Application of Least Squares Support Vector Machine Based on Fast Sparse Approximation in Radar Target Recognition

Dongbo Zhao [1, a], Hui Li [2]

[1]School of Electronic Engineering, Xi'an Aeronautical University, Xi'an, 710077, China;

[2]School of Electronic Information, Northwestern Polytechnical University, Xi'an, 710129, China.

[a]alien_dffy@163.com

Abstract

Least squares support vector machine (LS-SVM) has great computational complexity and the lack of sparsity. Aiming at this problem, this paper proposed a fast sparse approximation of least squares support vector machine algorithm (FSALS-SVM). The training speed of least square support vector machine (LS-SVM) can be further improved by combining the improved method of solving the problem of method and selection strategy of basis function based on reverse fitting, and the support vector is more global and superior. The sparsity of the detection function can be obtained by introducing the non sensitive criterion to reduce the computational complexity. Based on experiments of one-dimensional range profile of radar target classification and recognition show that FSALS-SVM have more sparse under the premise of same generalization performance.

Keywords

Kernel Principal Component Analysis (KPCA); Central moment; Radar Target Recognition; Least Squares Support Vector Machine(LS-SVM); Fast Sparse Approximation Least Squares Support Vector Machine(FSALS-SVo).

1. Introduction

Support vector machine (SVM) performs well in many classification problems as a classifier, but it have some shortcomings, such as large amount of computation, low speed, more experience in parameter selection and poor performance in multi classification problems. In order to solve the problem of slow speed of SVM classification, Lee et al put forward a variety of methods such as reducing SVM training time and improving training speed, but these algorithms are difficult to apply to the learning test of large data sets because of their lack of sparsity.

The least squares support vector machine (LS-SVM) maximizes the maximum interval between two classes at the same time minimizing the minimum square error of the training sample, and obtains better classification results in many classification problems. However, there are still two defects in LS-SVM. First, its training process is equivalent to the solution of linear equations [1]. When solving large scale linear equations, the classical method can not be used to solve the equations. Secondly, the solution of LS-SVM is different from the solution of SVM, and the solution of LS-SVM is not sparsity, which will lead to slower testing speed than SVM.

Based on the purpose of overcoming the two defects of LS-SVM, this paper presents a fast sparse approximation least squares support vector machine (FSALS-SVM) algorithm.

2. Feature Extraction

2.1 Kernel Principal Component Analysis

Kernel principal component analysis (KPCA) uses kernel theory to extract nonlinear structures in data and it's suitable for solving nonlinear problems. It provides a good method for nonlinear feature extraction, and has been widely used in radar target recognition field.

The training data $x_i(i=1,2,...N)$ is a set of zero mean radar range profiles, and the points of the range profiles are N dimensional. Set $\phi: R^n \to F$ is a nonlinear mapping, and After a nonlinear transformation of x_i, that is, transform $\phi(x_i)$, can get a set $A=\{\phi(x_1),\phi(x_2),.....,\phi(x_N)\}$.

The covariance matrix of samples in F space is:

$$\overline{C}=\frac{1}{N}\sum_{j=1}^{N}\phi(\overline{x_j})\phi(\overline{x_j})^T \tag{1}$$

Feature decomposition of C:

$$\overline{C}v=\lambda v \tag{2}$$

v is the eigenvector of the F space corresponding to the λ. The eigenvalue of C is nonnegative, Set $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N \geq 0$, and the corresponding eigenvector is the feature vector, which is formed by the sample of F space.

$$v=\sum_{j=1}^{N}\alpha_j\phi\left(\overline{x_j}\right) \tag{3}$$

A matrix K: $K_{ij}=k(\overline{x_i},\overline{x_j})=\phi^T(\overline{x_i})\phi(\overline{x_j})$ is defined to satisfy the Mercer kernel condition. First of all, the 0 mean processing of the data after mapping is [4]:

$$\overline{\phi}(\overline{x_i})=\phi(\overline{x_i})-\frac{1}{N}\sum_{i=1}^{N}\phi(\overline{x_i}) \tag{4}$$

At the same time, the corresponding centralization kernel matrix can be obtained:

$$\overline{K}=(I_{NN}-\frac{1}{N}1_{NN})K(I_{NN}-\frac{1}{N}1_{NN}) \tag{5}$$

The two sides of the formula (2) are left multiplied by $\phi(\overline{x_i})$, which can be obtained:

$$\phi(\overline{x_i})\overline{C}v=\lambda\phi(\overline{x_i})v \tag{6}$$

Replace (1) and (3) into (6):

$\overline{K}\alpha=N\lambda\alpha=\lambda'\alpha$

Set $\lambda_1' \geq \lambda_2' \geq ... \geq \lambda_p' \geq 0$, which $\alpha_1,\alpha_2,...\alpha_p$ is the corresponding eigenvector. λ_p' is the last non zero eigenvalue, so the first p eigenvalues and corresponding eigenvectors should be selected. Normalization conditions are used: normalization, so that the eigenvectors can be obtained. Normalization condition $\lambda'\alpha K^T\alpha_s=\delta_{ks},\forall k=1,2,...,p$ of α is used to normalize v so that the eigenvector v_k of $\overline{C}$ can be obtained. The projection of input samples in the feature space is the projection of x_i in the v_k direction.

$$y_k=[v_k\cdot\phi(x)]=\sum_{j=1}^{N}\alpha_j^k[\phi(\overline{x_j})\cdot\phi(x)]=\sum_{j=1}^{N}\alpha_j^k k(\overline{x_j},x),k=1,2,...p \tag{8}$$

k=1, 2,……,p

y_k is the corresponding component of the P principal component. All projection values $y_k(x)=[y_k(x),y_{k+1}(x),...,y_m(x)]^T$ constitute variables, which are the characteristic variables of samples.

The kernel function is introduced in the calculation of y_k. Using kernel function instead of dot product operation in F space, the computation complexity and complexity can be greatly reduced [4]. The selection of different kernel functions will have different effects. The Gauss kernel function is global

kernel, which can extract the global characteristics of range images. Therefore, the Gauss kernel function is used in this paper:

$$k(x,y)=\exp\left\{-\frac{\|x-y\|^2}{2\sigma^2}\right\} \tag{9}$$

2.2 Center Moment Feature Extraction

The center moment reflects the shape information of the target. It is a simple translation, rotation and scale invariant feature. $\{x(i),i=1,2,...,d\}$ is the first i high resolution range profile, where d is the number of distance units. Before solving the central moments, the following normalization operations are performed first:

$$\overline{x}(i)=\frac{x(i)}{\sum_{i=1}^{d}x(i)} \tag{10}$$

From this formula, we can see that $x(i)$ is the discrete probability distribution function . In this way, the central moment of the first i high resolution range profile can be expressed as:

$$u^{(p)}=\sum_{i=1}^{d}(i-i_0)^p\,\overline{x}(i) \tag{11}$$

In this way, the central moment vector of the first i high resolution range profile can also be written as:

$$f=\left[f(1),f(2),...,f(p_{\max-1}-1)\right]^T=\left[u^{(2)},u^{(3)},...,u^{(p_{\max})},\right]^T \tag{12}$$

Among them, $p_{\max}$ is the highest order of the center moment. In radar target recognition, it is necessary to identify suitable $p_{\max}$.

The magnitude of the central moments increases sharply with the order, and the central moments of each order are in an order of magnitude. If these values are used directly, the performance of the classifier will be degraded. The effect of it on classifiers can be eliminated by the range transformation shown in the following expressions:

$$\overline{f}_{nj}=\frac{f_{nj}-f_{n,\min}}{f_{n,\max}-f_{n,\min}},n=2,3,...,p_{\max} \tag{13}$$

the N in this formula is the number of high resolution range profiles.

However, the central moment characteristics of range profiles still have range sensitivity of targets. So in this paper, the average distance profiles is obtained from the coherent average of the training distance in each corner domain, then the center moment vector of the average distance profiles is extracted as the template feature. The mean distance profile is actually the self term in the range profiles, which remains stable in the range of corners. The average distance image can be approximated as:

$$\overline{x}\approx\frac{1}{M}\sum_{n=1}^{M}|x_n(n)|\approx\sqrt{\sum_{k=1}^{K}|a_{nk}^2|} \tag{14}$$

2.3 Feature Extraction Steps

(1) firstly, the omnidirectional angle is uniformly divided by the distance migration without scattering points, and the average distance profile is extracted at each interval.

(2) then the center moment representation of the mean distance profiles is extracted according to the central moment representation defined by the formula (11), and the eigenvectors can be obtained after the central moment is normalized.

(3) finally, the redundant information in the invariant distance profiles will affect the recognition rate, so the kernel principal component analysis (KPCA) is used to compress the feature vectors to remove the redundant features and to compress the feature dimension.

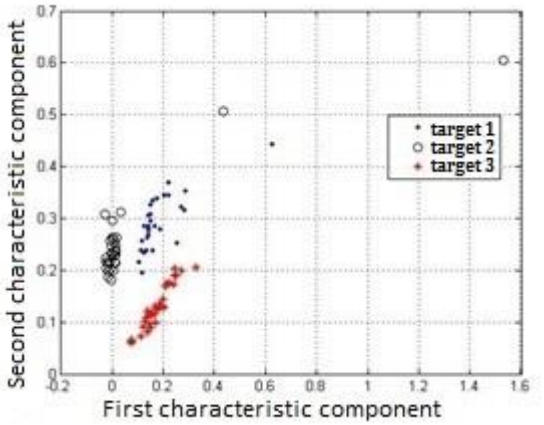

Fig 1. after extracting the central moment eigenvector, the result is compressed by KPCA feature.

3. Classifier Design

3.1 Least Squares Support Vector Machine Algorithm (Ls-Svm)

For a given training sample set $\{x_i \in R_n\}$ and corresponding classification set $\{y_i \in \pm 1\}$, the decision function of LS-SVM in the feature space H is :

$$y = \mathrm{sgn}(w^T \varphi(x) + b) \tag{15}$$

In the nonlinear mapping $\varphi(x)$, the input samples x are mapped to the feature space H, the dimension of H which is $n \le \dim(H) \le \infty$. Classifier LS-SVM solves the following optimization problems:

$$\begin{aligned} & y = \mathrm{sgn}(w^T w + \frac{\gamma}{2}\sum_{i=1}^{l} e_i^2) \\ & s.t \;\; y_i = w^T \varphi(x_i) + b + e_i \;\; (i = 1,...,l) \end{aligned} \tag{16}$$

The Wolfe duality problem is as follows:

$$\begin{aligned} & \min \; (\frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j \varphi^T(x_i)\varphi(x_j) + \sum_{i=1}^{l} \frac{\alpha_i^2}{2\gamma} - \sum_{i=1}^{l} \alpha_i y_i) \\ & s.t. \; \sum_{i=1}^{l} \alpha_i = 0 \end{aligned} \tag{17}$$

The skewness term b is introduced in the objective function of formula (17), which can be obtained as follows:

$$\min \; (\frac{1}{2}\sum_{i,j=1}^{l} \alpha_i \alpha_j k(x_i, x_j) + \sum_{i=1}^{l} \frac{\alpha_i^2}{2\gamma} + b\sum_{i=1}^{l} \alpha_i - \sum_{i=1}^{l} \alpha_i y_i) \tag{18}$$

According to the Mercer kernel principle [2], any positive definite kernel function can be expressed as the inner product of two vectors in a characteristic space, so the positive definite kernel function $k(x, x_i)$ can be introduced into the least squares support vector machine. The common kernel function has the Gauss kernel function. For a new sample x, the following formula can be classified into its category:

$$f(x) = \mathrm{sgn}(w^T \varphi(x_i) + b) = \mathrm{sgn}(\sum_{i=1}^{l} \alpha_i k(x, x_i) + b) \tag{19}$$

3.2 A Fast Sparse Approximation Least Squares Support Vector Machine Based on Backward Fitting.

3.2.1 Fast Sparse Approximation Least Squares Support Vector Machine

Given a kernel function $k(x, y)$, it defines a basis function for each sample in the training set. The set of fundamental functions $D = \{k(x, x_i), i = 1, 2, ..., l\}$ is a kernel dictionary. Fast sparse approximation least

squares support vector machine (SVM) is a greedy algorithm which iteratively constructs decision functions by adding one base function from the core dictionary. Given the empty index set $P=\varnothing$ and index set $Q=\{1,2,...,l\}$, first select a new base function $k(x,x_S)$ from D according to a criterion, then delete the index s from Q and add it to the middle. When the new basis function is added, the subproblems composed of the new added basis functions and all the previously selected basis functions are solved. This process is repeated until the stop criterion is satisfied, and a fast sparse approximation least squares support vector machine [1] is obtained.

3.2.2 Reverse Fitting Strategy

Reverse fitting is not to change the Lagrange multiplier of the selected base function when solving the subproblem, and then selects the base function that can minimize the objective function. For fixed $i\in Q$, inverse fitting, the sub problem can be expressed as:

$$f_i^{n+1}=\begin{cases}\min & f(\alpha_i,b,\alpha_{p_1},\alpha_{p_2},...,\alpha_{p_n})\\ s.t. & \alpha_{Q-\{i\}}=0 \quad (i\in Q)\\ & \alpha_p=\alpha_p^n\end{cases} \tag{20}$$

Turn it into:

$$f_i^{n+1}=\min_{\alpha_i}\left(\frac{1}{2}(k(x_i,x_i))+\frac{1}{2\gamma})\alpha_i^2+r_i^n\alpha_i\right)\quad (i\in Q) \tag{21}$$ among

$$r_i^n=\begin{cases}-y, & n=0\\ \sum_{k=1}^{n}\alpha_{p_k}k(x_i,x_i)+b^n-y_i, & n>0\end{cases} \tag{22}$$ The optimal sol

ution is as follows: $f_i^{n+1}=-(r_i^n)^2/(2(k(x_i,x_i)+\frac{1}{2\gamma}))\quad (i\in Q)$ (22).In this way, we can find the index value of the base function to be added by using the following formula (23):

$$s=\min_{i\in Q}\left(-(r_i^n)^2/(2(k(x_i,x_i)+\frac{1}{2\gamma}))\right) \tag{23}$$

The outage criterion selected by this algorithm is: if the insensitive criterion $\max(abs(r_Q^n))<\varepsilon$ is established, the Non selected training samples can be predicted with less than the error, and the iteration is stopped. This stopping criterion is similar to the early stop [1] in the neural network algorithm, which is an effective mechanism to avoid over fitting.

Reverse fitting provides an effective sparse approximation strategy for least squares support vector machines. The process is as follows:

① $\alpha^0=0^T, b=0, r^0=-y, Q=\{1,2,3,...,l\}, P=\varnothing, n=0.$

② If $Q=\varnothing$ or $\max(abs(r_Q^n))<\varepsilon$, stop or continue.

③ $s=\arg\min\left(-(r_i^n)^2/(2(k(x_i,x_i)+\frac{1}{2\gamma}))\right).$

④If n=0, then $R^{n+1}=\begin{bmatrix}0 & 1\\ 1 & k(x_s,x_s)+\frac{1}{2\gamma}\end{bmatrix}^{-1}, \begin{bmatrix}b^{n+!}\\ \alpha_s^{n+1}\end{bmatrix}=R^{n+1}\begin{bmatrix}0\\ y_s\end{bmatrix}.$

⑤ If n=0, according to the formula $R^{n+1}=\begin{pmatrix}R^n & 0\\ 0^T & 0\end{pmatrix}+\lambda\begin{pmatrix}\beta\\ -1\end{pmatrix}(\beta^T-1)$ and formula $\begin{bmatrix}b^{n+1}\\ \alpha_p^{n+1}\\ \alpha_s^{n+1}\end{bmatrix}=\begin{bmatrix}b^n\\ \alpha_p^n\\ 0\end{bmatrix}+\lambda\left(\beta^T\begin{bmatrix}0\\ y_p\end{bmatrix}-y_s\right)\begin{bmatrix}\beta\\ -1\end{bmatrix}$, compute $R^{n+1},\alpha_p^{n+1},\alpha_s^{n+1},b^{n+!}$.

⑥ $Q=Q-\{s\}, P=P+\{s\}$

⑦ $r_Q^{n+1} = K_{QP}\alpha_P^{n+1} + b^{n+1}$

⑧ $n = n+1$, Return to ②.

4. Algorithm Simulation

4.1 Simulation Data

The simulation data of three kinds of aircraft (ANN -26, commendation, Jacques 42) are used in the experiment. The bandwidth of radar transmitting pulse is 400MHz (range resolution is 1m and radar radial sampling interval is 0.5m). The target data with different pitch angles are selected as training and testing data respectively. The training samples were 50, 50 and 50 respectively, and the test samples were 400, 500 and 400 respectively.

4.2 Results and Analysis

The experiment uses Gauss function to construct the kernel matrix and takes 4 different values of C, it is $2^i(i=-1,1,2,3)$, ε=0.5. The table (1) compares the data of the two methods of FSALS-SVM and LS-SVM in the classification accuracy (the mean of three types of target recognition rates); the Table 1 compares the data of the training time between the two methods of FSALS-SVM and LS-SVM.

Table 1. FSALS-SVM and LS-SVM two methods in classification accuracy data.

$\log_2(\gamma)$	FSALS-SVM	LS-SVM
-1	0.9743	0.9638
1	0.9752	0.9734
2	0.9816	0.9792
3	0.9831	0.9815

Table 2. FSALS-SVM and LS-SVM two methods in training time data

$\log_2(\gamma)$	FSALS-SVM/s	LS-SVM/s
-1	9.22	9.54
1	8.38	8.46
2	7.24	7.53
3	6.57	6.82

It can be found that FSALS-SVM is slightly higher than LS-SVM in terms of classification accuracy, while FSALS-SVM training time is obviously less than LS-SVM in training time. It can be seen that the FSALS-SVM algorithm can improve the training speed of LS-SVM by selecting the base function in the random subset. At the same time, because of the insensitive criterion, FSALS-SVM can obtain the sparsity of the decision function with a relatively low computational complexity.

5. Concluding Remarks

LS-SVM is an effective classification method, but it has shortcomings in radar target recognition. FSALS-SVM improves the approximation speed of LS-SVM by introducing insensitive stopping criteria from sparsity. Through the classification and recognition of radar target range profile data, the fast-sparse approximation least squares support vector machine has good generalization.

References

[1] Jiao Licheng, Zhou Weida, Zhang Li et al. Intelligent target recognition and classification, (Science Press, China 2010), p. 116-119.

[2] Xu Bin, Chen Bohai, Liu Hongwei, Jinlin. Based on the recurrent neural network model, radar high resolution distance image target recognition. Journal of electronics and information, vol. 38(2016), No. 12, p. 2988-2995.

[3] Li Hui, Jin Baolong, Zhai Haitian. High resolution radar signal translation invariant KPCA feature extraction algorithm. computer simulation, vol. 29(2012), No. 1, p. 9-12.
[4] Cheng Liang, Julie. Millimeter wave radiometer target recognition based on PCA optimized radial basis function neural network. Acta microwave Sinica, vol. 10(2015), No. 10, p. 225-229
[5] Sun Shaoyuan, Li Linna, Zhao Haitao. Depth estimation of monocular vehicle infrared images using KPCA and BP neural networks. infrared and laser engineering, vol. 42(2013), No. 9, p. 2348-2352.
[6] Yang Wenxiu, Fu Wen Xing, Zhou Zhiwei, Yu Wei, Ma Jie. Lidar rapid target recognition based on projection dimension reduction. infrared and laser engineering, vol. 43(2014), No. 12, p. 1-6.
[7] Yongsheng Xiao, Lizhen Huang, Jianjiang Zhou, RATR of adaptive angular-sector segmentation based on grey incidence analysis model. Grey Systems: Theory and Application, vol. 7(2017), No. 1, p. 71-79.
[8] Wei Cao, Hui Zhou, Zhimin Zhou, Zuowei Fu, An Approach for High Resolution Radar Target Recognition Based on BP Neural Network. International Conference on Intelligent Computing, ICIC 2011: Advanced Intelligent Computing pp 33-39.
[9] Daiying Zhou, Radar target HRRP recognition based on reconstructive and discriminative dictionary learning, Signal Processing, vol. 126(2016), No. 11, p. 52-64.