

Signal Classification Method Based on Support Vector Machine and High-Order Cumulants

Xin ZHOU, Ying WU, Bin YANG

Zhengzhou Information Science and Technology Institute, Zhengzhou, China

Email: zx007_0_0@126.com

Received September 15, 2009; revised November 13, 2009; accepted November 18, 2009

Abstract

In this paper, a classification method based on Support Vector Machine (SVM) is given in the digital modulation signal classification. The second, fourth and sixth order cumulants of the received signals are used as classification vectors firstly, then the kernel thought is used to map the feature vector to the high dimensional feature space and the optimum separating hyperplane is constructed in space to realize signal recognition. In order to build an effective and robust SVM classifier, the radial basis kernel function is selected, one against one or one against rest of multi-class classifier is designed, and method of parameter selection using cross-validation grid is adopted. Through the experiments it can be concluded that the classifier based on SVM has high performance and is more robust.

Keywords: High-Order Cumulants, Support Vector Machine, Kernel Function, Signal Classification

1. Introduction

Automatic modulation classification (MC) is an intermediate step between signal detection and demodulation, and plays a key role in various civilian and military applications. It is also one of many key technologies in software radio and cognitive radio.

The recognition methods in early years are mainly about signal waveform, frequency, transient amplitude and transient phase [1]. The performances of these methods descend quickly when they face to low SNR. Statistical decision and pattern recognition based on statistics are two main methods in approaching MC problem in recent years [2]. The first method is based on hypothesis testing; problem it has to face is that needs to give proper hypothesis and strict data analysis to get the correct decision threshold. The Reference [3] uses neural net to solve MC problem and gets better effect. But because the sample length is limit, the neural net is easy to bring the phenomenon of over-learning and local minimal value. There are some researchers use support vector machine (SVM) to solve MC problem, and get higher classification accuracy [4,5]. But in the two references they neither gave how to select the optimal parameter of SVM classifier and how to construct multi-class SVM. In this paper, we introduce the support vector machine firstly, then research the selection methods of kernel function and its

parameter, and study on the multi-classes classification methods, and then apply them to digital signal classification. We also compare the SVM with other common classifiers.

The paper is organized as follows: In Section 2, the robust feature extraction based on high-order cumulants is proposed. In Section 3, the multi-classifier based on SVM is designed. The principle of SVM is introduced firstly, then the kernel and parameter selection are given, the method of decomposing multi-class classifier is used. In Section 4, we input the signal feature to multi-class SVM classifier to do experiment. In Section 5, the paper is concluded.

2. Feature Extraction Based on High-Order Cumulants

High-order cumulant is a tool of mathematics which describes the high order statistical characteristic of random process. It not only can remove the influence of Gauss noise, but also is robust to the rotation and excursion of the constellation diagram.

We suppose the classifier works in the interrelated and synchronization environment. The received signal has carried out carrier frequency synchronization and timing synchronization, but the unknown referenced phased offset exists. The output signal of receiver can be

Table 1. The cumulants of signals.

signal	$ C_{40} $	$ C_{42} $	$ C_{63} $	$\frac{ C_{40} }{ C_{42} }$	$\frac{ C_{63} ^2}{ C_{42} ^3}$
4ASK	$1.36E^2$	$1.36E^2$	$9.16E^3$	1	33.36
2PSK/ 2ASK	$2E^2$	$2E^2$	$13E^3$	1	21.125
4PSK	E^2	E^2	$4E^3$	1	16
2/4/8FSK	0	E^2	$4E^3$	0	16
16QAM	$0.68E^2$	$0.68E^2$	$2.08E^3$	1	13.76

Table 2. The cumulants of FSK signals.

signal	$ C_{21} $	$ C_{42} $	$\frac{ C_{21} ^2}{ C_{42} }$
2FSK	$E\Delta\omega^2$	$E^2\Delta\omega^4$	1
4FSK	$5E\Delta\omega^2$	$9E^2\Delta\omega^4$	2.78
8FSK	$21E\Delta\omega^2$	$105E^2\Delta\omega^4$	4.2

expressed as:

$$\begin{aligned} r(i) &= s(i) + n(i) \\ &= \sum_{k=1}^L \sqrt{E} a_k e^{j\theta_c} p(i-k) + n(i) \end{aligned} \quad (1)$$

where a_k is the sending symbol sequences, L is the observational symbol number, E is the signal average power, θ_c is referenced phase, $p(i)$ is channel remnant answer, $n(i)$ is assumed to be complex white Gaussian noise with power σ^2 .

Suppose the emanant signal serial is independent and identically distributed, the different average power has been normalized to 1, the ideal high-order cumulants of these signals can be expressed by Table 1 [6].

Because we calculate the high-order cumulants can not identify 2FSK, 4FSK and 8FSK signal directly, the ratio of $|C_{21}|^2$ and $|C_{42}|$ get from each signal in Table 2 is the signal after difference through median filter which is used to classify FSK signals, where $\Delta\omega$ is frequency offset.

3. The Classifier Based on Support Vector Machine

3.1. Support Vector Machine (SVM)

SVM is basically a two-class classifier based on the idea of “large margin” and “mapping data into a higher dimensional space” [7]. The principle of SVM is to make minimize the structure risk, in the high dimensional feature space, find an optimal discriminant hyperplane with low VC dimension to make the distance between the two

classes’ data have large margin. When the feature space is not linear dividable, SVM maps the data into high dimensional feature space with non-linear mapping, and finds the optimal classification hyperplane in high dimensional feature space.

Based on the principle of configuration risk minimization, suppose in inner product space F exists two kinds discriminable samples $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, $\mathbf{x}_i \in R^n$, $y_i \in \{-1, +1\}$, $i = 1, 2, \dots, n$. -1 and +1 denote two kinds; the optimal classification hyperplane can be expressed as:

$$\{\mathbf{x} \in F : (\mathbf{w} \cdot \mathbf{x}) + b = 0\} \quad (2)$$

where \mathbf{w} is support vector, b is translation vector. In order to make classification hyperplane and \mathbf{w} one-to-one correspondence, we standardize it and let the distance of the sample which is nearest to hyperplane is $1/\|\mathbf{w}\|$. So hyperplane after standardization satisfies:

$$\min_{i=1,2,\dots,n} |(\mathbf{w} \cdot \mathbf{x}_i) + b| = 1 \quad (3)$$

Solving the optimal classification hyperplane can be transformed into quadratic optimization problem:

$$\begin{aligned} \min \quad & \Phi(\mathbf{w}) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) \\ \text{s.t.} \quad & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{aligned} \quad (4)$$

The optimal hyperplane is discussed on the condition that samples can be classified linearly, if can not, we will use slack variables $\xi_i \geq 0$ and penalty factor C to resolve generalized optimal classification hyperplane (to classify samples farthest and make the largest classify margin at the same time):

$$\begin{aligned} \min \quad & \Phi(\mathbf{w}) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & \begin{cases} y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i \geq 0 \\ \xi_i \geq 0 \end{cases} \end{aligned} \quad (5)$$

where $i=1,2,\dots,n$, C is a certain constant, it is the control of the punishment of samples which are classified mistakenly. It is a compromise between the proportion of false classified samples and algorithm complexity.

According to the equation above and Lagrange theorem, use Kuhn-Tucker condition, the (5) can be transformed into duality problem:

$$\begin{aligned} \max_{\alpha} \quad & Q(\alpha) = -\frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j (\varphi(\mathbf{x}_i) \varphi(\mathbf{x}_j)) + \sum_{i=1}^n \alpha_i \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0, \quad i = 1, 2, \dots, n \\ 0 \leq \alpha_i \leq C \end{cases} \end{aligned} \quad (6)$$

Use kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$, the quadratic problem can be represented by [8]:

$$\max_{\alpha} Q(\alpha) = -\frac{1}{2} \sum_{j=1}^n \sum_{i=1}^n \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^n \alpha_i$$

$$s.t. \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \end{cases}, \quad i=1, 2, \dots, n \quad (7)$$

The classification threshold b can be gotten by any support vector use (8):

$$\begin{cases} \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\ y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \end{cases}, \quad i=1, 2, \dots, n \quad (8)$$

The optimal classification discriminant function expressed by kernel function is:

$$f(\mathbf{x}) = \text{sgn}\{(\mathbf{w} \cdot \mathbf{x}) + b\} = \text{sgn}\left\{ \sum_{\mathbf{x}_i \in SV} \alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) + b \right\} \quad (9)$$

where $\mathbf{w} = \sum_{\mathbf{x}_i \in SV} \alpha_i y_i \phi(\mathbf{x}_i)$.

According to optimal problem (7), the complexity of SVM has nothing to do with dimension of feature, but is restricted by the number of samples. SVM needs to compute the kernel functions between every two training samples, to generate a kernel function matrix which has n^*n elements, and n is the number of training samples.

3.2. The Selection of Kernel Function

In fact, changing kernel parameter is to change mapping function implicitly, and change the complexity of samples' distribution in feature space. So the selection of kernel function and parameters are very important. There are 3 kinds of kernels that are usually used [8]:

1) Dimensional polynomial kernel of degree d , the expression is:

$$k(\mathbf{x}, \mathbf{y}) = [(\mathbf{x}, \mathbf{y}) + p]^d \quad (10)$$

where p and d are custom parameters; If $p=0$ and $d=1$, it is called linear kernel function. The operation speed of kernel function is fast.

2) Radial basis function kernel, the expression is:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2}\right) \quad (11)$$

where $\sigma^2 > 0$, it controls the width of kernel function

and needs to be confirmed.

3) Neural Network kernel function, the expression is:

$$k(\mathbf{x}, \mathbf{y}) = \tanh(\mu(\mathbf{x}, \mathbf{y}) + \nu) \quad (12)$$

where μ and ν are parameters. Only some values satisfy Mercer condition can be used.

Because the feature space of radial basis function kernel is limitless, the limit samples in this feature space must be linearly discriminable, so it is most commonly used in classification. In this paper, we also select radial basis function kernel.

3.3. The Parameter Selection of SVM

In SVM classifier, the parameter selection of kernel function and penalty factor is very important. The penalty factor C is the optimal compromise with the distance between hyperplane and the nearest training point is farthest and the classification error is least. The parameters of kernel function determine the data mapping into higher dimensional space.

There are many parameter selection methods, such as grid searching, GD algorithm, gradient descent algorithm, genetic algorithm, simulated annealing algorithm and so on. The parameter evaluation criterion has k-fold cross-validation, leave-one-out (LOO), generalized approximate cross-validation (GACV), approximate span bound, margin-radius bound and so on.

In this paper, we use k-fold cross-validation to select parameter C , γ ($\gamma = 1/\sigma^2$) of RBF-SVM. Suppose we have n known samples, they construct sample set (\mathbf{x}_i, y_i) , $i=1, 2, \dots, n$, $y_i \in \{-1, +1\}$. In order to differentiate kernel function, we use l express the k value of the k-fold cross-validation. The steps of k-fold cross-validation are as followed:

1) Divide sample set contains n samples to l subset equally, each subset contains n/l samples.

2) Put from the first to $(l-1)$ subset of $(l-1)n/l$ samples as training ones, give a smaller value of parameter (C , γ), put in (7) and get the solution of Lagrange operator α_i^* , the samples corresponding to α_i^* which are more than zero are support vectors.

3) Put each α_i^* into:

$$\|\mathbf{w}\|^2 = \mathbf{w}^{*T} \mathbf{w}^* = \sum_{i,j=1}^{(l-1)n/l} \alpha_i^* \alpha_j^* y_i y_j \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

$$b^* = \frac{1}{|SV|} \sum_{k \in SV} \left(y_k - \sum_{i=1}^{(l-1)n/l} \alpha_i^* y_i \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_k\|^2\right) \right)$$

where $|SV|$ is the number of support vector, and the cal-

ulation of classification threshold b^* uses the mean of each SVM.

4) Put the α_i^* , b^* and test samples \mathbf{x}_u , $u = (l-1)n/l+1, \dots, n$ into classification function (9) to get the output $f(\mathbf{x}_u)$ of each kind, to validate whether $f(\mathbf{x}_u)$ is in accordance with real output y_u .

5) Take from the second to l subset of $(l-1)n/l$ samples as training ones, the first subset as test samples, repeat the steps 2)–4). According to the proposed mechanism, until all subsets are tested, it also repeats the above steps l times and calculates the accuracy of cross-validation.

6) Fix the parameter C , first increase γ gradually, repeat steps 2)–5); then increase γ gradually, repeat steps 2)–5) to get different accuracy of different parameter (C, γ) . The experience expresses that parameters increase as exponent is more effect.

7) Get the max validation classification accuracy and corresponding (C, γ) , if the accuracy satisfies requirement, then go to step 8); Or search in the range of (C, γ) continually which is gotten by the maximum validation accuracy and (C, γ) gotten by the second maximum validation accuracy. It also repeats step 2)–6) until satisfies the accuracy.

8) Use the satisfied parameter (C, γ) to train all training samples and get the final optimal parameter α_i^* ($i \in sv$) and b^* , then determine the optimal classification function.

3.4. The Design of Multi-Classifier

There are two ideas to solve multi-class classification problem of SVM [9]: One is to properly change the original optimal problem in order to compute all the multi-class classification discriminant functions at the same time; the other is to divide the multi-class problem into a series of binary problems which can be solved directly, and based on the results, gain the final discriminant results.

The first idea seems to be simpler, but because its computation is too complex and costly, and also hard to implement, it is not widely used. There are 5 kinds of multi-class methods based on the second idea: One Against Rest (OAR), One Against One (OAO), Binary Tree (BT), Error Correcting Output Code (ECOC), Directed Acyclic Graph (DAG). The OAO and OAR methods are often used.

4. Computer Simulation and Performance Analysis

4.1. Experiment Steps

The steps of signal classification of SVM based on grid searching parameters selection are as follows:

1) Extract cumulant features of the received signals, divide the feature vectors equally to training samples and test samples.

2) Select RBF kernel function and a certain multi-classifier design method; initialize σ^2 and C ; give the parameter search range, use k-fold cross-validation to get the optimal parameter of SVM.

3) Set the optimal parameter according to step 2) of RBF-SVM and train it using training samples.

4) After training, input features of await classification signals to classify them.

4.2. Classification Experiment

Parameter selection experiment: we create 200 every digital signal every 2dB from 0 to 20dB in awgn channel, extract cumulant feature and get new sample serial. Samples of each class are separated into training ones and test ones randomly. We use SVM one-against-one decomposition, chose RBF kernel, initialize $C_0 = 10$, $\sigma_0^2 = 10$ and disperse the parameter logarithmically, get the grid value $(\log \sigma^2, \log C)$. Where $\log C \in \{\log C_0 - 3, \log C_0 - 2, \dots, \log C_0 + 3\}$,

$\log \sigma^2 \in \{\log \sigma_0^2 - 3, \log \sigma_0^2 - 2, \dots, \log \sigma_0^2 + 3\}$. Then the isolines of classification accuracy are shown in Figure 1. The maximum accuracy is 99.1% and the optimal parameters are $C = 0.01$, $\sigma^2 = 0.1$.

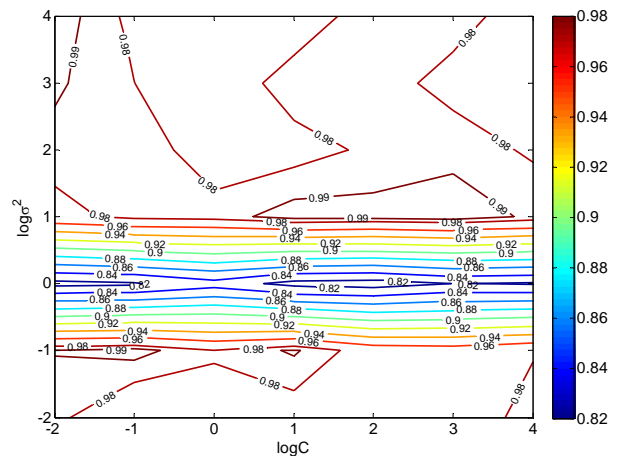


Figure 1. The cross-validation isolines of OAO-RBF-SVM.

Table 3. The simulation result at 5dB.

input	output(classification accuracy)						
	4ASK	2PSK/ 2ASK	4PSK	2FSK	4FSK	8FSK	16QAM
4ASK	92.2	7.8	0	0	0	0	0
2PSK/ 2ASK	100	0	0	0	0	0	0
4PSK	0	0.6	99.4	0	0	0	0
2FSK	0	0	0	100	0	0	0
4FSK	0	0	0	2.6	97.4	0	0
8FSK	0	0	0	0.4	2.6	97.0	0
16QAM	0.8	4.2	5.8	0	0	0	89.2

Table 4. Comparison of different classification methods.

Classifier	Classification accuracy(%)
Nearest distance classifier	80.2
Neural network	85.6
OAR-SVM $\sigma^2=0.01, C=1$	90.4
OAO-SVM $\sigma^2=0.01, C=0.1$	92.2

Test 1: In this experiment, we get the classification accuracy of different signals in awgn channel. The sample frequency is 40kHz and carrier frequency is 8kHz. The length of signal is 1200, the symbol rate is 2000Bd, and we do 500 Monte Carlo experiments at 5dB. The OAO-SVM is used to get the classification accuracy in Table 3.

From Table 3 we can see that SVM classifier can get higher classification accuracy at 5dB. The QAM classification accuracy is lowest and is 89.2%. This is because the feature extraction of QAM is close to the feature of 2PSK and 4PSK, so it is easy to judge to the two signals mistakenly.

Test 2: In this experiment, we compare SVM, neural network and the nearest distance discrimination classifier. The simulation assumption is the same as test1. We calculate the classification accuracy of 4ASK at 5 dB. The SVM uses RBF kernel and OAR and OAO classification algorithm, and then we do 500 Monte Carlo experiments. The classification accuracy is shown in Table 4.

From Test 2 we can see that the classification accuracy of the nearest distance classifier is lowest, and then is neural network and SVM is highest.

5. Conclusions

In this paper, we use the kernel thought of statistical learning theory for reference and use decomposition me-

thods of multi-class classifier and method of parameter selection using cross-validation grid search to build effective and robust SVM classifiers. We also use fourth and sixth cumulants of the received signals as the classification vectors, to realize digital signals classification. From the computer simulation and analysis, we can get the following conclusion:

1) The feature vector of cumulants can remove the influence of Gauss noise. It is robust and has high performance.

2) Classification method based on kernel function is less affected by dimension of input data. The classification capability of kernel classifier is affected by the kernel function and parameters, and a fine classification precision can only be obtained when kernel parameters are in special range. The classification stability can be effectively improved by parameter selection via cross-validate grid search method. If the proper parameters are chosen, the classification accuracy of SVM is high.

6. References

- [1] K. Nandi and E. E. Azzouz, "Automatic modulation recognition [J]," *Signal Processing*, Vol. 46, No. 2, pp. 211–222, 1995.
- [2] O. A. Dobre, A. Abdi, Y. Bar-Ness, *et al.*, "Survey of automatic modulation classification techniques: Classical approaches and new trends [J]," *IEE Communication*, Vol. 1, No. 2, pp. 137–156, 2007.
- [3] W. C. Han, H. Han, L. N. Wu, *et al.*, "A 1-dimension structure adaptive self-organizing neural network for QAM signal classification [C]," *Third International Conference on Natural Computation (ICNC 2007)*, HaiKou, August 24–27, 2007.
- [4] X. Z. Feng, J. Yang, F. L. Luo, J. Y. Chen, and X. P. Zhong, "Automatic modulation recognition by support vector machines using wavelet kernel [J]," *Journal of Physics, International Symposium on Instrumentation Science and Technology*, pp. 1264–1267, 2006.
- [5] H. Mustafa and M. Doroslovacki, "Digital modulation recognition using support vector machine classifier [C]," *Proceedings of The Thirty-Eighth Asilomar Conference on Signals, Systems & Computers*, November 2004.
- [6] O. A. Dobre, Y. B. Ness, and S. Wei, "Higher-order cyclic cumulants for high order modulation classification [C]," *IEEE MILCOM*, pp. 112–115, 2003.
- [7] Z. L. Wu, X. X. Wang, Z. Z. Gao, and G. H. Ren, "Automatic digital modulation recognition based on support vector machine [C]," *IEEE Conference on Neural Networks and Brain*, pp. 1025–1028, October 2005.
- [8] V. Vapnik, "Statistical learning theory [M]," Wiley, 1998.
- [9] B. Gou and X. W. Huang, "SVM multi-class classification [J]," *Journal of Southern Yangtze University*, Vol. 21, pp. 334–339, September 2006.