Advances in Engineering Research, volume 118
2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017)

Radar source identification method based on sample reduction and improved support vector machine

Chen Wengao[1,a], Jia Xin[1] and Tang Xiaojing[1]

[1]Equipment Academy, Beijing 101416, China.

[a]tingxue9278@163.com

Keywords: Radar emitter identification, Boundary extraction, Support vector machine

Abstract: Aiming at the problem of low efficiency of radar emitter identification method, a new method based on sample reduction and improved support vector machine is studied. Firstly, for removed redundant information, at the same time reduce the training data, the algorithm through the local normal vector to boundary extraction of sample prior information in the database. Then using the Sequential Minimal Optimization algorithm, multi classification and cross-validation to improve the original SVM. Through the improved algorithm train the reduced samples, and get the optimal model parameters. Finally using the optimal identification model to recognize the unknown pulse sequence information. Through simulation results and comparison, it is proved that the proposed radar source identification method based on sample reduction and improved support vector machine not only have high identification accuracy and robustness, but also have a good timeliness.

1. Introduction

Under the condition of the current high density and big data, the quadratic model optimization of SVM need large amount of computation, this can restrain the overall performance of the algorithm severely. To solve this problem, this paper proposed a radar emitter identification method based on sample reduction and improved SVM.

2. Sample reduction based on local normal vector

SVM algorithm need to select the appropriate data from the samples to restructure the support vector function, the appropriate data is only distributed on the contour of data set, the remaining data within the outline is redundant, without any contribution to SVM algorithm [1]. Therefore, in order to discard redundant data and reduce the computation of SVM, it is necessary to extract the contour of the sample dataset. Consider a two-dimensional projection with universal data distribution, including three typical boundary data points, as shown in figure 1.

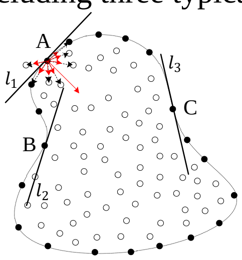


Fig. 1 Two-dimensional projection of the data set

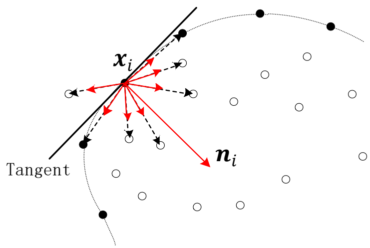


Fig. 2 The schematic of local normal vector generation

The solid dots in figure represent the contour data points, hollow dots are internal or external data points. A, B, C respectively represent three typical boundary distribution points, A point is located in the bulge of boundary, the hyper tangent plane l_1 divides internal data points into the same side; B, C points locate in the sunken boundary, the most points are divided into one side by the hyper tangent plane l_2, l_3, and the left points locate on the other side. In Figure 1, the A point is an example, and the unit vector and local normal vector of the nearest 7 data points are indicated. Observe all the vectors, we can find for the points of convex (including high-dimensional hyper convex) profile data set, corresponding to the unit vector and the local normal vector angle $\theta < \pi/2$ data points belonging to the same cluster [1]. Using the following method to determine whether data x_i is the boundary point or not:

$$p_i = \frac{1}{k}\sum_{j=1}^{k} l_j, \qquad l_j = \begin{cases} 1, & if\ \theta_{ij} \le \frac{\pi}{2}; \\ 0, & \text{else.} \end{cases} \tag{1}$$

For the boundary point x_i in Figure 2, if there is no effect on the outside interference data, $p_i = 1$. B and C in Figure 1, due to the contour concave region, there are few data points outside the super tangent plane, thus $p_i = 1 - \gamma$ (general γ value is smaller). Through the above analysis, if $p_i \ge 1 - \gamma$, corresponding to x_i is the boundary data points of dataset. After examining the dataset of noise reduction as above, simple boundary dataset can be obtained, and servicing for the follow-up SVM algorithm.

3. Construction of improved SVM model

Statistical learning theory is considered to be the best theory of finite sample statistics, and SVM is a machine learning method based on Structural Risk Minimization and VC dimension of this theory [2]. In the following example, the typical soft interval support vector machine C-SVM is introduced, and the model construction of SVM is introduced.

3.1 C-SVM model

Soft margin support vector machine C-SVM based on the original divided linear support vector machine, penalty factor C and slack variable ξ_i are introduced, allowing the abnormal value to a certain extent, is the most extensive model for SVM at present [3]. Under the guidance of the maximizing distance principle, the C-SVM classification problem can be described as:

$$\min\ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \qquad \text{s.t. } y_i\big((w \cdot x_i) + b\big) \ge 1 - \xi_i,\ \xi_i \ge 0, i = 1, \cdots, l \tag{2}$$

Where (x_i, y_i) is training sample, C is the penalty factor, ξ_i is the slack variable. After obtaining the decision function, we can identify the unknown information. But there are many problems can be found by analyzing the traditional C-SVM model. 1) When the training sample very large, solving convex quadratic programming problem will become extremely inefficient. 2) Only can handle two classification problem. 3) The selection of penalty factor and kernel function parameter without theoretical support. These above problems restrict the applicability and timeliness of the C-SVM algorithm seriously, so the following three methods are introduced to improve the traditional C-SVM algorithm.

3.2 Sequential Minimal Optimization-SMO

Sequential Minimal Optimization algorithm [2] is a heuristic algorithm for convex quadratic programming problems of SVM, analytic method can be used to solve this problem, thereby reducing the complexity of SVM greatly. The SMO algorithm consist of two parts, analytical method and variable selection, not repeat them here.

3.3 Multi class processing

SVM algorithm is designed to solve the problem of two classification, so if without improved, the multi classification problem cannot be solved, so as to limit the applicability of the algorithm greatly. In view of this problem, using the method of one-to-multi kind to solve the problem [4].

3.4 Cross validation

Firstly the sample dataset T is divided into k disjoint subsets $\mathrm{T} = T_1 \cup T_2 \cup \cdots \cup T_k, T_i \cap T_i = \Phi, i, j \in [1, k], i \ne j$ randomly and almost equal interval. For each cross validation, T_i is chosen as the training set, and the remaining subset as the test dataset. Using the decision function recognize the test dataset, comparing the true class label, recording the number of false identification samples as $\mathrm{e_i}$. At the end of the k cross validation, the kernel function width and penalty factor are selected as the optimal model parameters of decision function with the minimum number of the wrong identification samples.

4. Improved identification process

The method of Radar Emitter Identification Based on sample reduction and improved SVM is mainly divided into three parts: training sample reduction, training learning phase and identification decision stage. The process is shown in the following figure:

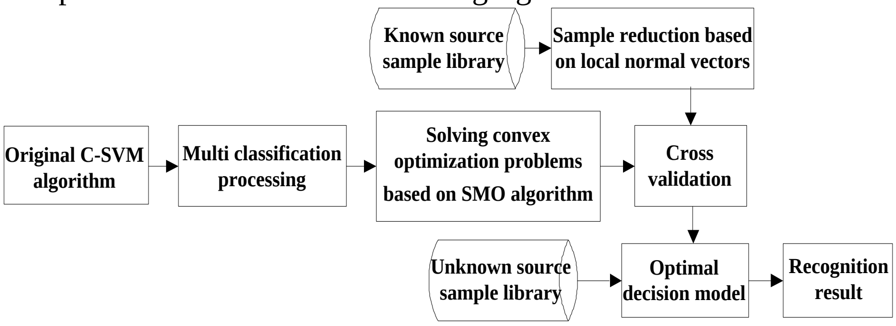


Fig. 3 The process of radar emitter identification based on sample reduction and improved SVM

5. Simulation experiment analysis

The RF, DOA, PW,PA, PRI, BW and the type of parameter are constructed the known radar source database, with 9 different kinds of parameters, the number of samples of each type are different, a total of 10000 samples. The parameters of each radar sources are shown in table 1.

(1)Identification capability analysis

In order to verify the ability of identification, in accordance with table1, 9 types of radar source parameter are generated respectively, it contains 4 test data, and each test data contain 9753 samples, random order, and have different parameter errors, the experimental process as follows.

Table 1 Known radar emitter sample characteristic information table

Radar	DOA/(°)	PA	PW/(μs)	RF/(MHz)		PRF/(Hz)		BW/(Hz)	Sample number
				Type	Range	Type	Range		
1	48-50	5-15	1.2-1.4	fixed	2104-2105	Group variable	580/680/740	10.3	1025
2	60-63	3-12	0.2-0.4	Agile	2750- 2850	Shake	360-540	7.3	1026
3	68-70	16-20	6.8-6.9	Group variable	2282/ 2297	Stagger	800/850/900/950/1000	18.6	1149
4	56-58	6-13	13.3-13.5	Agile	1550-1750	Shake	700-900	39.3	1091
5	30-32	33-38	6.6-6.8	fixed	1925-1926	Group variable	810/830/880	50.1	1104
6	49-51	23-31	3.3-3.4	Group variable	1925/1926	fixed	3499-3500	11.6	1122
7	31-33	50-56	60.1-60.3	Group variable	2202/2217/2263	fixed	2988-3000	13.8	1068
8	65-68	10-23	61-61.2	Group variable	3278/3282/3297	fixed	3199-3201	16.6	1045
9	36-38	40-50	61-61.1	fixed	2198-2199	fixed	1562-1564	10.2	1370

For the radar source sample information after normalization, using local vector method to eliminate redundant information, through many simulation experiments in order to achieve the purpose of reduction of the sample data, and try not to damage the contour information, set the parameters k=30, gamma =0.04 to extract boundary data. Then, to get the best parameters of the support vector model, the improved C-SVM algorithm is used to cross validation of the reduced samples. The cross validation process is shown in Fig 4.

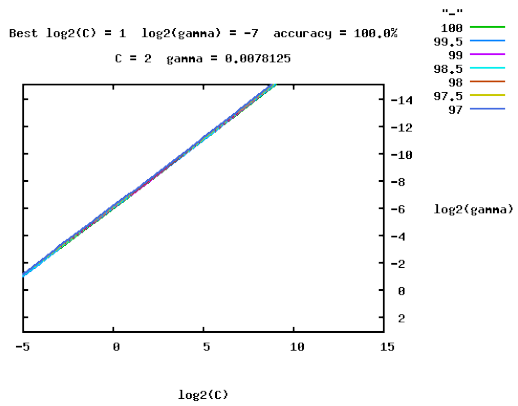


Fig.4 The process of Radar Emitter Identification Based on sample reduction and improved SVM

From Fig 4, the cross validation process when the penalty parameter C=2 and kernel width of g=0.0078125, the identification rate can reach 100%. Thus this model is selected as support vector decision model, increase the parameter error gradually, the identification results are shown in table 2.

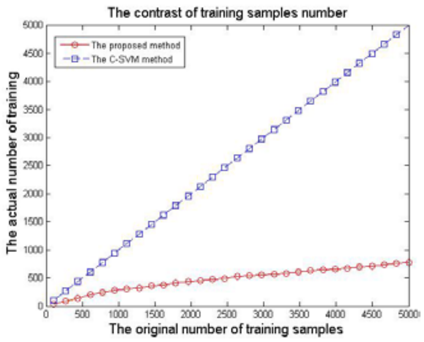
Table 2 Identification result statistics

Experimental batch	1	2	3	4
Parameter error (%)	0	10	20	30
Kernel width g		0.0078125		
Penalty factor C		2		
Correct identification of samples	10000	9745	8377	7188
Total number of test samples		10000		
Correct rate of identification (%)	100	97.45	83.77	71.88

From the identification results, when there is no error, the correct rate of identification can be to 100%, with increase the test parameter error, the identification accuracy rate decreased gradually, when the parameter error reach 30%, the algorithm still can achieve 71.88% identification accuracy. Under actual conditions, the parameter error of reconnaissance receiver is generally less than 10%, so the proposed identification method can achieve more than 97.45% accuracy, proved that the algorithm has advantages of high robustness and high identification accuracy.

(2)Timeliness comparative analysis

In order to verify the timeliness of proposed method, in accordance with9 radar sources parameter information of table1, simulated training and test samples, where of 9753 test samples, the parameter error is 10%. Using the proposed method and C-SVM algorithm to learn the training data respectively with the number of training samples increased gradually from 100 to 5000. Then using support vector decision function to recognize the test data with the optimal model parameters. The number of samples to participate in the training are shown in Fig 5, and the identification results are shown in Fig 6 and Fig 7.



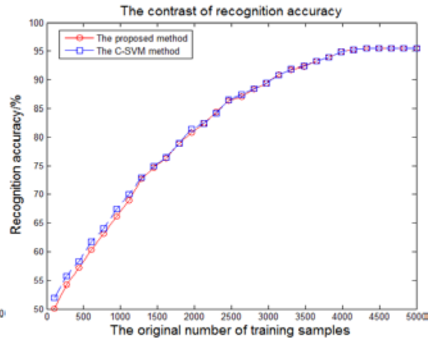


Fig.5 The number of training samples Fig.6 Contrast of training process Fig.7 Identification accuracy omparison

By comparing the identification process and results of figure 5~7 can be find that in the training samples, the C-SVM algorithm without the reduction of the sample data, with the increase of training samples, the computation time increases sharply. In this paper, a large number of redundant information is eliminated by boundary extraction, which reduces the sample data greatly, thus reducing the training time. Thus, the proposed method not only reduce the training time, but also achieve the same identification accuracy as the C-SVM algorithm.

6. Conclusions

In view of the poor timeliness problem of traditional radar source identification methods, this article studied a radar source identification method based on sample reduction and improved SVM. The method through the sample reduction to eliminate redundant information and reduce the training sample size, without affecting the identification accuracy and reducing the training time. Through simulation and analysis, it is proved that the proposed method not only has high identification accuracy and robustness, but also has good timeliness.

7.Reference

[1] Li Y. Selecting training points for one-class support vector machines[J]. Pattern Identification Letters, 2011, 32(11): 1517-1522.

[2] Chang Tiantian. Research on some problems of support vector machine learning algorithm [D]. Xidian University, 2010: 23-25.

[3]Wang Tinghua. Research on model selection for support vector machine [D]. Beijing: Beijing Jiaotong University, 2009: 36-37.

[4] Deng Naiyang, Tian Yingjie. New method in data mining-support vector machine [M]. Science Press, 2004: 214-216.