

# A Solving Algorithm of Fuzzy Support Vector Machines Based on Determination of Membership

Hua Duan, Yanmei Hou

College of Information Science and Technology Shandong University of Science and Technology  
QingDao, China  
huaduan59@163.com

**Abstract**—In order to overcome the issues that Support Vector Machine is sensitive to the outlier and noise points, Fuzzy Support Vector Machine (FSVM) is proposed. The key issue to solve the FSVM is determinate the fuzzy membership. This paper gives an overview of construction algorithm of the fuzzy membership. We also give an algorithm to solve FSVM that is derived from improved-SMO algorithm.

**Keywords**-Support Vector Machine; Fuzzy Support Vector Machine; the fuzzy membership; SMO

## I. INTRODUCTION

The theory of support vector machines<sup>[1]</sup> (SVMs) is a new classification technique, which is based on the idea of VC dimension and structural risk minimization of statistical learning theory. SVM exhibits many unique advantages in resolving the small samples, nonlinear and high dimensional pattern recognition problem. Because the SVM is mainly dependent on the support vectors, it is very sensitive to noise points and isolated points of the SVM training samples. Due to those reasons, SVM is unstable and have a poor tolerance to noise. To solve the problem, the Fuzzy Support Vector Machines<sup>[2]</sup> (FSVM) was proposed in 2002, which have used in many fields<sup>[9-11]</sup>. The method of FSVM applies a fuzzy membership to each input point so that different input points can make different contribution to the learning of decision surface. We use the fuzzy membership to measure the importance of training points, so one of the main problems of FSVM is how to find the suitable membership. In this paper, we summarize some main algorithms on membership and we also give an algorithm to solve FSVM.

## II. FUZZY SUPPORT VECTOR MACHINES THEORY

In order to more clearly understand the FSVM, we briefly review the basis of the theory of SVM and fuzzy set.

### A. Support Vector Machines

Given a training set:

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times Y)^l \quad (1)$$

each training point  $x_i \in R^n$  belongs to either of two classes and is given a label  $y_i \in \{1, -1\}$ , for  $i = 1, \dots, l$ . We call the points that  $y_i = 1$  as a positive class point and  $y_i = -1$  as a negative class point. According to the statistical learning theory, SVM finds a suitable hyperplane to make the space between the positive class and the negative class be the

largest. The SVM can be transformed into solving the following optimization problem:

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \\ \text{s.t.} \quad & y_i((\omega \cdot x_i) + b) \geq 1 - \xi_i, i = 1, \dots, l, \\ & \xi_i \geq 0, i = 1, \dots, l \end{aligned} \quad (2)$$

where  $\xi = (\xi_1, \dots, \xi_l)^T$  is a slack variable, which reflects the misclassification scale of the training set. The penalty parameter  $C$  represents the misclassification scale and the classification margin.

### B. Fuzzy Set

Definition 1 (Fuzzy Set): For a fuzzy subset  $A$  on the domain  $X$ , a mapping from  $X$  to  $[0,1]$  such that

$$\mu_A : X \rightarrow [0,1] \quad (3)$$

where  $\mu_A$  is the membership function of set  $A$ , and the degree that element  $x$  belongs to set  $A$  is called the membership of  $x$ .

We use the following method to express fuzzy set:

$$A = \{(x, \mu_A(x)) \mid x \in A\} \quad (4)$$

where  $\mu_A(x)$  is the membership of element  $x$ .

### C. Fuzzy Support Vector Machines

For fuzzy support vector machines, we give a set  $T$  of labeled training points with associated fuzzy membership

$$T = \{(x_1, y_1, \mu_1), \dots, (x_l, y_l, \mu_l)\} \in (R^n \times Y)^l \quad (5)$$

each training point  $x_i \in R^n$  is given a fuzzy membership  $\sigma \leq \mu_i \leq 1$  with  $y_i \in \{1, -1\}$  and sufficient small  $\sigma > 0$ .  $\mu_i$  is the fuzzy membership of training point  $(x_i, y_i)$ . Then the FSVM is transformed into the following quadratic programming problem :

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \mu_i \xi_i \\ \text{s.t.} \quad & y_i ((\omega \cdot \Phi(x_i)) + b) \geq 1 - \xi_i, \quad i = 1, \dots, l, \\ & \xi_i \geq 0, \quad i = 1, \dots, l \end{aligned} \quad (6)$$

From the above equation, we can see that  $C\mu_i$  represents the classification margin and misclassification scale. For noise and isolated point would be given a small weight, the coefficient  $C\mu_i$  of  $\xi_i$  would be also small so that it can be allowed that  $\xi_i$  becomes larger. The determination of the membership is the key to solving FSVM problem. In this paper, the algorithms are addressed to find the membership.

### III. DETERMINATION OF MEMBERSHIP

#### A. Time Series-Based Membership

FSVM was proposed in 2002 and at the same time the algorithm was given to determine the membership according to the order of the sample points' collection time<sup>[2]</sup>. In the model of time series-based membership, the data collected recently would be more important than the previous data, so they should be given a larger weight. The membership function is:

$$\mu_i = f(t_i), \quad i = 1, 2, \dots, l \quad (7)$$

This equation satisfies that  $\mu_1 = f(t_1) = \sigma$ ,  $\mu_l = f(t_l) = 1$ , where  $\sigma > 0$  is the lower limit of the membership, and  $t_1 \leq \dots \leq t_l$  are time that training point arrival system.  $f(t)$  is an increasing function for  $t$ .

#### B. Space Measure-Based Membership

1) *membership based on the distance of sample points and the sample center*

The basic idea is that, in the feature space, the more close samples from the class center, the greater membership is. Let  $x$  be the sample center and  $d(x_i) = \|x_i - x\|$ , then  $\mu(x_i)$  is a monotone decreasing function of  $d(x_i)$ . Here we introduce a typical s-type function<sup>[3]</sup>.

$$\mu_s(x_i) = \begin{cases} 1, & d(x_i) \leq a \\ 1 - 2 \cdot \left[ \frac{d(x_i) - a}{c - a} \right]^2, & a < d(x_i) \leq b \\ 2 \cdot \left[ \frac{d(x_i) - c}{c - a} \right]^2, & b < d(x_i) \leq c \\ 0, & c < d(x_i) \end{cases} \quad (8)$$

$a, c$  are parameters given in advance and  $b = \frac{a+b}{2}$ .

2) *Membership based on the distance of sample points and the class hyperplane*

In this method, the class hyperplane is used instead of sample center and the membership is a function about class hyperplane. The only difference from the above method is the determination of  $d(x_i)$ . To obtain  $d(x_i)$ , we just need to find class hyperplane<sup>[4]</sup>.

Let  $x^+$  and  $x^-$  be the center of the positive and negative sample points respectively and  $\tilde{\omega} = x^+ - x^-$  be the class hyperplane's normal vector, so class hyperplanes are

$$\tilde{\omega}^T (x - x^+) = 0, \quad \tilde{\omega}^T (x - x^-) = 0 \quad (9)$$

and the distances of the positive and negative respectively to their sample center is:

$$d_{i+} = \frac{|\tilde{\omega}^T (x - x^+)|}{\|\tilde{\omega}\|}, \quad d_{i-} = \frac{|\tilde{\omega}^T (x - x^-)|}{\|\tilde{\omega}\|} \quad (10)$$

so we can use  $d_{i+}$  and  $d_{i-}$  to define the memberships.

#### C. Spatial distribution-based membership

##### 1) Density-Based Membership

In this method, the point's positive and negative density is used to determinate the training point's membership<sup>[5]</sup>. The number of similar points in one point's neighborhood is defined as the point's positive density, and the number of different points is defines as negative density. When a point is outlier, its positive and negative density is small. When the positive density is large and the negative density is small, the positively that this point is noise is very large. If a point is normal sample, its positive density is large and negative density is small.

the density of point is defined:

$$\rho(x_i) = M(\{x | f(x, x_i) \leq T\}) \quad (11)$$

where  $x_i$  and  $x$  are sample point,  $\rho(x_i)$  is the density of  $x_i$ ,  $M(X)$  is the number of set  $X$ ,  $f(x, x_i)$  is the function of similarity measure of  $x_i$  and  $x$ , and  $T$  is a given threshold. In the same way, we can get the membership according the actual needs.

##### 2) Tightness-Based Membership

The tightness of sample set is measured by the smallest hypersphere that surround the sample set. In this method, the membership can be defined through the smallest hypersphere of the sample set. When there are some outlier points in the sample set, a small part of the sample points is allow to locate at the outside of the hypersphere. The follow formula is used to get the smallest hypersphere<sup>[6]</sup>:

$$\begin{aligned} \min_{(R, x, \xi)} \quad & R^2 + D \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \|x_i - x\|^2 \leq R^2 + \xi_i \\ & \xi_i > 0, \quad x_i \in X \end{aligned} \quad (12)$$

where  $R$  is radius and  $x$  is the center of hypersphere. Let  $(R^*, x^*)$  be a solution of the optimization equation and we can get it by solving the dual program of this optimization equation, then we use  $R^*$  and  $x^*$  to make every point's membership.

#### IV. A SOLVING ALGORITHM OF FSVM

FSVM can be transformed into the following quadratic programming problem:

$$\min_{\alpha} d(\alpha) = \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j (\Phi(x_i) \cdot \Phi(x_j)) \alpha_i \alpha_j - \sum_{j=1}^l \alpha_j \quad (13)$$

$$s.t. \sum_{i=1}^l y_i \alpha_i = 0,$$

$$0 \leq \alpha_i \leq \mu_i C, \quad i = 1, \dots, l.$$

Let  $\alpha^*$  be the solution of this quadratic programming, then we can get FSVM according to  $\alpha^*$ .

In this paper, to solve FSVM, we remake SMO<sup>[7][8]</sup> algorithm to Algorithm 1.

##### Algorithm 1:

**Step 1:** Given the initial vector

$$\alpha^0 = (\alpha_1^0, \dots, \alpha_l^0)^T = 0 \text{ and accuracy } \varepsilon. \text{ Let } k = 0;$$

**Step 2:** Solving the subscript of the working set  $i$  and  $j$  according to  $\alpha^k$ :

$$i = \arg \max_t \{-y_t [\nabla d(\alpha^k)]_t \mid t \in I_{up}(\alpha^k)\}$$

$$j = \arg \min_t \{-y_t [\nabla d(\alpha^k)]_t \mid t \in I_{low}(\alpha^k)\}$$

$$I_{up}(\alpha) = \{t \mid \alpha_t < \mu_i C, y_t = 1 \text{ or } \alpha_t > 0, y_t = -1\}$$

$$I_{low}(\alpha) = \{t \mid \alpha_t < \mu_i C, y_t = -1 \text{ or } \alpha_t > 0, y_t = 1\}$$

**Step 3:** We get  $\alpha_j^{new,unc} = \alpha_j^k + \frac{y_j E}{K}$ , where

$$E = \beta_i - \beta_j + \alpha_i^k y_i (\beta_{ii} - \beta_{ij}) + \alpha_j^k y_j (\beta_{jj} - \beta_{ij}) - y_i + y_j$$

$$K = \beta_{ii} + \beta_{jj} - 2\beta_{ij}$$

$$\beta_i = \sum_{k \neq i, j} y_k \alpha_k K(x_k, x_i), \beta_j = \sum_{k \neq i, j} y_k \alpha_k K(x_k, x_j)$$

$$\beta_{ii} = K(x_i, x_i), \beta_{jj} = K(x_j, x_j), \beta_{ij} = K(x_i, x_j)$$

**Step 4:**

$$\alpha_j^{new} = \begin{cases} V, & \text{if } \alpha_j^{new,unc} > V \\ \alpha_j^{new,unc}, & \text{if } U \leq \alpha_j^{new,unc} \leq V \\ U, & \text{if } \alpha_j^{new,unc} < U \end{cases}$$

we can get  $U$  and  $V$  by the following formula:

if  $y_i \neq y_j$ , then

$$U = \max(0, \alpha_j^k - \alpha_i^k)$$

$$V = \min(\mu_j C, \mu_i C - \alpha_i^k + \alpha_j^k)$$

if  $y_i = y_j$ , then

$$U = \max(0, \alpha_j^k + \alpha_i^k - \mu_i C)$$

$$V = \min(\mu_j C, \alpha_i^k + \alpha_j^k)$$

$$\text{then } \alpha_i^{new} = \alpha_i^k + y_i y_j (\alpha_j^k - \alpha_i^{new})$$

**Step 5:** We get  $\alpha^{k+1}$  by update  $\alpha^k$ :

$$\alpha_i^{k+1} = \alpha_i^k, \alpha_j^{k+1} = \alpha_j^k$$

$$\alpha_m^{k+1} = \alpha_m^k, m = 1, \dots, l \text{ and } m \neq i, j$$

#### V. SUMMARY

FSVM has better results in solving problems containing the outlier points and noise points. The determination of the membership is one of the two main issues in FSVM. There are different methods to calculate the membership, so we should choose the best membership function of the actual sample distribution. To solve FSVM needs a large of time and memory, so one of the important issues in the study of FSVM is how to reduce the time spent in solving problems.

#### ACKNOWLEDGEMENT

This paper is supported partly by the NSFC (61170079 and 61202152); Specialized Research Fund for the Doctoral Program of Higher Education of China (20103718110007) and Special Fund for Fast Sharing of Science Paper in Net Era by CSTD(2012107), Sci. & Tech. Development Fund of Qingdao(10-3-3-32-nsh and 2011-2-47), Excellent Young Scientist Foundation of Shandong Province of China (BS2009DX004).

#### REFERENCES

- [1] Lin Chunfu, Wang Shengde. Fuzzy support vector machines[J]. IEEE Transactions on Neural Networks, 2002, 3(2): 464-471.
- [2] Huang Hanpang, Liu Yihuang. Fuzzy support vector machines for pattern recognition and data mining. International Journal of Fuzzy Systems, 2002, 4(3):826-835.
- [3] Bian Zhaoqi, Zhang Xuegong. Pattern Recognition[M]. Beijing: Tsinghua University Press, 1999.
- [4] Zhe Du, Sanyang Liu, Xiaogang Qi. Fuzzy Support Vector Machine with New Membership Function [J]. Journal of System Simulation, 2009, 21(7): 1901-1903

- [5] Jinlong An, ZhengOu Wang, Zhenping Ma. Fuzzy Support Vector Machine Based on Density [J]. Journal of Tianjin University, 2004, 37(6): 544-548.
- [6] Xiang Zhang, Xiaoling Xiao, Guanyou Xu. Fuzzy Support Vector Machine Based on Affinity Among Samples [J]. Journal of Software, 2006, 17(5): 951-958.
- [7] Platt J C. Fast Training of support Vector Machines Using Sequential Minimal Optimization [M]. Cambridge, MA: MIT Press, 1999:185-208.
- [8] Keerthi S S, Shevade S K, Bhattaacharyya C, Murthy K R K. Improvements to Platt's SMO algorithm for SVM classifier design[J]. Neural Computation, 2001, 13( 3) : 637- 649.
- [9] Qi Wu, Rob Law. The complex fuzzy system forecasting model based on FSVM with triangular fuzzy number input and output[J]. Expert Systems with Applications, 2011, 38(10):12085-12093.
- [10] Abu Sayeed Md, Prabir Bhattacharya, Sudhir P. Classification of Ultrasound Medical Images Using Distance Based Feature Selection and Fuzzy-SVM[J]. Springerlink, 2011, 6669(2011):176-183.
- [11] Batuwita, R. Fuzzy Support Vector Machines for Class Imbalance Learning[J]. IEEE Transactions on Fuzzy Systems. 2011, 18(3):558-571.