squares TWSVM called WLSTWSVM. In WLSTWSVM, different weights were put on the error variables in order to eliminate the impact of noise data and obtain the robust estimation. Compared to LSTWSVM, WLSTWSVM has better performance. In order to make better use of the information contained in prior knowledge, Arun Kumar [76] considered prior knowledge in the form of multiple polyhedral sets and incorporates the same into the formulation of LSTWSVM and term them as knowledge based LSTWSVM (KBLSTWSVM). Both of these formulations were capable of generating non-parallel hyperplanes based on real-world data and prior knowledge. Computational experiments showed that KBLSTWSVM was a versatile classifier whose solution was extremely simple compared with LSTWSVM. In order to make LSTWSVM suitable for manifold learning problems, a manifold LSTWSVM (MLSTWSVM) algorithm was proposed by Wang et al. [77–78] in 2010. First, a new and diverse manifold regularization term was added to the LSTWSVM optimization model. This simple transformation allowed two hyperplane objective functions to be directly transformed into two linear equations, and then the local information of the data was used as the first Know the new model of knowledge. This method was superior to TWSVM and LSTWSVM, and its computation time was faster than TWSVM but slower than LSTWSVM. While LSTWSVM showed a strong advantage in dealing with large-scale data sets, its efficiency would be further improved in dealing with high-dimensional data sets. In order to improve the efficiency of solving high dimensional problems, Gao et al. [79] introduced the Tikhonov regular term into the original problem of LSTWSVM and then transformed the two linear equations into two linear programming problems. Furthermore, the Newton method with fast convergence ability was used to solve the problem of external penalty in the linear programming dual problem. Thus, a 1-norm least squares twin support vector machine (NLSTWSVM) learning algorithm that can automatically select a sample feature was proposed. NLSTWSVM not only retains the advantages of LSTWSVM, but also reduces the dimension of the sample. The experimental results on the UCI dataset showed the feasibility and effectiveness of the NLSTWSVM, which was an effective method to deal with the large-scale and high-dimensional data classification problem.

Considering that the different positions of the samples had different effects on the topological structure of the classification hyper-plane, Ye et al. [80] proposed a density-weighted least squares twin support vector machines in 2012. In this method, LSTWSVM was weighted according to the density of the sample. That is, a region with a high density gave a higher weight, and a low density region gave a lower weight, which help to reduce the effect of noise on the hyper-plane. Experimental results showed that compared with TWSVM and LSTWSVM, the proposed algorithm could get a better hyper-plane and the classification accuracy was improved. In order to avoid the possibility of matrix singularity that can occur when solving the LSTWSVM objective function, Peng [81] modified the objective function of LSTWSVM to avoid the singularity of the matrix in the solution process, and proposed another version of least squares twin support vector machines, namely least squares twin support vector hyper-sphere (LS-TSVH). Theoretical analysis and experimental results showed that LS-TSVH could get higher classification accuracy in shorter time than LSTWSVM. In order to better solve the semi-supervised multi-category k-class classification problem, Khemchandani and Pal [82] formulated a Laplacian LSTWSVM called Lap-LST-KSVC in 2016. The proposed method evaluated all the training samples into "1-versus-1- versus-rest" classification paradigm, so as to generated ternary output {−1,0,+1}. Experimental results proved the efficacy of Lap-LST-KSVC over LSTWSVM in terms of classification accuracy and computational time. In order to better use LSTWSVM to solve human activity recognition, Khemchandani and Sharma [83] introduced a robust LSTWSVM, namely RLS-TWSVM. RLS-TWSVM handled the heteroscedastic noise and outliers present in activity recognition framework. Incremental RLS-TWSVM was proposed to speed up the training phase. Further, the authors introduced the hierarchical approach with RLS-TWSVM to deal with multi-category activity recognition problem. Experimental results showed that RLS-TWSVM was robust in order to handle heteroscedastic noise and outliers.

3.2. Research on projection twin support vector machine

In 2011, Chen et al. [84] proposed a recursive projection twin support vector machine (PTSVM) introducing the variance minimization method. The idea of PTSVM was to seek two projection directions, one for each class, such that the projected samples of one class were well separated from those of the other class in its respective subspace. The optimization problem of PTSVM is formulated as [84]:

$$\begin{aligned} \min_{w_1} \ & \frac{1}{2}\sum_{i=1}^{m_1}\left(w_1^T x_i^{(1)} - w_1^T \frac{1}{m_1}\sum_{j=1}^{m_1} x_j^{(1)}\right)^2 + c_1\sum_{k=1}^{m_2}\xi_k \\ s.t. \ & w_1^T x_k^{(2)} - w_1^T \frac{1}{m_1}\sum_{j=1}^{m_1} x_j^{(1)} + \xi_k \geq 0 \quad k = 1, 2, \cdots, m_2 \end{aligned} \tag{39}$$

and

$$\begin{aligned} \min_{w_1} \ & \frac{1}{2}\sum_{i=1}^{m_2}\left(w_2^T x_i^{(2)} - w_2^T \frac{1}{m_2}\sum_{j=1}^{m_2} x_j^{(2)}\right)^2 + c_2\sum_{k=1}^{m_1}\eta_k \\ s.t. \ & w_2^T x_k^{(1)} - w_2^T \frac{1}{m_2}\sum_{j=1}^{m_2} x_j^{(2)} + \xi_k \geq 0 \quad k = 1, 2, \cdots, m_1 \end{aligned} \tag{40}$$

where c_1 and c_2 are both trade-off constants, and ξ_k and η_k are all nonnegative slack variables. The authors gave a detailed explanation of the optimization problem (39). The first term in the objective function of Eq. (39) makes the variance of the projected samples of Class 1 as small as possible, that is, they are clustered around its mean. The second term reduces the hinge loss. The constraints in Eq. (39) require the projected samples of the other class, i.e. Class 2, to be at a distance of at least 1 from the projected center of Class 1. This term aims at maximizing the separation between the two classes. Slack variables ξ_k are used to measure the error when the constraints are not feasible. Eq. (40) is similar to Eq. (39). In order to further boost performance, a recursive algorithm for PTSVM was proposed to generate more than one projection axis for each class. At the same time, in order to avoid the problem of matrix singularity, the principal component analysis method was used to compress the dimension of the sample, so that the objective function of PTSVM was convex in low dimensional space. The experimental results show the effectiveness of the proposed algorithm. In 2012, Shao et al. [85] proposed least squares projection twin support vector machine (LSPTSVM) in order to improve the training speed of PTSVM. Different from PTSVM, a regularization term was added, which ensured the optimization problems in LSPTSVM were positive definite and resulted better generalization ability. In addition, the secondary loss function of PTSVM was replaced by the quadratic loss function, and the inequality constraint condition of PTSVM was modified to equality constraint. This clever modification could transform the quadratic programming problem of PTSVM into two linear equations. The experimental results showed that the training time of LSPTSVM was lower than that of PTSVM under considerable classification accuracy. In order to further improve the learning performance of PTSVM, Hua and Ding [86] proposed matrix pattern based projection twin support vector machines (MPPTSVM) introducing the idea of a matrix pattern. MPPTSVM had some advantages in improving the PTSVM generalization capability. In 2013, Hua and Ding [87] proposed a Recursive least squares projection twin support vector machines (RLSPTSVM) using recursive idea and least squares method. The combination of the two ideas makes the RLSPTSVM better than the PTSVM. Also in 2013, Shao et al. [88] proposed a regularized projection twin support vector machines (RPTSVM). In RPTSVM, the maximum interval regularization term was added to the objective function of the PTSVM. This simple change could make the objective function of the PTSVM be transformed from the principle of empirical risk minimization to the principle of structural risk minimization, avoiding the occurrence of learning problems and singularity problems. Furthermore, the successive super-relaxation technique and genetic algorithm were introduced as the parameter selection method of RPTSVM. Theoretical analysis and experimental results showed that RPTSVM had better classification accuracy and higher computational efficiency than PTSVM. In 2015, Hua and Ding [89] found that LSPTSVM failed to exploit the underlying correlation or similarity information between any pair of data points with the same labels that may be important for classification performance as much as possible. To mitigate this deficiency, the authors proposed a novel binary classifier based on LSPTSVM, namely weighted LSPTSVM with local information (LIWLSPTSVM). Because of measuring the importance of samples in the same class bi density weighting method, weight mean, instead of standard mean in LSPTSVM. The experimental results showed the effectiveness of LIWLSPTSVM. The above algorithms are generally applicable to solve the two classification problem. It is computationally expensive since it needs solving a series of quadratic programming problems when solving multi-class classification. To relieve the training burden, Yang et al. [90] proposed a novel multiple least squares recursive projection twin support vector machines (MLSPTSVM) for multi-class classification problem. For a K classes classification problem, MLSPTSVM aimed at seeking K groups of projection axes, one for each class that separated it from all the other. Due to solving a series of linear equations, MLSPTSVM tended to relatively simple and fast. Experimental results showed that MLSPTSVM had comparable classification accuracy while took significantly less computing time compared with RPTSVM for multi-class classification problem. Similarly, in order to solve multi-class classification problem, Li et al. [91] proposed a novel multiple projection twin support vector machine (Multi-PTSVM) using the similar idea of the above algorithm. The experimental results showed that this algorithm was effective for solving the multi-classification problem.

3.3. Research on model selection method of twin support vector machine

Similar to SVM, the model selection problem (including the selection of kernel functions, parameters and features) of TWSVM is also an urgent problem to be solved. The kernel function, which is an important part of the TWSVM model, determines the nonlinear processing capability of TWSVM and the quality of the constructed classification function. Therefore, the kernel function occupies the extremely important position in the TWSVM theory research, which is the key point of the mature development of TWSVM theory. Likewise, the selection of parameters and features has a crucial role in improving the training speed and classification accuracy of TWSVM. At present, some scholars have studied the TWSVM model selection problem. In 2009, Jayadeva et al. [92] treated the kernel selection problem for TWSVM as an optimization problem over the convex set of finitely many basic kernels, and formulated the same as an iterative alternating optimization problem. The experimental results showed that this method could get better classification performance compared with a single kernel function. In 2012, Yu et al. [93] proposed a twin support vector machine based on rough set in order to improve the training speed of TWSVM. In this paper, the author used the rough set to extract the feature of the data set and compress the dimension of the sample. The experimental results showed that this algorithm was effective especially for solving high-dimensional problems. In 2013, Ding et al. [94,95] respectively used particle swarm optimization (PSO) and quantum particle swarm optimization (QPSO) as the parameter optimization methods to optimize TWSVM. PSO and QPSO are the swarm intelligence optimization algorithms with strong robustness and good global searching ability. Compared with the classical TWSVM, experimental results show that the proposed methods have higher classification accuracy.

3.4. Research on twin parametric-margin support vector machine

In order to improve the ability of TWSVM to solve the heteroscedasticity of structural noise, in 2011, Peng [96] constructed a pair of nonparallel parametric interval hyper-graphs by referring to the idea of parametric interval support vector machine and proposed Twin Parametric-Margin Support Vector Machine (TPMSVM). The optimization problem of TPMSVM is formulated as [96]:

$$\begin{aligned} \min \ & \frac{1}{2}\|w_1\|_2^2 + \frac{v_1}{l_2} e_2^T (x_2 w_1 + e_2 b_1) + \frac{c_1}{l_1} e_1^T \xi \\ s.t. \ & x_1 w_1 + b_1 e_1 \geq 0 - \xi, \\ & \xi \geq 0, \end{aligned} \tag{41}$$

and

$$\begin{aligned} \min \ & \frac{1}{2}\|w_2\|_2^2 - \frac{v_2}{l_1} e_1^T (x_1 w_2 + e_1 b_2) + \frac{c_2}{l_2} e_2^T \eta \\ s.t. \ & x_2 w_2 + b_2 e_2 \leq 0 + \eta, \\ & \eta \geq 0, \end{aligned} \tag{42}$$

where c_1 and c_2 are the regularization parameters which determine the penalty weights, and ξ, η are the slack variables. The first term in the objective function of (41) is to control the model complexity for finding the positive parametric margin hyper-plane. The second term in the objective function of (41) is to minimize the sum of projection values of negative training points. Opimizing this term leads the negative training points to be as possible as far from the positive parametric-margin hyper-plane. The constraints require the projection values of positive training samples on the positive parametric-margin hyper-plane be not less than zeros. Otherwise, slack variable ξ is introduced to measure the error. Eq. (42) is similar to Eq. (41).

The experimental results showed that TPMSVM had a unique advantage in solving the data with heteroscedastic structural noise. In 2013, Shao et al. [97] introduced the least squares method and proposed Least Squares Twin Parametric-Margin Support Vector Machine (LSTPMSVM). In LSTPMSVM, the inequality constraints of TPMSVM were modified to equality constraints, and the TPMSVM quadratic programming problem was transformed into two linear equations. Furthermore, PSO was introduced as the parameter optimization method of LSTPMSVM. The experimental results showed the effectiveness and feasibility of LSTPMSVM. In order to make effective use of the prior structure information of the samples, in 2013, Peng et al. [98] proposed Structural Twin Parametric-Margin Support Vector Machine (STPMSVM). In STPMSVM, the granularity of each sample was clustered, and the structural information of each sample was excavated and trained according to the obtained structure information. The experimental results showed that the algorithm had good generalization performance. In 2013, Wang et al. [99] introduced a smooth technique and then proposed Smooth Twin Parametric-Margin Support Vector Machine (STPMSVM) using genetic algorithm as a parameter selection method. The introduction of smooth technology make the TPMSVM model can use the gradient algorithm with fast convergence ability to solve in the original space, which greatly improves its training speed. At the same time, At the same time, the use of genetic algorithms to optimize the TPMSVM four parameters make STPMSVM had better learning ability. TPMSVM obtained a significant performance, however, its decision function lost the sparsity. In order to overcome this problem, Peng et al. [100] proposed an improved TPMSVM, named centroid-based twin parametric-margin support vector machine (CTPSVM). The significant advantage of CTPSVM over TPMSVM was that its decision hyper-plane was sparse by optimizing simultaneously the projection values of the centroid points of two classes on its pair of nonparallel hyper-planes.

3.5. Research on v-twin support vector machines

TWSVM minimizes empirical risks of those training data samples which affect its generalization ability. In order to solve this problem, a v-twin support vector machine (v-TWSVM) was proposed by Peng [101] in 2010. This v-TWSVM introduced a pair of parameters (v) to control the bounds of the fractions of the support vectors and the error margins. The optimization problem of v-TWSVM is formulated as [101]:

$$\begin{aligned} \min_{w^{(1)}, b^{(1)}, \xi^{(2)}} \ & \frac{1}{2}\left\|Aw^{(1)} + e_1 b^{(1)}\right\|^2 - v_1 \rho_1 + \frac{1}{l_2}\xi^{(2)} \\ s.t. \ & -\left(Bw^{(1)} + e_2 b^{(1)}\right) + \xi^{(2)} \geq e_2 + \rho_1, \\ & \xi^{(2)} \geq 0, \rho_1 \geq 0. \end{aligned} \tag{43}$$

$$\begin{aligned} \min_{w^{(2)}, b^{(12}, \xi^{(1)}} \ & \frac{1}{2}\left\|bw^{(2)} + e_2 b^{(2)}\right\|^2 - v_2 \rho_2 + \frac{1}{l_1}\xi^{(1)} \\ s.t. \ & -\left(Aw^{(2)} + e_1 b^{(2)}\right) + \xi^{(1)} \geq e_1 + \rho_2, \\ & \xi^{(1)} \geq 0, \rho_2 \geq 0. \end{aligned} \tag{44}$$

Instead of c_1 and c_2 parameters, v-TWSVM uses two new parameters v_1 and v_2. It also introduces two additional factors ρ_1

and ρ_2 which are needed to be optimized. l_1 and l_2 are the size of positive and negative training sample, respectively. The theoretical analysis showed that this method can be interpreted as a pair of minimum generalized Mahalanobis-norm problems on two reduced convex hulls. In order to further improve the performance of v-TWSVM, Khemchandani et al. [102] proposed two novel binary classifiers termed as "Improvements on v-twin support vector machine: Iv-TWSVM and Iv-TWSVM (Fast)" that were motivated by v-TWSVM. The significant advantage of Iv-TWSVM over v-TWSVM was that Iv-TWSVM solved one smaller-sized Quadratic Programming Problem (QPP) and one Unconstrained Minimization Problem (UMP). Furthermore, Iv-TWSVM (Fast) avoided solving a smaller sized QPP and transformed it as a unimodal function, which can be solved using line search methods. In view of the effectiveness of v-TWSVM, Xu et al. [103] used this method to solve the universum data classification problems and obtained a good computation result.

3.6. Research on some other improved TWSVM

In 2015, Shao et al. [104] formulated a twin-type support vector machine for large-scale classification problems, called weighted linear loss twin support vector machine (WLTSVM). By introducing the weighted linear loss, WLTSVM only needed to solve simple linear equations with lower computational cost, and meanwhile, maintained the generalization ability. Therefore, it was able to deal with large-scale problems efficiently without any extra external optimizers. In TWSVM, the local information of samples is ignored, and then each sample shares the same weight when constructing the separating hyper-planes. Inspired by the studies above, Xu [105] proposed a K-nearest neighbor (KNN)-based weighted multi-class twin support vector machine (KWMTWSVM). Weight matrix W was employed in the objective function to exploit the local information of intra-class. Meanwhile, the weight vectors were introduced into the constraints to exploit the information of inter-class. Experimental results demonstrated the validity of this method. The traditional TWSVM adopted hinge loss which easily led to its sensitivity of the noise and instability for resampling. To solve this problem, in 2017, Xu et al. [106] presented a novel TWSVM with the pinball loss (Pin-TWSVM) which dealt with the quantile distance and was less sensitive to noise points. Then they further investigated its properties, including the noise insensitivity, between-class distance maximization, and within-class scatter minimization. Numerical experiments on a synthetic data set and 14 benchmark data sets with different noises demonstrated the feasibility and validity of the proposed method.

3.7. Research on twin support vector regression

As for regression problem, there are also many improved support vector regression (SVR) algorithms, such as Smooth SVR [107–118], LS-SVR [119–130], the SMO [131], etc. In 2010, Shevade et al. [132] introduced a new nonparallel plane regression in the spirit of TWSVM, termed as the twin support vector regression (TSVR). TSVR also aims at generating two nonparallel functions such that each function determines the ε-insensitive down- or up- bounds of the unknown regressor. Similar to TWSVM, TSVR only needs to solve a pair of smaller QPPs, instead of solving the large one in SVR. Furthermore, the number of constraints of each QPP in STVR is only half of the classical SVR, which makes TSVR work faster than SVR. Similar to SVR, TSVR solves the QPPs in the dual space. However, this solving method will be affected by time and memory constraints when dealing with the large datasets, which would make the learning speed of TSVR low. In order to improve the learning speed of TSVR, Peng [133] proposed a primal version for TSVR, termed primal TSVR (PTSVR). PTSVR directly optimizes the pair of QPPs of TSVR in the primal space based on a series of sets of linear equations by introducing a quadratic function to approximate its loss function. PTSVR can obviously improve the learning speed of TSVR without loss of the generalization. In 2011, Peng [134] proposed a reduced TSVR (RTSVR) that used the notion of rectangular kernels to obtain significant improvements in execution time over TSVR, thus facilitating its application to larger sized datasets. In 2012, Singh et al. [135] proposed a weighted TSVR, where samples in the different positions were proposed to give different penalties. The final regressor can avoid the over-fitting problem to a certain extent and yield great generalization ability. TSVR may incur suboptimal solution since it is objective function is positive semi-definite and the lack of complexity control. In order to address this shortcoming, Xu and Wang [136] developed a novel TSVR algorithm termed as smooth TSVR (STSVR). The idea is to adopt a smoothing technique to convert the original constrained quadratic programming problems into unconstrained minimization problems, and then use the well-known Newton-Armijo algorithm to solve the smooth TSVR. Experiments demonstrated its effectiveness. Aiming at the low approximation ability of sigmoid function of STSVR, using CHKS (Chen-Harker-Kanzow-Smale) function which has better approximation ability as the smooth function, a new version of smooth TSVR called smooth CHKS twin support vector regression (SCTSVR) model was proposed in [137]. A least squares version for TSVR (PLSTSVR) was also considered in the primal space [138]. However, least squares TSVR (LSTSVR) is not sparse, which would make its learning speed low. Huang et al. [139] proposed a novel nonparallel plane regressor, which can automatically select the relevant features. This method can suppress input features so that it can obtain comparable regression performance when using fewer computational time. Numerical experiments on artificial dataset and benchmark datasets demonstrated the feasibility and validity of the proposed algorithm. Huang et al. [140] improved TSVR by formulating it as a pair of linear programming problems. The use of 1-norm distance in the linear programming TSVR as opposed to the square of the 2-norm in the quadratic programming TSVR led to the better generalization performance and less computational time. Zhao et al. [141] proposed a new regressor termed as twin least squares support vector regression (TLSSVR). TLSSVR owns faster computational speed. Zhao et al. [142] proposed a new regressor, termed as ε-twin support vector regression (ε-TSVR). ε-TSVR determined a pair of ε-insensitive proximal functions by solving two related SVM-type problems. Experimental results showed this algorithm had remarkable improvement of generalization performance with short training time.

In this paper, we also mentioned that compared the performance on these algorithms. It is observed that the TPMSVM performs better in terms of the accuracy, but it takes more time as compared to the other existing approaches. So, improvement in speed and accuracy both may become an area for further research.

4. Application research progress on TWSVM

TWSVM was presented at a relatively short time, so there was less research on its application. Ganesh and Arjunan et al. [143–152] applied TWSVM to solve the biomedicine problem. In this paper, TWSVM was used to determine the accuracy of the TWSVM on the unbalanced data set generated by the subdivision of the sEMG surface electromyogram. The prediction accuracy was significantly higher than that of other popular algorithm. Speaker recognition means extracting the characteristics of personal identity from the speech signal and then identifying the identity of the speaker, but the content of the speech is invalid. Currently, GMM and SVM are frequently combined together for speaker recognition. This combined approach can be divided into three main methods [153–157]. One was to use GMM to get the kernel function in SVM, the second was using SVM as GMM preprocessing module, and the third was to use SVM as GMM post-processor, while mediating the results. Ding et al. [72] proposed that the TWSVM model applied to the speaker recognition method was different from the above three methods, which used the GMM model to extract the characteristic parameters as the input of the TWSVM model. This method can be used to train the insufficient sample and the large scale, which would give better robustness. These showed that the proposed new method can achieve better recognition performance in speaker recognition. Hua and Ding [73] applied TWSVM to image analysis and recognition in MCs detection, and then improved TWSVM by Boosting algorithm, and proposed Boosting-TWSVM [158,159]. The experimental results showed that the proposed method improved the detection rate and detection accuracy. Furthermore, they proposed BB-TWSVM [71] to further improve the detection accuracy. Ding et al. [160] applied TWSVM to intrusion detection. It not only to get faster detection speed and higher recognition accuracy, but also greatly reduce the time complexity of the algorithm. Yang et al. [161] also applied it to function estimation to solve the overfitting problem of SVR. In 2017, Gu et al. [162] proposed a L1-norm twin support vector machine (TPSVM-L1) for robust representation and recognition of images. The robustness of TPSVM-L1 was mainly driven by the L1-norm based distance metric that was proven to be robust to noise and outliers in data. Simulations on real image datasets verified the validity of TPSVM-L1.

5. Conclusions and prospects

At present, TWSVM has become one of the popular methods because of its excellent learning performance. Because TWSVM is a relatively new theory in the field of machine learning, it is not mature and perfect. such as, it is not sparse. At the same time, it has a lower generalization ability. Furthermore, its theory lacks practical application background. Even more, the frame of TWSVM learning theory has yet to be established. Therefore, TWSVM needs further study and improvement. Further research includes:

(1) How to construct the perfect kernel function for TWSVM. The kernel function, which is an important part of TWSVM, determines the level of non-linear processing ability. So the kernel function in the TWSVM occupies an extremely important position, which is the key to the mature development of TWSVM theory. So far, there is not yet a clear theoretical result to guide us on how to choose a good performance based on a specific data set kernel function. The selection and construction of kernel functions and the optimization of the corresponding parameters are still an open and urgent problem.

(2) How to better solve the sparseness of TWSVM. Lack of sparseness, it will reduce the accuracy of the algorithm. The lack of sparseness is a key problem in the study of TWSVM theory.

(3) How to better apply TWSVM to multi-classification. At present, the traditional TWSVM algorithm model is generally applied to two classification problems, how to extend the TWSVM algorithm to multi-classification problem, and achieve the desired computing efficiency, which is worthy of our study.

(4) How to extend the scope of TWSVM applications. At present, TWSVM research focused on the theoretical research. How to expand its practical application areas, will be the focus of the next step.

Acknowledgments

This work is supported by the National Natural Science Foundation under Grants nos. 61662005, 61463007, 61563008. The Science and Technology Research Project of Guangxi University under Grant no.KY2015YB076, the Talent Research Projects of Guangxi University for Nationalities under Grant no.2014MDQD018 and the open fund of Key Laboratory of Guangxi High Schools for Complex System & Computational Intelligence (No. 15CI01D).

References

[1] C. Cortes, V.N. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[2] H. Pei, Y. Chen, Y. Wu, et al., Laplacian total margin support vector machine based on within-class scatter, Knowl. Based Syst. 119 (2017) 152–165.
[3] Ch. Sudheer, R. Maheswaran, B.K. Panigrahi, et al., A hybrid SVM-PSO model for forecasting monthly streamflow, Neural Comput. Appl. 24 (6) (2014) 1381–1389.
[4] S.C. Huang, B. Chen, Automatic moving object extraction through a real-world variable-bandwidth network for traffic monitoring system, IEEE Trans. Ind. Electron. 61 (4) (2014) 2099–2112.
[5] J.X. Wu, Efficient HIK SVM learning for image classification, IEEE Trans. Image Process. 21 (10) (2012) 4442–4453.
[6] F. Zhu, N. Ye, W. Yu, Boundary detection and sample reduction for one-class support vector machine, Neurocomputing 123 (2014) 174–184.
[7] J.X. Du, D.S. Huang, X.F. Wang, X. Gu, Shape recognition based on neural networks trained by differential evolution algorithm, Neurocomputing 70 (4–6) (2007) 896–903.
[8] D.S. Huang, Systematic Theory of Neural Networks for Pattern Recognition, Publishing House of Electronic Industry of China, 1996.
[9] S.P. Deng, L. Zhu, D.S. Huang, Predicting hub genes associated with cervical cancer through gene co-expression networks, IEEE/ACM Trans. Comput. Biol. Bioinf. 13 (1) (2016) 27–35.
[10] D.S. Huang, Radial basis probabilistic neural networks: model and application, Int. J. Pattern Recognit. Artif. Intell. 13 (7) (1999) 1083–1101.
[11] D.S. Huang, J.X. Du, A constructive hybrid structure optimization methodology for radial basis probabilistic neural networks, IEEE Trans. Neural Netw. 19 (12) (2008) 2099–2115.
[12] D.S. Huang, A constructive approach for finding arbitrary roots of polynomials by neural networks, IEEE Trans. Neural Netw. 15 (2) (2004) 477–491 2004.
[13] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, et al., Detection and classification of single and combined power quality disturbances using neural networks, IEEE Trans. Ind. Electron. 61 (5) (2014) 2473–2482.
[14] D. Huang, H.H.S. Ip, K.C.K. Law, Z. Chi, Zeroing polynomials using modified constrained neural network approach, IEEE Trans. Neural Netw. 16 (3) (2005) 721–732.
[15] Z.Q. Zhao, D.S. Huang, Palmprint recognition with 2DPCA+PCA based on modular neural networks, Neurocomputing 71 (1–3) (2007) 448–454.
[16] Z.Q. Zhao, D.S. Huang, A mended hybrid learning algorithm for radial basis function neural networks to improve generalization capability, Appl. Math. Model. 31 (7) (2007) 1271–1281.
[17] J.X. Du, D.S. Huang, G. Zhang, Z. Wang, A novel full structure optimization algorithm for radial basis probabilistic neural networks, Neurocomputing 70 (1–3) (2006) 592–596.
[18] L. Shang, D.S. Huang, J. Du, C. Zheng, Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network, Neurocomputing 69 (13–15) (2006) 1782–1786.
[19] J.X. Mi, D.S. Huang, B. Wang, X. Zhu, The nearest-farthest subspace classification for face recognition, Neurocomputing 113 (2013) 241–250.
[20] C.Y. Lu, D.S. Huang, Optimized projections for sparse representation based classification, Neurocomputing 113 (2013) 213–219.
[21] Y. Zhao, D.S. Huang, Completed local binary count for rotation invariant texture classification, IEEE Trans. Image Process. 21 (10) (2012) 4492–4497.
[22] B. Li, C. Wang, D.S. Huang, Supervised feature extraction based on orthogonal discriminant projection, Neurocomputing 73 (1–3) (2009) 191–196.
[23] X.F. Wang, D.S. Huang, J. Du, H. Xu, Laurent Heutte, Classification of plant leaf images with complicated background, Appl. Math. Comput. 205 (2) (2008) 916–926.
[24] X.F. Wang, D.S. Huang, A novel multi-layer level set method for image segmentation, J. Univers. Comput. Sci. 14 (14) (2008) 2428–2452.
[25] B. Li, D.S. Huang, C. Wang, K. Liu, Feature extraction using constrained maximum variance mapping, Pattern Recognit. 41 (11) (2008) 3287–3294.
[26] D.S. Huang, J. Mi, A new constrained independent component analysis method, IEEE Trans. Neural Netw. 18 (5) (2007) 1532–1535.
[27] J.X. Du, D.S. Huang, X.F. Wang, X. Gu, Computer-aided plant species identification (CAPSI) based on leaf shape matching technique, Trans. Inst. Meas. Control 28 (3) (2006) 275–284.
[28] Z. Sun, D.S. Huang, C. Zheng, L. Shang, Optimal selection of time lags for temporal blind source separation based on genetic algorithm, Neurocomputing 69 (7–9) (2006) 884–887.
[29] C.H. Zheng, D.S. Huang, Z. Sun, M.R. Lyu, T. Lok, Nonnegative independent component analysis based on minimizing mutual information technique, Neurocomputing 69 (7–9) (2006) 878–883.
[30] L. Shang, D.S. Huang, C. Zheng, Z. Sun, Noise removal using a novel non-negative sparse coding shrinkage technique, Neurocomputing 69 (7–9) (2006) 874–877.

[31] X.Z. Xu, S.F. Ding, Z.Z. Shi, et al., A novel optimizing method for RBF neural network based on rough set and AP clustering algorithm, J. Zhejiang Univ. SCI. C. 13 (2) (2012) 131–138.
[32] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, et al., Detection and classification of single and combined power quality disturbances using neural networks, IEEE Trans. Ind. Electron. 61 (5) (2014) 2473–2482.
[33] H.G. Han, J.F. Qiao, Nonlinear model-predictive control for industrial processes: an application to wastewater treatment process, IEEE Trans. Ind. Electron. 61 (4) (2014) 1970–1982.
[34] H. Ozturk, M.E. Kutay, An artificial neural network model for virtual superpave asphalt mixture design, Int. J. Pavement Eng. 15 (2) (2014) 151–162.
[35] O. Ludwig, U. Nunes, et al., Eigenvalue decay: a new method for neural network regularization, Neurocomputing 124 (2014) 33–42.
[36] Z. Chen, Z. Qi, B. Wang, et al., Learning with label proportions based on nonparallel support vector machines, Knowl. Based Syst. 119 (2017) 126–141.
[37] D.S. Huang, The Study of Data Mining Methods for Gene Expression Profiles, Science Press China, 2009.
[38] X.F. Wang, D.S. Huang, H. Xu, An efficient local Chan–Vese model for image segmentation, Pattern Recognit. 43 (3) (2010) 603–618.
[39] B. Li, C. Zheng, D.S. Huang, Locally linear discriminant embedding: an efficient method for face recognition, Pattern Recognit. 41 (12) (2008) 3813–3821.
[40] W. Jiang, D.S. Huang, L. Shenghong, Random walk-based solution to triple level stochastic point location problem, IEEE Trans. Cybern. 46 (6) (2016) 1438–1451.
[41] Z.L. Sun, D.S. Huang, Y. Cheung, Extracting nonlinear features for multispectral images by FCMC and KPCA, Digit. Signal Process. 15 (4) (2005) 331–346.
[42] Z.Y. Chen, Z.P. Zhi, Distributed customer behavior prediction using multiplex data: a collaborative MK-SVM approach, Knowl. Based Syst. 35 (2012) 111–119.
[43] R. Moraes, J.F. Valiati, N. Gaviao, P. Wilson, Document-level sentiment classification: an empirical comparison between SVM and ANN, Expert Syst. Appl. 40 (2) (2013) 621–633.
[44] S.F. Ding, Y.Z. Han, J.Z. Yu, Y.X. Gu, A fast fuzzy support vector machine based on information granulation, Neural Comput. Appl. 23 (1) (2013) S139–S144.
[45] C.J. Lin, On the convergence of the decomposition method for support vector machine, IEEE Trans. Neural Netw. 12 (2001) 1288–1298.
[46] X.K. Zhang, S.F. Ding, Y. Xue, An improved multiplebirth support vector machine for pattern classification, Neurocomputing 225 (2017) 119–128.
[47] S.F. Ding, X.K. Zhang, Y.X. An, Y. Xue, Weighted linear loss multiple birth support vector machine based on information granulation for multi-classification, Pattern Recognit. 67 (2017) 32–46.
[48] S.F. Ding, Z.Z. Shi, D.C. Tao, B. An, Recent advances in support vector machines, Neurocomputing 211 (2016) 1–3.
[49] S.F. Ding, H.J. Huang, J.Z. Yu, H. Zhao, Research on the hybrid models of granular computing and support vector machine, Artif. Intell. Rev. 43 (4) (2015) 565–577.
[50] S.F. Ding, H.J. Huang, X.Z. Xu, J. Wang, Polynomial smooth twin support vector machines, Appl. Math. Inf. Sci. 8 (4) (2014) 2063–2071.
[51] S.F. Ding, H. Li, C.Y. Su, J.Z. Yu, F.X. Jin, Evolutionary artificial neural networks: a review, Artif. Intell. Rev. 39 (3) (2013) 251–260.
[52] H.H. Tsai, Y.S. Lai, S.C. Lo, A zero-watermark scheme with geometrical invariants using SVM and PSO against geometrical attacks for image protection, J. Syst. Softw. 86 (2) (2013) 335–348.
[53] H.J. Huang, S.F. Ding, Z.Z. Shi, Primal least squares twin support vector regression, J. Zhejiang Univ. – SCI. C 14 (9) (2013) 722–732.
[54] D.S. Huang, W. Jiang, A general CPL-AdS methodology for fixing dynamic parameters in dual environments, IEEE Trans. Syst. Man Cybern. B 42 (5) (2012) 1489–1500.
[55] X.F. Wang, D.S. Huang, A novel density-based clustering framework by using level set method, IEEE Trans. Knowl. Data Eng. 21 (11) (2009) 1515–1531.
[56] W. Jiang, D.S. Huang, S. Li, Random-walk based solution to triple level stochastic point location problem, IEEE Trans. Cybern. 46 (6) (2016) 1438–1451.
[57] B. Boser, I. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proceedings of Fifth Annual Workshop Computation on Learning Theory, ACM, Pittsburgh, PA, 1992.
[58] E. Osuna, R. Freund, An improved training algorithm for support vector machines, in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, IEEE Press, New York, 1997, pp. 276–285.
[59] S.S. Keerthi, S.K. Shevade, et al., Improvements to Platt's SMO Algorithm for SVM classifier design, Neural Comput. 13 (2001) 637–649.
[60] L. Mangasarian O, W. Wild Edward, Multi-surface proximal support vector machine classification via generalized eigenvalues, IEEE Trans. Pattern Anal. Mach. Intell. 28 (1) (2006) 69–74.
[61] K.R. Jayadeva, C. Suresh, Twin support vector machines for pattern classification, IEEE Trans. Pattern Anal. Mach. Intell. 29 (5) (2007) 905–910.
[62] Cong H.H, Yang C.F, Pu X.R. Efficient speaker recognition based on multi-class twin support vector machines and GMMs. Proceedings of the 2008 IEEE Conference on Robotics, Automation and Mechatronics, 2008, 348–352.
[63] X.S. Zhang, X.B. Gao, Y. Wang, Twin support vector machine for MCs detection, J. Electron. (China) 26 (3) (2009) 318–325.
[64] S.F. Ding, H.J. Huang, Z.Z. Shi, Weighted smooth CHKS twin support vector machines, J. Softw. 24 (11) (2013) 2548–2557.
[65] W.X. Bian, S.F. Ding, Y. Xue, Combining weighted linear project analysis with orientation diffusion for fingerprint orientation field reconstruction, Inf. Sci. 396 (2017) 55–71.
[66] H.J. Huang, S.F. Ding, F.L. Wu, Invasive weed optimization algorithm for optimization the parameters of mixed kernel twin support vector machines, J. Comput. 8 (8) (2013) 2077–2084.
[67] X.P. Hua, S.F. Ding, Weighted least squares projection twin support vector machines with local information, Neurocomputing 160 (2015) 228–237.
[68] S.F. Ding, X.P. Hua, Recursive least squares projection twin support vector machines, Neurocomputing 130 (2014) 3–9.
[69] S.F. Ding, X.K. Zhang, J.Z. Yu, Twin support vector machines based on fruit fly optimization algorithm, Int. J. Mach. Learn. Cybern. 7 (2) (2016) 193–203.
[70] X.K. Zhang, S.F. Ding, T.F. Sun, Multi-class LSTMSVM based on optimal directed acyclic graph and shuffled frog leaping algorithm, Int. J. Mach. Learn. Cybern. 7 (2) (2016) 241–251.
[71] S.F. Ding, H.J. Huang, Forecasting method of stock price based on polynomial smooth twin support vector regression, in: Proceedings of the Ninth International Conference on Intelligent Computing Theories, ICIC2013, 2013, pp. 96–105.
[72] S.F. Ding, Y.X. An, X.K. Zhang, Y. Xue, Wavelet twin support vector machines based on glowworm swarm optimization, Neurocomputing 225 (2017) 157–163.
[73] X.P. Hua, S.F. Ding, Locality preserving twin support vector machines, J. Comput. Res. Dev. 51 (3) (2014) 590–597.
[74] M. Arun Kumar, M. Gopal, Least squares twin support vector machines for pattern classification, Expert Syst. Appl. 36 (4) (2009) 7535–7543.
[75] J. Chen, G.G. Ji, Weighted least squares twin support vector machines for pattern classification, in: Proceedings of the Second International Conference on Computer and Automation Engineering, 2, Singapore, 2010, pp. 242–246.
[76] M. Arun Kumar, K. Reshma, M. Gopal, C. Suresh, Knowledge based least squares twin support vector machines, Inf. Sci. 180 (23) (2010) 4606–4618.
[77] D. Wang, L Ye Q, N. Ye, Localized multi-plane TWSVM classifier via manifold regularization, in: Proceedings of the Second International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2, 2010, pp. 70–73.
[78] D. Wang, N. Ye, Q.L. Ye, Twin support vector machines via fast generalized Newton refinement, in: Proceedings of the Second International Conference on Intelligent Human-Machine Systems and Cybernetics, 2, Nanjing Jiangsu, [s.n.], 2010, pp. 62–65.
[79] S.B. Gao, Q.L. Ye, N. Ye, 1-norm least squares twin support vector machines, Neurocomputing 74 (2011) 3590–3597.
[80] Q.L. Ye, N. Ye, S.B. Gao, Density-based weighting multi-surface least squares classification with its applications, Knowl. Inf. Syst. 33 (2012) 289–308.
[81] X.J. Peng, Least squares twin support vector hypersphere (LS-TSVH) for pattern recognition, Expert Syst. Appl. 37 (2010) 8371–8378.
[82] R. Khemchandani, A. Pal, Multi-category Laplacian least squares twin support vector machine, Appl. Intell. 45 (2) (2016) 458–474.
[83] R. Khemchandani, S. Sharma, Robust least squares twin support vector machine for human activity recognition, Appl. Soft Comput. 47 (2016) 33–46.
[84] X.B. Chen, J. Yang, Q.L. Ye, J. Liang, Recursive projection twin support vector machine via within-class variance minimization, Pattern Recognit. 44 (2011) 2643–2655.
[85] Y.H. Shao, N.Y. Deng, Z.M. Yang, Least squares recursive projection twin support vector machine for classification, Pattern Recognit. 45 (2012) 2299–2307.
[86] X.P. Hua, S.F. Ding, Matrix pattern based projection twin support vector machines, Int. J. Digit. Content Technol. Appl. 6 (20) (2012) 172–181.
[87] S.F. Ding, X.P. Hua, Recursive least squares projection twin support vector machines, Neurocomputing 130 (2014) 3–9.
[88] Y.H. Shao, Z. Wang, W.J. Chen, N.Y. Deng, A regularization for the projection twin support vector machine, Knowl. Based Syst. 37 (2013) 203–210.
[89] X.P. Hua, S.F. Ding, Weighted least squares projection twin support vector machines with local information, Neurocomputing 160 (2015) 228–237.
[90] Z.M. Yang, H.J. Wu, et al., Least squares recursive projection twin support vector machine for multi-class classification, Int. J. Mach. Learn. Cybern. 7 (3) (2016) 411–426.
[91] C.N. Li, Y.F. Huang, et al., Multiple recursive projection twin support vector machine for multi-class classification, Int. J. Mach. Learn. Cybern. 7 (3) (2016) 729–740.
[92] R.K. Jayadeva, Optimal kernel selection in twin support vector machines, J. Optim. Lett. 3 (2009) 77–88.
[93] J.Z. Yu, S.F. Ding, F.X. Jin, et al., Twin support vector machines based on rough sets, Int. J. Digit. Content Technol. Appl. 6 (20) (2012) 493–500.
[94] S.F. Ding, J.Z. Yu, H.J. Huang, et al., Twin support vector machines based on particle swarm optimization, J. Comput. 8 (9) (2013) 2296–2303.
[95] S.F. Ding, F.L. Wu, R. Nie, et al., Twin support vector machines based on quantum particle swarm optimization, J. Softw. 8 (7) (2013) 1743–1750.
[96] X.J. Peng, TPMSVM: a novel twin parametric-margin support vector machine for pattern recognition, Pattern Recognit. 44 (2011) 2678–2692.
[97] Y.H. Shao, Z. Wang, W.J. Chen, et al., Least squares twin parametric-margin support vector machine for classification, Appl. Intell. 39 (2013) 451–464.
[98] X.J. Peng, Y.F. Wang, D. Xu, Structural twin parametric-margin support vector machine for binary classification, Knowl. Based Syst. 49 (2013) 63–72.
[99] Z. Wang, Y.H. Shao, T.R. Wu, A GA-based model selection for smooth twin parametric-margin support vector machine, Pattern Recognit. 46 (2013) 2267–2277.
[100] X.J. Peng, L.Y. Kong, et al., Improvements on twin parametric-margin support vector machine, Neurocomputing 151 (2015) 857–863.
[101] X.J. Peng, A ν-twin support vector machine (ν-TSVM) classifier and its geometric algorithms, Inf. Sci. 180 (2010) 3863–3875.

[102] R. Khemchandani, P. Saigal, et al., Improvements on v-twin support vector machine, Neural Netw. 79 (2016) 97–107.
[103] Y.T. Xu, M. Chen, et al., V-twin support vector machine with Universum data for classification, Appl. Intell. 44 (4) (2016) 956–968.
[104] Y.H. Shao, W.J. Chen, et al., Weighted linear loss twin support vector machine for large-scale classification, Knowl. Based Syst. 73 (2015) 276–288.
[105] Y.T. Xu, K-nearest neighbor-based weighted multi-class twin support vector machine, Neurocomputing 205 (2016) 430–438.
[106] Y.T. Xu, Z.J. Yang, et al., A novel twin support vector machine with pinball loss, IEEE Trans. Neural Netw. Learn. Syst. 28 (2) (2017) 359–370.
[107] Y.J. Lee, W.F. Hsieh, C.M. Huang, ε-SSVR: a smooth support vector machine for ε-insensitive regression, IEEE Trans. Knowl. Data Eng. 17 (5) (2005) 678–685.
[108] S.F. Zheng, A fast algorithm for training support vector regression via smoothed primal function minimization, Int. J. Mach. Learn. Cybern. 6 (1) (2015) 155–166.
[109] W.C. Hong, Application of seasonal SVR with chaotic immune algorithm in traffic flow forecasting, Neural Comput. Appl. 21 (3) (2012) 583–593.
[110] Y. Mo, Applications of SVR to the Aveiro discretization method, Soft Comput. 19 (7) (2015) 1939–1951.
[111] K. Ucak, An adaptive support vector regressor controller for nonlinear systems, Soft Comput. 20 (7) (2016) 2531–2556.
[112] S. Balasundaram, Y. Meena, A new approach for training Lagrangian support vector regression, Knowl. Inf. Syst. 49 (3) (2016) 1097–1129.
[113] I.F. Chen, C.J. Lu, Sales forecasting by combining clustering and machine learning techniques for computer retailing, Neural Comput. Appl. 28 (9) (2017) 2633–2647.
[114] K.N. Wang, W.X. Zhu, P. Zhong, Robust support vector regression with generalized loss function and applications, Neural Process. Lett. 41 (1) (2015) 89–106.
[115] W.Y. Choi, D.H. Choi, K.J. Cha, Robust estimation of support vector regression via residual bootstrap adoption, J. Mech. Sci. Technol. 29 (1) (2015) 279–289.
[116] R.H. Laskar, K. Banerjee, et al., A pitch synchronous approach to design voice conversion system using source-filter correlation, Int. J. Speech Technol. 15 (3) (2012) 419–431.
[117] S. Balasundaram, R. Singh, On finite Newton method for support vector regression, Neural Comput. Appl. 19 (7) (2010) 967–977.
[118] Y.M. Guo, C.B. Ran, X.L. Li, J.Z. Ma, Adaptive online prediction method based on LS-SVR and its application in an electronic system, J. Zhejiang Univ. SCI. C 13 (12) (2012) 881–890.
[119] Q.L. Li, Y. Song, Z.G. Hou, Estimation of lower limb periodic motions from sEMG using least squares support vector regression, Neural Process. Lett. 41 (3) (2015) 371–388.
[120] H.W. Chen, W. Ser, Sound source DOA estimation and localization in noisy reverberant environments using least-squares support vector machines, J. Signal Process. Syst. 63 (3) (2011) 27–300.
[121] S.Q. Xu, Z. Liu, Y. Zhang, Least squares support vector regression and interval type-2 fuzzy density weight for scene denoising, Soft Comput. 20 (4) (2016) 1459–1470.
[122] X.N. Zhang, S.J. Song, J.B. Li, et al., Robust LS-SVM regression for ore grade estimation in a seafloor hydrothermal sulphide deposit, Acta Oceanol. Sin. 32 (8) (2013) 16–25.
[123] M. Ölmez, C. Güzeliş, Exploiting chaos in learning system identification for nonlinear state space models, Neural Process. Lett. 41 (1) (2015) 29–41.
[124] X.L. Zhang, T. Wu, Y. Shao, J. Song, Structure optimization of wheel force transducer based on natural frequency and comprehensive sensitivity, Chin. J. Mech. Eng. 30 (4) (2017) 973–981.
[125] L.J. Xie, Y.B. Ying, Use of near-infrared spectroscopy and least-squares support vector machine to determine quality change of tomato juice, J. Zhejiang Univ. SCI. B 10 (6) (2009) 465–471.
[126] X.J. Wen, Y.N. Zhang, W.W. Yan, et al., Nonlinear decoupling controller design based on least squares support vector regression, J. Zhejiang Univ. SCI. A 7 (2) (2006) 275–284.
[127] H.B. Huo, Y. Ji, X.J. Zhu, et al., Control-oriented dynamic identification modeling of a planar SOFC stack based on genetic algorithm-least squares support vector regression, J. Zhejiang Univ. SCI. A 15 (10) (2014) 829–839.
[128] Y.Q. Wang, Smooth nonparametric copula estimation with least squares support vector regression, Neural Process. Lett. 38 (1) (2013) 81–96.
[129] X.J. Zhou, T. Jiang, Enhancing least squares support vector regression with gradient information, Neural Process. Lett. 43 (1) (2016) 65–83.
[130] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.
[131] J.A.K. Suykens, L. Lukas, P. van Dooren, et al., Least squares support vector machine classifiers: a large scale algorithm, in: Proceedings of European Conference of Circuit Theory Design, 1999, pp. 839–842.
[132] S.K. Shevade, S.S. Keerthi, C. Bhattacharyya, et al., Improvements to the SMO algorithm for SVM regression, IEEE Trans. Netw. 11 (5) (2000) 1188–1193.
[133] X.J. Peng, TSVR: an efficient twin support vector machine for regression, Neural Netw. 23 (2010) 365–372.
[134] X.J. Peng, Primal twin support vector regression and its sparse approximation, Neurocomputing 73 (2010) 2846–2858.
[135] M. Singh, J. Chadha, P. Ahuja, et al., Reduced twin support vector regression, Neurocomputing 74 (2011) 1474–1477.
[136] Y.T. Xu, L.S. Wang, A weighted twin support vector regression, Knowl. Based Syst. 33 (2012) 92–101.
[137] X.B. Chen, J. Yang, J. Liang, et al., Smooth twin support vector regression, Neural Comput. Appl. 21 (2012) 505–513.
[138] H.J. Huang, S.F. Ding, Z.Z. Shi, Smooth CHKS twin support vector regression, J. Comput. Res. Dev. 52 (3) (2015) 561–568.
[139] H.J. Huang, S.F. Ding, Z.Z. Shi, Primal least squares twin support vector regression, J. Zhejiang Univ. SCI. C 14 (9) (2013) 722–732.
[140] H.J. Huang, X.X. Wei, Y.Q. Zhou, A sparse method for least squares twin support vector regression, Neurocomputing 211 (2016) 150–158.
[141] P. Zhong, Y.T. Xu, Y.H. Zhao, Training twin support vector regression via linear programming, Neural Comput. Appl. 21 (2012) 399–407.
[142] Y.P. Zhao, J. Zhao, M. Zhao, Twin least squares support vector regression, Neurocomputing 118 (2013) 225–236.
[143] Y.H. Shao, C.H. Zhang, Z.M. Yang, et al., An ε-twin support vector machine for regression, Neural Comput. Appl. 23 (2013) 175–185.
[144] S.P. Arjunan, D.K. Kumar, G.R. Naik, A machine learning based method for classification of fractal features of forearm sEMG using twin support vector machines, in: Proceedings of the IEEE Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), 2010, pp. 4821–4824.
[145] R.N. Ganesh, K.K. Dinesh, Jayadeva, Twin SVM for gesture classification using the surface electromyogram, IEEE Trans. Inf. Technol. Biomed. 14 (2) (2010) 301–308.
[146] M. Liu, Y. Xie, Z. Yao, et al., A new hybrid GMM/SVM for speaker verification, Int. Conf. Pattern Recognit. 4 (2006) 314–317.
[147] M.H. Liu, B.Q. Dai, Y.L. Xie, et al., Improved GMM-UBM/SVM for speaker verification, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, 2006, pp. 1925–1928.
[148] S. Fine, J. Navratil, R. Gopinath, A hybrid GMM/SVM approach to speaker identification, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, 2001, pp. 417–420.
[149] X.J. Ding, G.L. Zhang, Y.Z. Ke, B.L. Ma, Z.C. Li, High efficient intrusion detection methodology with twin support vector machines, in: Proceedings of the International Symposium on Information Science and Engieering, 1, 2008, pp. 560–564.
[150] S.F. Ding, J. Zhang, X.K. Zhang, Y.X. An, Survey on multi class twin support vector machines, J.Softw. 29 (1) (2018) 89–108 (Chinese).
[151] S.P. Arjunan, D.K. Kumar, G.R. Naik, A machine learning based method for classification of fractal features of forearm sEMG using twin support vector machines, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2010, pp. 4821–4824.
[152] R.N. Ganesh, K.K. Dinesh, Jayadeva, Twin SVM for gesture classification using the surface electromyogram, IEEE Trans. Inf. Technol. Biomed. 14 (2) (2010) 301–308.
[153] M. Liu, Y. Xie, Z. Yao, et al., A new hybrid GMM/SVM for speaker verification, Int. Conf. Pattern Recognit. 4 (2006) 314–317.
[154] M.H. Liu, B.Q. Dai, Y.L. Xie, et al., Improved GMM-UBM/SVM for speaker verification, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, 2006, pp. 1925–1928.
[155] S. Fine, J. Navratil, R. Gopinath, A hybrid GMM/SVM approach to speaker identification, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1, 2001, pp. 417–420.
[156] D.S. Huang, H.J. Yu, Normalized feature vectors: a novel alignment-free sequence comparison method based on the numbers of adjacent amino acids, IEEE/ACM Trans. Comput. Biol. Bioinf. 10 (2) (2013) 457–467.
[157] Z.L. Sun, D.S. Huang, Y.M. Cheung, J.M. Liu, G.B. Huang, Using FCMC, FVS and PCA techniques for feature extraction of multispectral images, IEEE Geosci. Remote Sens. Lett. 2 (2) (2005) 108–112.
[158] X.S. Zhang, Boosting twin support vector machine approach for MCs detection, Asia Pac. Conf. Inf. Process. 46 (2009) 149–152.
[159] X.S. Zhang, X.B. Gao, MCs detection approach using bagging and boosting based twin support vector machine, in: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TXUSA, 2009, pp. 5000–5005.
[160] X.J. Ding, G.L. Zhang, Y.Z. Ke, B.L. Ma, Z.C. Li, High efficient intrusion detection methodology with twin support vector machines, in: Proceedings of the International Symposium on Information Science and Engineering, 1, 2008, pp. 560–564.
[161] Yang C.F, Zhang Y, Lin Z. Function approximation based on twin support vector machines. Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems, 2008, 259–264.
[162] Z.F. Gu, Z. Zhang, et al., Robust image recognition by L1-norm twin-projection support vector machine, Neurocomputing 223 (2017) 1–11.
[163] S.F. Ding, J.Z. Yu, B.J. Qi, An overview on twin support vector machines, Artif. Intell. Rev. 42 (2) (2014) 245–252.

Huajuan Huang received the Ph.D. degree from School of Computer Science and Technology, China University of Mining and Technology in 2014. Now she is lecturer at College of Information Science and Engineering, Guangxi University for Nationalities. She has published more than 10 academic papers on machine learning and support vector machine. Her current research interests include twin support vector machines, data mining, pattern recognition, etc.

Xiuxi Wei received the M.S. degree from School of Computer and Electronic Information, Guangxi University in 2009. Now he is lecturer at Information Engineering Department, Guangxi International Business Vocational College. He has published more than 8 academic papers on artificial intelligence field. His current research interests include support vector machine, data mining, pattern recognition, etc.

Yongquan Zhou, Ph.D & Prof. He received the MS degree in computer science from Lanzhou University, Lanzhou, China, in 1993 and the Ph.D degree in Computation Intelligence from the Xiandian University, Xi'an, China, in 2006. He is currently a professor in Guangxi University for Nationalities. His research interests include computation intelligence, neural networks, and intelligence information processing et al. He has published 2 books, and more than 150 research papers in journals.