Research on a new membership functions based on fuzzy SVM

Qiuhua Xu

Zhoukou vocational college of science and technology, Henan, 466000, China

Corresponding author's e-mail: xuqiuhuazk@163.com

Received 01 December 2013, www.cmnt.lv

Abstract

Traditional membership functions in fuzzy SVM (FSVM) were designed based on the distance between a sample and its cluster center, which are irrational for dataset with non-spherical-shape distribution. A new membership function was proposed based on the distance between a sample and a hyper plane within the class. It overcomes disadvantages of traditional designing methods and improves the generalization ability of FSVM, while reducing the dependence of membership function on the geometric shape of sample data. Numerical experiments show that, compared with the traditional SVM and three FSVM with different membership functions, FSVM with new membership function has better classification accuracy. The New method is simple and its computation time is less.

Keywords: support vector machine, fuzzy support vector machine, membership function, classification

1 Introduction

Because of its good performance and broad application [3, 4], Support vector machines (SVM) [1, 2] is the research focus in recent years. In order to reduce its sensitivity to noise and outliers and improve the generalization ability one way is the traditional support vector machine to introduce new parameters, and the typical approach is fuzzy support vector machine. Fuzzy functions will support vector machine for each sample by introducing fuzzy membership parameters to achieve. Because the standard support vector machine are all treated the same input samples, and samples from selected to form part of the final classification hyper plane, so it is very sensitive to noise and outliers. When introducing the membership parameters for each sample hyper plane contribution is not the same, compared with the traditional support vector machine, fuzzy support vector function better and reduce the effects of noise caused by outliers and improve classification accuracy. In the fuzzy support vector machine theory, fuzzy membership function design is the most critical step of the algorithm to achieve the degree of difficulty, and the final classification results have a very important impact of different membership function design. This requires the design of membership functions must be able to accurately reflect the sample distribution system and the uncertainty. In [5] In the original input space, the membership is seen as a linear function of the sample distance where the geometric center of its kind between the sample, which is equivalent to all similar samples taken on its cover a super ball, the farther away from the center of the sphere of membership smaller. In [6], this idea will be carried out by the kernel function in the feature space. [7] Seek an appropriate feature space over the ball; the same sample selectively covering samples outside the sphere is considered outliers or noise. Then membership is regarded as a linear function of the distance between the sample and the center of the sphere, which is different from the previous full coverage. However, due to the need to solve two quadratic programming, requires a lot of time and space overhead. Overall, these methods are based on the distance from the center of the sample with the class, and their pros and cons, for a sample set of more favorable spherical distribution.

This article instead of using class hyper plane class center, the sample point to a linear function of the distance of the plane in the class as a membership function, membership function which reduces dependence on the geometry of the sample set, more in line with SVM hyper plane classification principle, Jane and the method to calculate the speed.

2 Fuzzy SVM

Assuming the training set can be noted as $\{x_i, y_i\}_{i=1}^{n} \in R^m \times \{+1,-1\}$, Kernel function's anonymous mapping is $\wp(\mathrm{x})$. Then the training set transfers to $(\wp(x_i), y_i)$, and hyper plane is $\omega \bullet \wp(x_i)+b=0$. Fuzzy factor s_i ($0< s_i <=1$, i=1,2,…,n) represents normal level of the i-th sample. So training set is transformed into the training sample set with fuzzy factor $(\wp(x_i), y_i, s_i)$ where $s_i\xi_i$ is the relaxation factor with different weights. And thus we can do the formation of fuzzy support vector machine (FSVM). Solving optimization problems the hyper plane is:

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n} S_i\xi_i \,, \tag{1}$$

$$y_i\left[\omega \bullet \varphi(x_i)+b\right]-1+\xi_i \geq 0 \,. \tag{2}$$

where $\xi_i \geq 0$ i=1,2,3,…,n and C is a constant. With standard support vector machine is similar to the solution process, by constructing the Lagrange function and saddle point condition, obtain the original problem (1) of dual programming:

$$\max W(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$

$$s.t.\sum_{i=1}^{n}\alpha_i y_i=0, 0\le\alpha_i\le s_i C, i=1,2,...,n. \quad (3)$$

The $K(x_i,x_j)=\varphi(\mathrm{x}_i)\bullet\varphi(\mathrm{x}_j)^T$ in the formula is kernel function. Clearly distinguish fuzzy support vector machines and standard support vector machine is that the upper bound constraints of variable α_i are changed with fuzzy factor s_i in the dual problem. This is equivalent to using a Penalty factor $s_i C$ in per sample. So the define of fuzzy factor s_i become the crucial decision weather fuzzy support vector machine performance is good or bad.

3 Class - plane based membership functions

SVM hyper plane is mainly determined by the surface near the support vector classification, but the fuzzy support vector machine is left hyper plane sample distance (relaxation factor) weighted. So, traditional membership functions and classes from the center of the sample design based on real inadequacies would be prompted. But away from the center of the sphere is bound to reduce the role of classification from the sample point close to the surface.

Solid sample points in Figure 1(a) classification as shown near the surface. Because of its distance the center where the class and will be reducing its membership values. Additional Shown in Figure 1(b) a class of sample distribution conditions, Sample points shown for two solid future contribution classification surfaces is similar, but the distance they are completely different from the cluster center. Such use of the conventional design method will provide them with a big difference. In addition it will be given a small membership values for the key box sample points, using paper [5] approach is more likely to be mistaken into noise points. As for the differences in the distribution of such non-spherical data is more evident. For radial basis function frequently used terms, implicit mapping all samples are mapped into feature space over a unit sphere, while covering the same sample points over the ball might coincide with the unit ball or be included at this time of any class-based design methodology centers are not very good samples reflect the contribution of hyper plane.

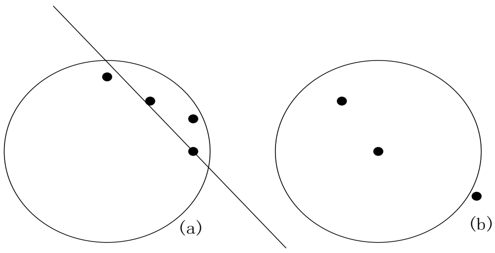

FIGURE 1 Class center design methodology-based schematic

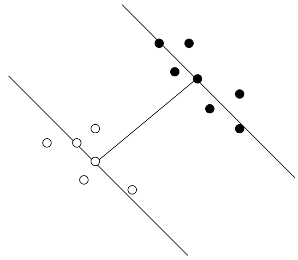

FIGURE 2 Membership design methods schematic

SVM classification is classifying using the maximum interval hyper plane in the original space or feature space. Hyper plane in this article can instead of the class using the class center. With the sample point to a linear function hyper plane distance to design membership function, get a more realistic approximation of the contribution of the sample, thus avoiding the above disadvantages surface classification.

As shown in Figure 2, the mean positive and negative points for class training samples is x_+, x_-. $\varpi=\mathrm{x}_+-x_-$ is normal vector. Two within hyper planes crossing over x_+, x_- are:

$$\varpi^T x-\varpi^T x_+=0,\ \varpi^T x-\varpi^T x_-=0.$$

Scilicet:

$$\varpi^T(x-x_+)=0,\ \varpi^T(x-x_-)=0. \quad (4)$$

Now positive and negative class category within each sample point to the hyper plane distances are

$$\eta_{i+}=\frac{|\varpi^T(x-x_+)|}{\|\varpi\|},\ \eta_{i-}=\frac{|\varpi^T(x-x_-)|}{\|\varpi\|}. \quad (5)$$

$D_+=\max\{\eta_{i+}\}$ and $D_-=\max\{\eta\text{-}\}$ respectively positive and negative deviation from their class in the class samples within a maximum distance hyper plane. Membership calculating function is

$$\mathrm{s}_i=1-\frac{\eta_{i+}}{D_+ +\delta}, \mathrm{y}_i=+1,\ \mathrm{s}_i=1-\frac{\eta_{i-}}{D\text{-}+\delta}, \mathrm{y}_i=-1. \quad (6)$$

δ is previously given a small positive number in order to ensure $0<s_i\le 1$. Because $\eta_{i\pm}$ and $D_{i\pm}$ both contains $\|\varpi\|$, so in order to calculate the simple, we can get from Equaiton (4) that $\eta_{i+}=|\varpi^T(x-x_+)|$, $\eta_{i-}=|\varpi^T(x-x_-)|$.

For nonlinear case, we use $\partial(\mathrm{x}\ )$. CKS class feature space center is $\partial(\mathrm{x}_+)=\frac{1}{\mathrm{n}}\sum_{i=1}^{n_+}\partial(\mathrm{x}_i)$, and negative class center is $\partial(\mathrm{x}_-)=\frac{1}{\mathrm{n}}\sum_{i=1}^{n_-}\partial(\mathrm{x}_i)$. n_+ and n_- are respectively the number of positive and negative sample class. Then get this:

$$\eta_{i+}=(\partial(\mathrm{x}_+)-\partial(\mathrm{x}_-))^T(\partial(\mathrm{x}\ )-\partial(\mathrm{x}_+)),$$

$$\eta_{i-}=(\partial(\mathrm{x}_+)-\partial(\mathrm{x}_-))^T(\partial(\mathrm{x}\ )-\partial(\mathrm{x}_-)). \quad (7)$$

Also according to Equation (6) in the feature space we can obtain membership function design.

4 Numerical experiments

The following numerical experiments are based on the CPU2.6G, 512M of inner memory microcomputer. The Matlab7.0.1 is also used to write membership function. SVM training primary function takes adaptation LIBsvm [8] package.

Then we make a comparison of a standard SVM [5], the class-based Center CFSVM and HFSVM classification. Classification results can be seen that HFSVM is the best. Then HFSVM and paper [5] CFSVM, [6] in KFSVM, paper [7] AFSVM different membership function of fuzzy support vector machine classification results were compared. Data are from UCI database standard test set; comparison index is classification accuracy (percentage) and training time (sec). The number of samples for training and testing data sets first column. When the experimental nuclear radial basis function uses cross method parameters, dataset with 5-fold cross-validation (5-fold cross validation) after averaging, the results is shown in Table 1. The article can be seen HFSVM have a good classification accuracy. Due to extra computing membership, all fuzzy SVM training time than the standard SVM much, but in exchange for the higher classification accuracy. As can be seen from Table 1, when the membership function is poorly designed, it will take lower than the standard classification accuracy of SVM (eg CFSVM for anana set). HFSVM time and CFSVM is basically similar, but the high classification accuracy than CFSVM. Also AFSVM classification accuracy and HFSVM fairly basic, but the former requires solving two quadratic programming to determine the membership, and debugging a lot of parameters, compared HFSVM easy, fast.

TABLE 1 The run time for five edge detection algorithms

	Dataset	Banana 400/4900	Splice 1000/2175	Diabetics 468/300	German 700/300
accuracy	SVM	88.47	86.43	76.47	76.39
	CFSVM	88.10	87.03	76.42	76.39
	KFSVM	87.12	87.90	77.02	76.39
	AFSVM	88.87	88.05	77.96	77.33
	HFSVM	88.86	89.50	77.42	77.52
time	SVM	0.84	3.00	0.92	2.08
	CFSVM	0.88	3.12	0.92	2.68
	KFSVM	1.02	3.96	1.12	2.01
	AFSVM	2.51	5.86	2.73	4.62
	HFSVM	0.87	3.12	0.97	2.07

5 Conclusions

This paper presents a design method of membership function based on the distance of the sample to the hyper plane. This method essentially from the classification SVM departure, reduce support vector machine classification capability based on traditional membership functions from the center of the sample to the class dependent on the geometry of the sample set, more rational design of membership functions, improved. By artificial data sets and real data sets UCI tests show that compared with several existing membership function design method, this method is simple, so fuzzy support vector machine to get higher classification accuracy, but the running time is relatively less, so you can expand the scope of application of fuzzy support vector machine.

References

[1] Vapnik V 1995 The Nature of Statistical Learning Theory *New York USA Springer* **5**(1) 69-77

[2] Cristianini N, Schawe-Taylor J 2000 An Introduction to Support Vector Machines *Cambridge England Cambridge University Press* 169-71

[3] Brown J 1994 Some motivational issues in computer-based instruction *Educational Technology* **26**(4) 27-9

[4] Zhang Z, Zhang S, Zhang C X, Chen Y Z 2005 Density Estimation Based on Support Vector Machines *Journal of System Simulation* (S1004-731X) **17**(10) 2355-7

[5] Lin C F, Wang S D 2005 *IEEE Transaction on Neural Networks* (S1045-9227) **13**(2) 2355-7

[6] Jiang X F, Yi Z, Lv J C 2012 Fuzzy SVM with a New Fuzzy Membership Function *Neural Computation and Application* (S0941-0643) **3**(15) 268-76

[7] Nakagawa Y, Rosenfeld Y 1978 *IEEE Transaction Systems, Man & Cybernetics* **8**(8) 632-5

[8] Chang C C, Lin C J 2007 Libsvm-A Library for Support Vector Machines [CP/OL] http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.htm

Author

Xu Qiuhua, 1980, Henan Province of China.

Current position, grades: lecturer.
University studies: Zhengzhou University of Light Industry.
Scientific interest: control theory.