ICIC Express Letters
Volume **11**, Number **2**, February **2017**

ICIC International ©2017 ISSN 1881-803X
pp. 445–449

FUZZY TWIN SUPPORT VECTOR MACHINE BASED ON AFFINITY AMONG SAMPLES

Qing Wu[1,2,*], Feng Jiang[3] and Leyou Zhang[4]

[1]School of Automation
[3]School of Telecommunications and Information Engineering
Xi'an University of Posts and Telecommunications
No. 563, Chang'an South Road, Yanta District, Xi'an 710061, P. R. China
jfeng910915@163.com

[2]State Key Laboratory for Strength and Vibration of Mechanical Structures
Xi'an Jiaotong University
No. 28, Xianning West Road, Xi'an 710049, P. R. China
*Corresponding author: xiyouwuq@126.com

[4]School of Mathematics
Xidian University
No. 2, South Taibai Road, Xi'an 710071, P. R. China
lyzhang@mail.xidian.edu.cn

Received July 2016; accepted October 2016

Abstract. *Compared with traditional support vector machine (SVM), twin support vector machine (TWSVM) has faster speed. The same penalties are given to the samples in TWSVM. In fact, samples in the different positions have different effects on the bound function. Then, dual fuzzy parameters are introduced and a fuzzy twin support vector machine based on the affinity of dual membership (DM-AFTWSVM) is presented. Numerical experiments on UCI datasets demonstrate the classification accuracy of twin support vector machine is improved.*

Keywords: Fuzzy twin support vector machine, Bound function, Dual membership, Affinity

1. **Introduction.** SVM is very sensitive to outliers and noises in the training set [1,2]. In order to avoid the affection of outliers and noises, fuzzy support vector machine (FSVM) is proposed in [3]. In twin support vector machine (TWSVM), the same penalties are given to the samples. In fact, every sample has different effects on the design of hyperplanes.

In 2007, Jayadeva et al. proposed twin support vector machine (TWSVM) [4] classifier for binary classification, motivated by multisurface proximal support vector machine classification via generalized eigenvalues (GEPSVM) [5]. 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.

Motivated by FSVM, a dual fuzzy membership is constructed and a fuzzy affinity twin support vector machine with dual membership (DM-AFTWSVM) is proposed in this paper. Different sample points have different effects on the separating hyperplanes which can handle the problem of outliers and noise effectively [6,7]. Theoretical analysis and experimental results demonstrate 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 $\omega_1, \omega_2 \in R^n$ 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:

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

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

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

The two nonparallel hyperplanes in n-dimensional input space are as follows:

$$x^T\omega_1 + b_1 = 0, \tag{3}$$

$$x^T\omega_2 + b_2 = 0. \tag{4}$$

3. **DM-AFTWSVM.** In TWSVM, the same penalties are given to the samples. In fact, they have different effects on the design of hyperplanes [8,9]. DM-AFTWSVM also finds two nonparallel hyperplanes in R^n : $x^T\omega_1 + b_1 = 0$, $x^T\omega_2 + b_2 = 0$. The DM-AFTWSVM classifier is obtained by solving the following pair of QPPs:

DM-AFTWSVM 1

$$\begin{array}{ll} \min\limits_{\omega_1, b_1, \xi_2} & \frac{1}{2}(A\omega_1 + e_1b_1)^T(A\omega_1 + e_1b_1) + c_1s_2e_2^T\xi_2 \\ \text{s.t.} & -(B\omega_1 + e_2b_1) + \xi_2 \geq e_2, \;\; \xi_2 \geq 0 \end{array}, \tag{5}$$

DM-AFTWSVM 2

$$\begin{array}{ll} \min\limits_{\omega_2, b_2, \xi_1} & \frac{1}{2}(B\omega_2 + e_2b_2)^T(B\omega_2 + e_2b_2) + c_2s_1e_1^T\xi_1 \\ \text{s.t.} & (A\omega_2 + e_1b_2) + \xi_1 \geq e_1, \;\; \xi_1 \geq 0 \end{array}, \tag{6}$$

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 Lagrangian based on the optimization problem (5) is given by

$$\begin{aligned} L(\omega_1, b_1, \xi_2, \alpha, \beta) = & \frac{1}{2}(A\omega_1 + e_1b_1)^T(A\omega_1 + e_1b_1) + c_1s_2e_2^T\xi_2 \\ & + \alpha^T((B\omega_1 + e_2b_1)^T - \xi_2 + e_2) - \beta^T\xi_2. \end{aligned} \tag{7}$$

By setting the gradient of (7) with respect to ω_1, b_1 and ξ_2 equal to zero, we obtain

$$\nabla_{\omega_1} L = A^T(A\omega_1 + e_1b_1) + \beta^T\alpha = 0, \tag{8}$$

$$\nabla_{b_1} L = e_1^T(A\omega_1 + e_1b_1) + e_2^T\alpha = 0, \tag{9}$$

$$\nabla_{\xi_2} L = c_1s_2e_2 - \alpha - \beta = 0. \tag{10}$$

The Karush-Kuhn-Tucker (K.K.T) necessary and sufficient optimality conditions [14] for DM-AFTWSVM 1 are given by

$$\alpha^T((B\omega_1 + e_2b_1)^T - \xi_2 + e_2) = 0, \tag{11}$$

$$\beta^T\xi_2 = 0, \tag{12}$$

$$\alpha \geq 0, \; \beta \geq 0. \tag{13}$$

Because of the inequation $\beta \geq 0$, according to (10) we can obtain

$$0 \leq \alpha \leq c_1 s_2. \tag{14}$$

Using (7) and the K.K.T conditions (8)-(12), the Wolfe dual of DM-AFTWSVM 1 is expressed as follows:

$$\begin{array}{ll} \underset{\alpha}{\text{Max}} & e_2^T \alpha - \dfrac{1}{2} \alpha^T G (H^T H)^{-1} G^T \alpha \\ \text{s.t.} & 0 \leq \alpha \leq c_1 s_2 \end{array}. \tag{15}$$

Similarly, we can obtain the Wolfe dual of DM-AFTWSVM 2:

$$\begin{array}{ll} \underset{\beta}{\text{Max}} & e_1^T \beta - \dfrac{1}{2} \beta^T H (G^T G)^{-1} H^T \beta \\ \text{s.t.} & 0 \leq \beta \leq c_2 s_1 \end{array}, \tag{16}$$

where $H = [A \ \ e_1]$, $G = [B \ \ e_2]$, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{l_2})^T$ 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, $x^T \omega_r + b_r = \min\limits_{l=1,2} \left| x^T \omega_l + b_l \right|$, where $|\cdot|$ is the perpendicular distance of point x from the plane: $x^T \omega_l + b_l = 0$ $(l = 1, 2)$.

DM-AFTWSVM described above can be extended to solve the nonlinear problems with kernel technique [10]. Once the surfaces of the nonlinear DM-AFTWSVM 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 with a dual membership [11] is constructed to overcome the noise sensitivity of TWSVM. That is to say, the sample points near by the class centers and the sample points far away from the class centers are given different memberships [12]. The method uses an effective membership function [13,14]. The sample mean of positive class is used as the center of positive class, denoted by $x_+ = \frac{1}{l_1} \sum_{i=1}^{l_1} x_i$ $(i = 1, 2, \ldots, l_1)$. And the sample mean of negative class is used as the center of negative class, denoted by $x_- = \frac{1}{l_2} \sum_{i=1}^{l_2} x_i$ $(i = 1, 2, \ldots, l_2)$, where l_1 and l_2 are the numbers of the samples in the positive class and the negative class, respectively. Define the positive class radius and negative class one respectively as

$$R^+ = \max |x_i - x_+|, \tag{17}$$

$$R^- = \max |x_i - x_-|. \tag{18}$$

The distance between the positive class center and the negative class center is $T = |x_+ - x_-|$. The distance between each positive sample to the positive class center is $D^+ = |x_+ - x_i|$ and the distance between each negative sample to the negative class center is $D^- = |x_- - x_i|$. In addition, ε is a control factor of radius that satisfies $\varepsilon > 0$, $T * \varepsilon < R^+$ and $T * \varepsilon < R^-$. The membership function is defined as

$$s_1 = \begin{cases} (\delta + D_i^+)/R^+, & D_i^+ \leq T * \varepsilon \\ \delta, & D_i^+ > T * \varepsilon \end{cases}, \tag{19}$$

$$s_2 = \begin{cases} (\delta + D_i^-)/R^-, & D_i^- \leq T * \varepsilon \\ \delta, & D_i^- > T * \varepsilon \end{cases}. \tag{20}$$

where δ is a very small positive number given in advance as the membership of the noises and outliers which can ensure $s_1 > 0$ and $s_2 > 0$.

5. **Experiment.** To compare the performance of DM-AFTWSVM with SVM, TWSVM and FSVM, we perform some experiments on UCI data sets [15]. All algorithms are running on Personal Computer with 2.5 GHz and a maximum of 4 GB of the memory available. The computer runs Win7 with MATLAB 2012a. The Gaussian kernel function in the experiments is employed. Tenfold cross-validation is used to select c_1 and c_2 to find the optimal parameters over the range $\{2^i|i = -4, \ldots, 4\}$. We also use tenfold cross-validation to find the optimal ε over the range $(0, 4]$. The accuracy of the experiment on the average values of training set is 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 classification accuracy and efficiency of the proposed algorithm are better than SVM and FSVM on all UCI datasets, especially in nonlinear case. Compared with TWSVM, the classification accuracy of the proposed algorithm is improved obviously while the CPU time is slightly increasing.

Table 1. Results of linear DM-AFTWSVM on UCI data sets

Data Set	Performance	SVM	FSVM	TWSVM	DM-AFTWSVM
Breast cancer	Accuracy (%)	68.83	73.50	75.68	**76.62**
200×9	Time (s)	1.21	2.00	0.34	0.46
German	Accuracy (%)	75.00	76.13	76.70	**76.74**
700×20	Time (s)	70.10	75.00	1.79	2.14
Thyroid	Accuracy (%)	89.33	93.28	93.18	**94.55**
140×5	Time (s)	1.50	1.57	0.19	0.28
Ionosphere	Accuracy (%)	86.30	86.97	88.97	**90.03**
351×34	Time (s)	13.92	19.03	0.31	0.61
Sonar	Accuracy (%)	75.89	76.36	76.50	**78.03**
208×60	Time (s)	3.66	4.58	0.18	0.44
Heart	Accuracy (%)	80.00	80.47	83.12	**84.81**
270×13	Time (s)	6.63	11.70	0.22	0.57
Bupa	Accuracy (%)	66.28	67.48	68.20	**68.97**
345×6	Time (s)	11.28	23.02	0.23	0.28

Table 2. Results of nonlinear DM-AFTWSVM on UCI data sets

Data Set	Performance	SVM	FSVM	TWSVM	DM-AFTWSVM
Breast cancer	Accuracy (%)	76.80	66.23	74.55	**77.59**
200×9	Time (s)	1.76	3.21	0.25	0.48
German	Accuracy (%)	77.32	72.00	76.49	**77.86**
700×20	Time (s)	83.40	88.92	1.59	2.26
Thyroid	Accuracy (%)	95.17	90.67	97.19	**97.35**
140×5	Time (s)	2.08	4.02	0.18	0.29
Ionosphere	Accuracy (%)	90.70	93.11	95.36	**95.42**
351×34	Time (s)	16.80	22.17	0.52	0.67
Sonar	Accuracy (%)	83.92	81.09	85.77	**87.97**
208×60	Time (s)	4.72	7.24	0.31	0.48
Heart	Accuracy (%)	82.70	83.02	77.45	**84.20**
270×13	Time (s)	7.69	12.03	0.38	0.66
Bupa	Accuracy (%)	68.44	66.17	64.91	**74.69**
345×6	Time (s)	12.42	23.29	0.63	0.45

6. **Conclusions.** In this paper, DM-AFTWSVM is proposed for binary classification. Fuzzy memberships are assigned to each training sample to indicate its membership degree to different classes. We also use tenfold cross-validation to determine the parameters of membership function. We demonstrate that the superiority of our method over the conventional TWSM. Therefore, how to construct a new membership to further improve the classification performance of the algorithm is our future work.

Acknowledgements. This work was supported in part by the Natural Science Foundation of China under Grants (61100165, 61100231, 61472307, 61562001), Natural Science Basic Research Plan in Shaanxi Province of China (Program 2014JM8313, 2016JM6004), Academics Propulsion Technology Transfer Projects of Xi'an Science and Technology Bureau (CXY1516(6)), New Star Team of Xi'an University of Posts and Telecommunications and Open Fund of State Key Laboratory for Strength and Vibration of Mechanical Structures (SV2015-KF-04).

REFERENCES

[1] V. N. Vapnik, *Statistical Learning Theory*, Springer-Verlag, Wiley, New York, 1998.
[2] N. Deng and Y. Tian, *A New Method of Data Mining: SVM*, Science Press, Beijing, 2004.
[3] C. Lin and S. Wan, Fuzzy support vector machines, *IEEE Trans. Neural Networks*, vol.13, no.2, pp.464-471, 2002.
[4] 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.
[5] 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.
[6] M. Ha et al., Intuitionistic fuzzy support vector machine, *Journal of Hebei University (Natural Science Edition)*, vol.21, no.3, pp.225-229, 2011.
[7] J. Chang and P. Hao, A new kernel based fuzzy clustering approach: Support vector clustering with cell growing, *IEEE Trans. Fuzzy Systems*, vol.11, no.4, pp.518-527, 2003.
[8] Y. Mao, Z. Xia, Z. Yin et al., Fault diagnosis based on fuzzy support vector machine with parameter tuning and feature selection, *Chinese Journal of Chemical Engineering*, vol.15, no.2, pp.233-239, 2009.
[9] H. Tang, Y. Liao, F. Sun et al., Fuzzy support vector machine with a new fuzzy membership function, *Journal of Xi'an Jiaotong University*, vol.43, no.7, pp.40-43, 2010.
[10] S. Ding and Y. Gu, A fuzzy support vector machine algorithm with dual membership based on hypersphere, *Journal of Computational Information Systems*, vol.7, no.6, pp.2028-2034, 2011.
[11] Q. Ye, C. Zhao, S. Gao and H. Zheng, Weighted twin support vector machines with local information and its application, *Neural Networks*, vol.35, no.3, pp.31-39, 2012.
[12] Y. Shao, W. Chen, J. Zhang et al., An efficient weighted lagrangian twin support vector machine for imbalanced data classification, *Pattern Recognition*, vol.47, no.1, pp.3158-3167, 2014.
[13] Q. Wu, S. Liu and Z. Du, Fuzzy support vector machine method based on border vector extraction, *Pattern Recognition and Artificial Intelligence*, vol.21, no.3, pp.332-337, 2008 (in Chinese).
[14] O. L. Mangasarian, *Nonlinear Programming*, SIAM, Philadelphia, PA, 1994.
[15] C. L. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, http://www.ics.uci.edu/∼ mlearn/MLRepository.html, 1998.