A Fuzzy Optimization Method for Data Mining

Zhimin Yang*1，XiaoYang*2，Guangli Liu*3
**1, Zhijiang College，Zhejiang University of Technology, Hangzhou P.R.. China, 310024*
**2，Corresponding author College of Business and Administration, Zhejiang University of Technology，*
Hangzhou P.R.. China, 310023
**3, College of Information and Electrical Engineering, China Agricultural University, Beijing P.R..China,100083*
yzm9966@126.com,yangxiao9618@163.com,liugl@cau.edu.cn
doi: 10.4156/jcit.vol5.issue5.9

Abstract

This paper is concerned with the fuzzy support vector classification, in which both of the type of the output training point and the value of the final fuzzy classification function are triangle fuzzy number. First, the fuzzy classification problem is formulated as a fuzzy chance constrained programming. Then, we transform this programming into its equivalence quadratic programming. Final, a fuzzy support vector classification algorithm is proposed to deal with the problem. An example is presented to illustrate rationality of the algorithm.

Keywords: *Machine learning; fuzzy support vector classification; possibility measure; triangle fuzzy number*.

1. Introduction

Support vector machines (SVMs) proposed by Vapnik, is a novel method for machine learning ([1]，[2]，[3]，[4]). Due to its excellent learning capability, It has become a hot spot of the machine learning field and applied successfully to many application fields. But as an immature novel technique, it still needs further improvements; when it is concerned to solve the problem with fuzzy information([9],[10],[11]). for instance, the output $y_j (j = 1, \cdots, l)$ of the training points in the training set $S = \{(x_1, y_1), \cdots, (x_l, y_l)\}$, are fuzzy. In 2002, Lin and Wang [12] proposed a Fuzzy Support Vector Machines(FSVM) technique, but they only added the fuzzy membership grade to the penalty parameters of the quadratic programming in the Fuzzy Support Vector classification rather than building up a Fuzzy Support Vector classification in terms of the algorithm's mathematical essentially. And the value of the final obtained classification functions (the output of the testing points) are still 1 or -1 (positive class or negative class). In 2004 and 2005, Tao Qing, Wang Jueand Lee K Y et.al. respectively proposed A New Fuzzy Support Vector Machine Based on the Weighted Margin（NFSVM）[13] and Possibilistic Support Vector Machine(PSVM) [14]. In NFSVM, the training points' fuzzy informations are represented by fuzzy member and fuzzy margin thus the fuzzy classification problem is transformed to a quadratic programming problem. In PSVM, the training points' fuzzy informations are represented by possibilistic member and possibilistic margin thus the fuzzy classification problem is also transformed to a quadratic programming problem. But the value of the final obtained classification functions (the output of the testing points) for both of the two techniques are still 1 or -1 (positive class or negative class).

In this paper, we study the fuzzy classification problem , in which the training points include the full fuzzy information, i.e., the input of the training points is the membership grades of the positive and negative class, the sum of which is 1. In this technique, both of the training points' output and the value of the constructed fuzzy optimal classification function are triangle fuzzy numbers. This technique is similar to the above three techniques for using the optimization method to solve the fuzzy classification problem in which the training points include fuzzy information. But the difference is that, in this method, the fuzzy optimization method is used to solve the fuzzy classification problem in which the training points include fuzzy information so that the fuzzy information in the training points are included naturally in the fuzzy programming and the obtained fuzzy optimal classification function is still fuzzy. That is to say, given an arbitrary testing point input, substitute it to the fuzzy optimal classification function, and the output are triangle fuzzy numbers. Thus the testing points and the training points are matched in the form and reach agreement in the logic. While the other three techniques are based on ordinary optimization method to solve the fuzzy classification problem with the training points including fuzzy information; thus the final obtained classification function is determinate (without fuzzy feature), which leads to the output of the testing points are determinate ,i.e., positive class or negative class, and not matched to form of the training points' output (fuzzy membership grades); so that the error would occur when the training points are used as testing points to make the LOO error estimation[4].

2. Fuzzy Support Vector Classification

A kind of special triangle fuzzy number is introduced as an extension of positive class symbol 1 or negative class symbol -1 .when the input of a training point is positive class membership grade $\delta^+ (0.5 \le \delta^+ \le 1)$, we define the corresponding output as the following triangle fuzzy number

$$\tilde{y}=(r_1,r_2,r_3)=(\frac{2(\delta^+)^2+\delta^+-2}{\delta^+},2\delta^+-1,\frac{2(\delta^+)^2-3\delta^++2}{\delta^+}),0.5\le\delta^+\le1. \quad (1)$$

Similarly, when the input of a training point is negative class membership grade $\delta^- (0.5 \le \delta^- \le 1)$, the corresponding output triangle fuzzy number is

$$\tilde{y}=(r_1,r_2,r_3)=(\frac{2(\delta^-)^2-3\delta^-+2}{-\delta^-},-2\delta^-+1,\frac{2(\delta^-)^2+\delta^--2}{-\delta^-}),0.5\le\delta^-\le1. \quad (2)$$

Thus we can use $(x,\tilde{y})$ to represent a training point's input x and the output $\tilde{y}$, where $\tilde{y}$ is a triangle fuzzy number as shown in (1) or (2). Also the following Corresponding relationship can be obtained.

$$\delta=\begin{cases}\delta^+ & if \quad x \quad is \quad positive,\\ -\delta^- & if \quad x \quad is \quad negative.\end{cases} \quad (3)$$

And (x,δ) also can be used to represent a training point's input x and the output δ.

Based on the above triangle fuzzy numbers, given the training set of classification is

$$S=\{(x_1,\tilde{y}_1),\cdots,(x_p,\tilde{y}_p),(x_{p+1},\tilde{y}_{p+1}),\cdots,(x_l,\tilde{y}_l)\} \quad (4)$$

or

$$S_\delta=\{(x_1,\delta_1),\cdots,(x_p,\delta_p),(x_{p+1},\delta_{p+1}),\cdots,(x_l,\delta_l)\} \quad (5)$$

where $x_j \in R^n$ is an usual input, $\tilde{y}_j (j=1,\cdots,l)$ is a triangle fuzzy number as shown in (1) or (2), δ_j is shown in (3), $j=1,\cdots,l$, $(x_t,\tilde{y}_t)$ and (x_t,δ_t) are fuzzy positive points ($t=1,\cdots,p$), $(x_i,\tilde{y}_i)$ and (x_i,δ_i) are fuzzy negative points ($i=p+1,\cdots,l$).

Definition 1 Suppose a fuzzy training set as shown in (6) or (7), given a confidence level λ

（$0 < \lambda \le 1$）, if there exist $w \in R^n$ and $b \in R$ so that

$$Pos\{\tilde{y}_j((w \cdot x_j)+b) \ge 1\} \ge \lambda,\ j=1,\cdots,l \qquad (6)$$

where $Pos\{\cdot\}$ is a feasibility estimate of the fuzzy event [5]. Then fuzzy training set (6) or (7)is considered to be fuzzy linearly separable, and the corresponding fuzzy classification problem be fuzzy linearly separable.

Theorem 1. Inequality (8) in Definition 1 is equivalent to the following real inequalities:

$$\begin{cases} ((1-\lambda)r_{t3}+\lambda r_{t2})((w \cdot x_t)+b) \ge 1, t=1,\cdots,p, \\ ((1-\lambda)r_{i1}+\lambda r_{i2})((w \cdot x_i)+b) \ge 1, i=p+1,\cdots,l. \end{cases} \qquad (7)$$

Where r_{t2}, r_{t3} are the center and right end points of the fuzzy positive point $y_t(t=1,\cdots,p)$ (triangular fuzzy number); r_{i2}, r_{i1} are the center and left end points of the fuzzy positive point $y_i(i=p+1,\cdots,l)$ (triangular fuzzy number) [5].

Under the confidence level $\lambda(0<\lambda\le 1)$, fuzzy linearly separability problem of the fuzzy training set as shown in (4)or (5), can be transformed to fuzzy chance constrained programming with decision variable $(w,b,)^T$:

$$\begin{cases} \min\limits_{w,b} \frac{1}{2}\|w\|^2, \\ s.t. Pos\{\tilde{y}_j((w \cdot x_j)+b) \ge 1\} \ge \lambda, j=1,\cdots,l, \end{cases} \qquad (8)$$

Theorem 2 under the confidence level $\lambda(0<\lambda\le 1)$, the certain equivalence programming (common programming equivalent to(8) of the fuzzy chance constrained programming（8）is the quadratic programming:

$$\begin{cases} \min\limits_{w,,b} \frac{1}{2}\|w\|^2, \\ s.t. ((1-\lambda)r_{t3}+\lambda r_{t2})((w \cdot x_t)+b) \ge 1, t=1,\cdots,p, \\ ((1-\lambda)r_{i1}+\lambda r_{i2})((w \cdot x_i)+b) \ge 1, i=p+1,\cdots,l. \end{cases} \qquad (9)$$

Theorem 3 There exists an optimal solution of quadratic programming (9).

Theorem 4 The duality programming of quadratic programming (9) is quadratic programming with decision variable $(\beta,\alpha)^T$:

$$\begin{cases} \min\limits_{\beta,\alpha} \frac{1}{2}(A+2B+D)-(\sum_{t=1}^{p}\beta_t+\sum_{i=p+1}^{l}\alpha_i), \\ s.t. \sum_{t=1}^{p}\beta_t((1-\lambda)r_{t3}+\lambda r_{t2})+\sum_{i=p+1}^{l}\alpha_i((1-\lambda)r_{i1}+\lambda r_{i2})=0, \\ \beta_t \ge 0, t=1,\cdots,p, \\ \alpha_i \ge 0, i=p+1,\cdots,l, \end{cases} \qquad (10)$$

where $A=\sum_{t=1}^{p}\sum_{s=1}^{p}\beta_t\beta_s((1-\lambda)r_{t3}+\lambda r_{t2})((1-\lambda)r_{s3}+\lambda r_{s2})(x_t \cdot x_s)$,

$$B=\sum_{t=1}^{p}\sum_{i=p+1}^{l}\beta_t\alpha_i((1-\lambda)r_{t3}+\lambda r_{t2})((1-\lambda)r_{i1}+\lambda r_{i2})(x_t \cdot x_i),$$

$$D=\sum_{i=p+1}^{l}\sum_{q=p+1}^{l}\alpha_i\alpha_q((1-\lambda)r_{i1}+\lambda r_{i2})((1-\lambda)r_{q1}+\lambda r_{i2})(x_i\cdot x_q),$$

$\beta=(\beta_1,\cdots,\beta_p)^T\in R_+^p$, $\alpha=(\alpha_{p+1},\cdots,\alpha_l)^T\in R_+^{l-p}$, $(\beta,\alpha)^T$ is decision variable. Programming (10) is a convex quadratic programming. After getting its optimal solution $(\beta^*,\alpha^*)^T=(\beta_1^*,\cdots,\beta_p^*,\alpha^*{}_{p+1},\cdots\alpha^*{}_l)^T$, we can get the certain optimal classification hyperplane (see [4]) :

$$(w^*\cdot x)+b^*=0, x\in R^n \tag{11}$$

$$w^*=\sum_{t=1}^{p}\beta_t^*((1-\lambda)r_{t3}+\lambda r_{t2})x_t+\sum_{i=p+1}^{l}\alpha_i^*((1-\lambda)r_{i1}+\lambda r_{i2})x_i,$$

$$b^*=((1-\lambda)r_{s3}+\lambda r_{s2})-(\sum_{t=1}^{p}\beta_t^*((1-\lambda)r_{t3}+\lambda r_{t2})(x_t\cdot x_s)+\sum_{i=p+1}^{l}\alpha_i^*((1-\lambda)r_{i1}+\lambda r_{i2})(x_i\cdot x_s))$$

$s\in\{s\mid\beta_s^*>0\}$, or

$$b^*=((1-\lambda)r_{q1}+\lambda r_{q2})-(\sum_{t=1}^{p}\beta_t^*((1-\lambda)r_{t3}+\lambda r_{t2})(x_t\cdot x_q)+\sum_{i=p+1}^{l}\alpha_i^*((1-\lambda)r_{i1}+\lambda r_{i2})(x_i\cdot x_q))$$

$q\in\{q\mid\alpha_q^*>0\}$.

For the Fuzzy Support Vector classification, however, we expect to obtain an fuzzy optimal classification function whose value is a triangular fuzzy number $\tilde{y}_j$ as shown in (1) and (2), which can be achieved through $g(x)=(w^*\cdot x)+b^*$ and the following function (12):

$$\delta=\delta(u)=\begin{cases}\varphi_+(u), 0<u\le\varphi_+^{-1}(1),\\ 1, u>\varphi_+^{-1}(1),\\ -\varphi_-(u), \varphi_-^{-1}(1)\le u<0,\\ -1, u<\varphi_-^{-1}(1),\end{cases} \tag{12}$$

where $\varphi_+^{-1}(u)$ and $\varphi_-^{-1}(u)$ are respectively the inverse function of $\varphi_+(u)$ and $\varphi_-(u)$. $\varphi_+(u)$ is a regression function (monotonous increasing function on u) obtained from the $\varepsilon-$support vector regression. And $\varepsilon-$ support vector regression is constructed by the following method.

(i)Build up the Construct training set of the regression problem

$$\{(g(x_1),\delta_1),\cdots,(g(x_p),\delta_p)\} \tag{13}$$

(ii) Use (13) as training set , and select appropriate $\varepsilon>0$, penalty parameter $C>0$,choose a linear kernel as the kernel function ,thus to construct $\varepsilon-$support vector regression.

Similarly, $\varphi_-(u)$ is a regression function (monotonous decreasing function on u) obtained from the $\varepsilon-$support vector regression which is constructed with the same method.

According to the corresponding relationship (3) and the transformation rules (1) and (2), the fuzzy optimal classification function will be achieved by transforming the function $\delta=\delta(g(x))$ in (12) to a triangular fuzzy number $\tilde{y}=\tilde{y}(x)$.

Given an arbitrary testing point input $\bar{x}$, substitute it to the fuzzy optimal classification function, and then obtain a output are triangle fuzzy numbers $\tilde{y}$ which is the output of the testing points and can objectively reflect the fuzzy classification situation of the testing points

$(\overline{x}, \widetilde{y})$ (shown the membership grade of the positive or negative testing input).

3. Conclusion

From the above discussion, we can achieve the following algorithm.

（1）Given a fuzzy training set (4) or (5) ,and select an appropriate confidence level $\lambda(\sigma \le \lambda \le 1)$, $C > 0$ and a kernel function $K(x, x')$,then construct quadratic programming:

$$\begin{cases} \min\limits_{\beta,\alpha} \dfrac{1}{2}(A_K + 2B_K + D_K) - (\sum\limits_{t=1}^{p} \beta_t + \sum\limits_{i=p+1}^{l} \alpha_i) \\ s.t. \sum\limits_{t=1}^{p} \beta_t((1-\lambda)r_{t3} + \lambda r_{t2}) + \sum\limits_{i=p+1}^{l} \alpha_i((1-\lambda)r_{i1} + \lambda r_{i2}) = 0 \\ 0 \le \beta_t \le C, t = 1, \cdots, p \\ 0 \le \alpha_i \le C, i = p+1, \cdots, l, \end{cases} \tag{14}$$

where $A_K = \sum\limits_{t=1}^{p}\sum\limits_{s=1}^{p} \beta_t \beta_s ((1-\lambda)r_{t3} + \lambda r_{t2})((1-\lambda)r_{s3} + \lambda r_{s2})K(x_t, x_s)$,

$$B_K = \sum_{t=1}^{p}\sum_{i=p+1}^{l} \beta_t \alpha_i ((1-\lambda)r_{t3} + \lambda r_{t2})((1-\lambda)r_{i1} + \lambda r_{i2})K(x_t, x_i),$$

$$D_K = \sum_{i=p+1}^{l}\sum_{q=p+1}^{l} \alpha_i \alpha_q ((1-\lambda)r_{i1} + \lambda r_{i2})((1-\lambda)r_{q1} + \lambda r_{i2})K(x_i, x_q),$$

$\beta = (\beta_1, \cdots, \beta_p)^T \in R_+^p$, $\alpha = (\alpha_{p+1}, \cdots, \alpha_l)^T \in R_+^{l-p}$, $(\beta, \alpha)^T$ is a decision variable.

（2）Solve quadratic programming(14), get optimal solution
$(\beta^*, \alpha^*)^T = (\beta_1^*, \cdots, \beta_p^*, \alpha_{p+1}^*, \cdots, \alpha_l^*)^T$.

（3）Select the positive component $\beta_s^* \in (0, C)$ of β^*,or $\alpha_q^* \in (0, C)$ of α^*，then compute

$$b^* = ((1-\lambda)r_{s3} + \lambda r_{s2}) - (\sum_{t=1}^{p} \beta_t^*((1-\lambda)r_{t3} + \lambda r_{t2})K(x_t, x_s) + \sum_{i=p+1}^{l} \alpha_i^*((1-\lambda)r_{i1} + \lambda r_{i2})K(x_i, x_s))$$

or

$$b^* = ((1-\lambda)r_{q1} + \lambda r_{q2}) - (\sum_{t=1}^{p} \beta_t^*((1-\lambda)r_{t3} + \lambda r_{t2})K(x_t, x_q) + \sum_{i=p+1}^{l} \alpha_i^*((1-\lambda)r_{i1} + \lambda r_{i2})K(x_i, x_q))$$

（4）Construct function :

$$g(x) = \sum_{t=1}^{p} \beta_t^*((1-\lambda)r_{t3} + \lambda r_{t2})K(x_t, x_s) + \sum_{i=p+1}^{l} \alpha_i^*((1-\lambda)r_{i1} + \lambda r_{i2})K(x_i, x_s).$$

（5）Consider $\{(g(x_1), \delta_1), \cdots, (g(x_p), \delta_p)\}$ and $\{(g(x_{p+1}), -\delta_{p+1}), \cdots, (g(x_l), -\delta_l)\}$ as training set respectively, construct $\varepsilon-$ support vector regression (select appropriate $\varepsilon > 0$, penalty parameter $C > 0$,choose a linear kernel as the kernel function), then obtain regression functions $\varphi_+(u)$ and $\varphi_-(u)$,and construct function(16).

（6）According to corresponding relationship (3) and the transformation rules (1) and (2), the fuzzy optimal classification function will be achieved by transforming the function $\delta = \delta(g(x))$ in (12) to a triangular fuzzy number $\tilde{y} = \tilde{y}(x)$.

Note: If the outputs of all fuzzy training points in fuzzy training set(6) or (7) are all real number 1 or -1, then the fuzzy training set degenerates to the common training set, so the fuzzy support vector classification degenerates to the support vector classification .

4. Acknowledgement

This paper is supported by National Natural Science Foundation of China (No.10926198)

5.References

[1] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag,1995.
[2] Cristianini N,et al. Introduction to Support Vector Machines. Cambridge: Cambridge University Press，2000.
[3] Mangasarian O L. Generalized Support Vector Machines. Advances in Large Margin Classifiers. MIT Press, 1999.
[4] Deng Naiyang, Tian Yingjie. The New Method in Data Mining——Support Vector Machines. Beijing,China: Science Press,2004.
[5] Liu Baoding ZhaoRuqing. Random Programming and Fuzzy Programming. Beijing, China: Tsinghua University Press,1998.
[6] Yang Zhimin, Liu Guangli. Principium and Applications of Uncertain Support Vector Machines. Beijing,China: Science Press,2007.
[7] Yuan Yaxiang, Sun Wenyu. Optimal Theories and Methods. Beijing,China: Science Press,1997.
[8] Zadeh L A. Fuzzy Sets as a Basis for a Theory of Possibility . Fuzzy Sets and Systems.1978，1(1)：3-28.
[9] Liu Baoding. Chance Constrained Programming with Fuzzy Parameters. Fuzzy Sets and Systems, 1998 ，94(2)：227-237.
[10] D.Latha, Md.Abdul Rahim,G.Ganesan. Intuitionistic rough fuzziness and its generalizations. Journal of Convergence Information Technology,2008,3(3):82-88.
[11] Hrudaya Ku.Tripathy, B.K.Tripathy, Pradip K Das. A Knowledge based Approach Using Fuzzy Inference Rules for Vowel Recognition. Journal of Convergence Information Technology,2008,3(1):51-56.
[12] Lin Chunfu , Wang Shengde . Fuzzy Support Vector Machines. IEEE Transactions on Neural Networks. 2002，13（2）：464-471.
[13] Tao Qing,.Wang Jue. A New Fuzzy Support Vector Machine Based on the Weighted Margin.Neural Procession Letters,2004,20(3):139-150.
[14] Lee K Y,Kim D W,Lee K H,Lee D. Possibilistic Support Vector Machine, Pattern Rrcognition,2005,38(3):1325-1327.