ICIC Express Letters ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **5**, May **2016** pp. **1087–1091**

FUZZY MARGIN TWIN SUPPORT VECTOR MACHINE WITH DUAL MEMBERSHIP

Qing Wu[1], Fei Xue[2] and Yanyan Wang[2]

[1]School of Automation
[2]School of Telecommunications and Information Engineering
Xi'an University of Posts and Telecommunications
Chang'an West St., Chang'an District, Xi'an 710121, P. R. China
xiyouwuq@126.com; happinessxf@qq.com

Received October 2015; accepted January 2016

Abstract. *Twin Support Vector Machine (TWSVM) has faster speed than traditional support vector machine, but without the structure risk minimization (SRM) and not considering the effect of different sample points for optimal classification plane. To solve this problem, a method based on fuzzy margin twin support vector machine with dual membership (DM-FMTWSVM) algorithm is presented. In order to weaken the impact of noise and isolated points of TWSVM classification, the fuzzy parameters and the regularization term are introduced in TWSVM. Therefore, it improves range and classification accuracy of twin support vector machine.*
Keywords: Fuzzy twin support vector machine, Dual membership, Margin, SRM

1. **Introduction.** Support Vector Machine (SVM) [1] proposed by Vapnik is a promising machine learning technique. It has shown powerful ability in machine learning on a small sample, nonlinearity and high dimension. Therefore, it is widely used in the field of text classification [2], image classification [3] and face recognition [4]. In 2007, Jayadeva et al. proposed twin support vector machine (TWSVM) [5] classifier for binary classification, motivated by multisurface proximal support vector machine classification via generalized eigenvalues (GEPSVM) [6]. TWSVM aims at generating two non-parallel hyperplanes by solving two smaller-sized quadratic programming problems (QPPs) compared with the classical SVM, such that each hyperplane is closer to one class and as far as possible from the other. A major advantage of TWSVM is that it is 4 times faster than SVM, since each of its QPPs is only roughly of half size. In recent years, some extensions to the TWSVM have been proposed such as ν-twin support vector machine (ν-TWSVM) [7], least squares twin support vector machine (LSTWSVM) [8], twin support vector hyphersphere (TSVH) [9].

However, these algorithms make it difficult to handle the effect of noise samples and fuzzy samples on the optimal hyperplanes. In this paper, a fuzzy margin twin support vector machine with dual membership (DM-FMTWSVM) is proposed. To handle the problem of outliers and noise, a dual fuzzy membership based on the density method [10,11] is constructed, so that different sample points have different effects on the classification of the surface. In addition, the empirical risk is minimized in the classical TWSVM. However, for our DM-FMTWSVM, the structural risk is minimized by adding a regularization term with the idea of maximizing some margin [12]. The computational comparison of the DM-FMTWSVM against traditional TWSVM in classification accuracy has been made on several UCI datasets for both linear and nonlinear kernels, which demonstrates the feasibility and validity of our proposed algorithm.

2. TWSVM. The basic thought of TWSVM is to construct a set of non-parallel hyperplanes in n-dimension real space R^n: $x^T\omega_1 + b_1 = 0$, $x^T\omega_2 + b_2 = 0$, where ω_1, $\omega_2 \in R^n$ and b_1, $b_2 \in R$ are determined. Consider a binary classification problem with l_1 samples belonging to class $+1$ and l_2 samples belonging to class -1 in the n-dimensional real space R^n. The positive samples and the negative samples are represented by matrix $A \in R^{l_1\times n}$ and $B \in R^{l_2\times n}$, respectively. Formally, TWSVM solves the following two QPPs for finding hyperplanes of the positive and negative, respectively:

$$\begin{array}{ll} \min\limits_{\omega_1,b_1,\xi_2} & \frac{1}{2}(A\omega_1 + e_1b_1)^T(A\omega_1 + e_1b_1) + c_1e_2^T\xi_2 \\ \text{s.t.} & -(B\omega_1 + e_2b_1) + \xi_2 \geq e_2, \quad \xi_2 \geq 0 \end{array}, \tag{1}$$

$$\begin{array}{ll} \min\limits_{\omega_2,b_2,\xi_1} & \frac{1}{2}(B\omega_2 + e_2b_2)^T(B\omega_2 + e_2b_2) + c_2e_1^T\xi_1 \\ \text{s.t.} & (A\omega_2 + e_1b_2) + \xi_1 \geq e_1, \quad \xi_1 \geq 0 \end{array}, \tag{2}$$

where $c_1 > 0$ and $c_2 > 0$ are penalty parameters, ξ_1 and ξ_2 are the slack variables. $\xi_i = \max\left(0, 1 - y_i(\omega^T x_i + b)\right)$ $(i = 1, 2)$ is the loss function. e_1 and e_2 are vectors of ones of appropriate dimensions.

3. DM-FMTWSVM. In TWSVM, the same penalties are given to the samples. In fact, they have different effects on the design of hyperplanes. In addition, noticing the primal problems of TWSVM, only the empirical risk is implemented. In order to overcome these two disadvantages, a dual fuzzy margin TWSVM is proposed based on structural risk minimization principle in this section.

DM-FMTWSVM also finds two nonparallel hyperplanes in R^n: $x^T\omega_1 + b_1 = 0$, $x^T\omega_2 + b_2 = 0$. The DM-FMTWSVM classifier is obtained by solving the following pair of QPPs:

DM-FMTSVM1

$$\begin{array}{ll} \min\limits_{\omega_1,b_1,\xi_2} & \frac{1}{2}(A\omega_1 + e_1b_1)^T(A\omega_1 + e_1b_1) + \frac{1}{2}\gamma_1\omega_1^T\omega_1 + c_1s_2e_2^T\xi_2 \\ \text{s.t.} & (B\omega_1 + e_2b_1) + \xi_2 \geq e_2, \quad \xi_2 \geq 0 \end{array}, \tag{3}$$

DM-FMTSVM2

$$\begin{array}{ll} \min\limits_{\omega_2,b_2,\xi_1} & \frac{1}{2}(B\omega_2 + e_2b_2)^T(B\omega_2 + e_2b_2) + \frac{1}{2}\gamma_2\omega_2^T\omega_2 + c_2s_1e_1^T\xi_1 \\ \text{s.t.} & (A\omega_2 + e_1b_2) + \xi_1 \geq e_1, \quad \xi_1 \geq 0 \end{array}, \tag{4}$$

where s_1, s_2 are the fuzzy membership values of sample sets A and B, respectively. c_1, c_2 are pre-specified penalty factors. The regularization terms $\frac{1}{2}\gamma_1\omega_1^T\omega_1$ and $\frac{1}{2}\gamma_2\omega_2^T\omega_2$ in the proximal problems (3) and (4) imply to maximize margin between the proximal hyperplane and the bounding hyperplane as much as possible. Minimization of the regularization terms of the problem in (3) and (4) is equivalent to maximizing the distance. Using the K. K. T conditions [13], the Wolfe dual of DM-FMTWSVM is expressed as follows:

$$\begin{array}{ll} \max\limits_{\alpha} & e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH + \gamma_1 E)^{-1}G^T\alpha \\ \text{s.t.} & 0 \leq \alpha \leq c_1s_2 \end{array}, \tag{5}$$

$$\begin{array}{ll} \max\limits_{\beta} & e_1^T\beta - \frac{1}{2}\beta^T H(G^TG + \gamma_2 E)^{-1}H^T\beta \\ \text{s.t.} & 0 \leq \beta \leq c_2s_1 \end{array}, \tag{6}$$

where $H = [A\ \ e_1]$, $G = [B\ \ e_2]$ and $E = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}$. $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{l_2})^T$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_{l_2})^T$ are the vectors of Lagrange multipliers. A data sample $x \in R^n$ is classified as class r $(r = 1, 2)$, depending on which one of the two hyperplanes it is closest to, that is $x^T\omega_r + b_r = \min_{l=1,2} \left|x^T w_l + b_l\right|$, where $|\cdot|$ is the perpendicular distance of point x from the plane: $x^T\omega_l + b_l = 0$ $(l = 1, 2)$.

DM-FMTWSVM described above can be extended to solve the nonlinear problems with kernel technique. Once the surfaces of the nonlinear DM-FMTWSVM are obtained, $x \in R^n$ is assigned to class $+1$ or class -1 in a manner similar to the linear case.

4. **Membership Function.** In this paper, a fuzzy membership function is constructed with a dual membership to overcome the problem of noise-sensitive TWSVM. That is, the sample points near by the class centers and the sample points far away from the class centers are given different memberships. The one class is determined by the distance between the point and its class center. The other is determined by the proportion between the number of its congeneric points and the number of its heterogeneous points in its neighborhood.

Let the training sample set for $T = \{(x_1, y_1, s_1), (x_2, y_2, s_2), \ldots, (x_l, y_l, s_l)\}$, where $x_i \in R^n$ and corresponding binary class labels $y_i \in \{-1, +1\}$. $s_i \in (0, 1]$ indicates the degree of x_i belongs to y_i and l is the number of the samples. The following are the definitions corresponding to the fuzzy membership function.

The distance between sample points is defined as $D(x_i, x_j) = \|x_i - x_j\|$. The density of sample points in the same class and the density of sample points in the other class are defined as

$$\rho^+(x_i, R) = M\left|\{x_j | D(x_j, x_i) \le R, y_j = y_i\}\right|,$$
$$\rho^-(x_i, R) = M\left|\{x_j | D(x_j, x_i) \le R, y_j \ne y_i\}\right|,$$

where $M|x|$ is the number of samples in the set, R is the radius of neighborhood of the sample points, and its size can be regulated.

It is defined the two class centers as follows $O^+ = \frac{\sum_{y_i=1} x_i}{l^+}$, $O^- = \frac{\sum_{y_i=-1} x_i}{l^-}$, where l^+ and l^- are the numbers of the samples in the positive class and the negative class, respectively.

Adjust to the radius of the neighborhood, according to the following formulas:

$$\rho(O^+) = \left|\left\{x_j | D(x_j, O^+) \le R^+, y_j = 1\right\}\right| = a\% l^+,$$
$$\rho(O^-) = \left|\left\{x_j | D(x_j, O^-) \le R^-, y_j = -1\right\}\right| = a\% l^-,$$

where $0 \le a \le 100$ can be regulated. Define the radius of the positive class and the negative class: R^+ and R^-, respectively. In this way the hypersphere composed of class center and the radius of the class, only contains $a\%$ class sample points. Hence, it greatly reduces the impact of noise and outliers on class radius.

The sample points are divided into two classifications depending on the distances between the points and their class centers. The membership functions of two sample points can be designed in two different methods.

According to the following formulas $D(x_i, O^+) \le m_1 R^+$, $y_i = +1$, $D(x_i, O^-) \le m_2 R^-$, $y_i = -1$, the membership of sample points near the center of the class is given as follows

$$s_i = \frac{1}{1 + \frac{cD(x_i, O^+)}{R^+}}, \quad y_i = +1, \quad s_i = \frac{1}{1 + \frac{cD(x_i, O^-)}{R^-}}, \quad y_i = -1, \tag{7}$$

where $0 < m_1 \le 1$, $0 < m_2 \le 1$ and $c > 0$ can be regulated. Then according to the following formulas $D(x_i, O^+) > m_1 R^+$, $y_i = +1$, $D(x_i, O^-) > m_2 R^-$, $y_i = -1$, the membership of sample points far from the center of the class is given as follows.

$$s_i = \frac{d\rho^+(x_i, R)}{\rho^+(x_i, R) + \rho^-(x_i, R)}, \tag{8}$$

where $0 < d \le 1$ and $R > 0$ can be regulated.

5. **Experiment.** To compare the performance of DM-FMTWSVM with SVM, TWSVM and DMFTWSVM (no margin DM-FMTWSVM), we perform some experiments on UCI data sets [14]. All algorithms are running on Personal Computer with 3 GHz and a maximum of 4 GB of the memory available. The computer runs Win7 with MATLAB 2011b. The Gaussian kernel function in the experiments is employed. Tenfold cross-validation is used to select all the optimal parameters over the range $\{2^i|i = -4, \ldots, 4\}$. The accuracy of the experiment is the average of the values of training set repeated 10 times and the CPU time is running one time with the optimum parameters.

In Table 1 and Table 2, it is easy to see that the accuracy of the proposed algorithm is better than TWSVM on all datasets, especially in nonlinear case. It is shown that DM-FMTWSVM has the best binary classification accuracy.

6. **Conclusions.** In this paper, DM-FMTWSVM is proposed for binary classification. It is more powerful than the traditional TWSVM by adding a dual membership and the regularization term in the primal problems of DM-FMTWSVM. This can achieve better

Table 1. Training set performance for linear classifiers

Data Set	Performance	SVM	TWSVM	DMFTWSVM	DM-FMTWSVM
Breast cancer	Accuracy (%)	68.83	75.68	75.75	**76.29**
200×9	Time (s)	1.21	0.34	0.43	0.37
German	Accuracy (%)	75.00	76.70	76.78	**76.82**
700×20	Time (s)	70.10	1.79	2.32	2.54
Thyroid	Accuracy (%)	89.33	93.18	94.67	**94.87**
140×5	Time (s)	1.50	0.19	0.22	0.21
Ionosphere	Accuracy (%)	86.30	88.97	88.77	**89.57**
351×34	Time (s)	13.92	0.31	0.48	0.47
Sonar	Accuracy (%)	75.89	76.50	77.50	**78.33**
208×60	Time (s)	3.66	0.18	0.25	0.25
Heart	Accuracy (%)	80.00	83.12	83.32	**83.63**
270×13	Time (s)	6.63	0.22	0.28	0.28
Bupa	Accuracy (%)	66.28	68.20	69.01	**69.22**
345×6	Time (s)	11.28	0.23	0.40	0.39

Table 2. Training set performance for nonlinear classifiers

Data Set	Performance	SVM	TWSVM	DMFTWSVM	DM-FMTWSVM
Breast cancer	Accuracy (%)	77.50	74.55	77.28	**78.00**
200×9	Time (s)	1.76	0.25	0.26	0.24
German	Accuracy (%)	77.57	76.49	77.88	**78.29**
700×20	Time (s)	98.40	1.59	2.12	2.06
Thyroid	Accuracy (%)	95.17	97.19	97.30	**97.30**
140×5	Time (s)	2.08	0.18	0.17	0.18
Ionosphere	Accuracy (%)	90.70	95.36	95.36	**96.21**
351×34	Time (s)	16.80	0.52	0.60	0.59
Sonar	Accuracy (%)	83.92	85.77	86.85	**87.84**
208×60	Time (s)	4.72	0.31	0.32	0.30
Heart	Accuracy (%)	82.70	77.45	86.92	**88.69**
270×13	Time (s)	7.69	0.38	0.41	0.36
Bupa	Accuracy (%)	68.44	64.91	71.52	**72.76**
345×6	Time (s)	12.42	0.63	0.52	0.55

performance on reducing the effects of outliers than some existing methods. Experimental results demonstrate the superiority of the approach over the classical TWSVM.

Acknowledgement. This work was supported in part by the Nature Science Foundation of China under Grant (61100165, 61100231, 61472307, 51205309), Natural Science Foundation of Shaanxi Province (2014JM8313), Foundation of Education Department of Shaanxi Province (2013JK1023) and New Star Team of Xi'an University of Posts and Telecommunications.

REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer Science & Business Media, 2000.
[2] X. L. Liu, S. F. Ding, H. Zhu et al., Appropriateness in applying SVMs to text classification, *Computer Engineering Science*, vol.32, no.6, pp.106-108, 2010.
[3] B. F. Guo, S. R. Gunn, R. I. Damper et al., Customizing kernel functions for SVM-based hyperspectral image classification, *IEEE Trans. Image Processing*, vol.17, no.4, pp.622-629, 2008.
[4] L. Zheng, Application of SVM in face recognition, *Computer Simulation*, vol.27, no.12, pp.271-273, 2010.
[5] Jayadeva, R. Khemchandani and S. Chandra, Twin support vector machines for pattern classification, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, no.5, pp.905-910, 2007.
[6] O. L. Mangasarian and E. W. Wild, Multisurface proximal support vector machine classification via generalized eigenvalues, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, no.1, pp.69-74, 2006.
[7] X. Peng, A v-twin support vector machine (v-TWSVM) classifier and its geometric algorithms, *Information Sciences*, vol.180, no.20, pp.3863-3875, 2010.
[8] M. A. Kumar and M. Gopal, Least squares twin support vector machines for pattern classification, *Expert Systems with Applications*, vol.36, no.4, pp.7535-7543, 2009.
[9] X. Peng, Twin support vector hypersphere (TSVH) classifier for pattern recognition, *Neural Computing and Application*, vol.24, pp.1207-1220, 2014.
[10] H. Zhang, Improved fuzzy support vector machine based on density clustering, *Computer Engineering*, vol.35, no.5, pp.194-196, 2009.
[11] M. H. Ha, Intuitionistic fuzzy support vector machine, *Journal of Hebei University (Natural Science Edition)*, vol.31, no.3, pp.225-229, 2011.
[12] Y. H. Shao and N. Y. Deng, A noval margin-based twin support vector machine with unity norm hyperplanes, *Neural Computer & Application*, vol.22, pp.1627-1635, 2013.
[13] O. L. Mangasarian, *Nonlinear Programming*, Philadelphia, PA, 1994.
[14] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, http://www.ics.uci.edu/~mlearn/MLRepository.html, Irvine, CA, University of California, 1998.