Automatic Modulation Classification Using Support Vector Machines and Error Correcting Output Codes

Jie Li[1], Qingda Meng[2], Ge Zhang[1], Yang Sun[1], Lede Qiu[2], Wei Ma[1]
1. China Academy of Space Technology (Xi'an), 710000 Xi'an, China
2. Institute of Telecommunication Satellite, CAST, 100094 BeiJing, China
xnkjlj@163.com, ralon@163.com, shanhualj@126.com

***Abstract*— A new approach, which because of support vector machines (SVMs) and error correcting output codes (ECOC), is proposed for automatic digital modulation classification (MC). Combining statistics and spectral feature sets, SVM classification and ECOC techniques are used to improve the generalization ability of classifiers. In contrast to the popular artificial neural networks (ANN) approach, this method avoids the problem of local optimum and requires less training data. Simulation results show that the proposed method is more efficient than other methods in low signal-to-noise ratio.**

Keywords—Modulation Classification; Support Vector Machines; Error Correcting Output Codes;

I. INTRODUCTION

Automatic digital modulation classification has been increasing requirement now. MC is first used in many military communication intelligence applications such as interference identification, source identification, spectrum surveillance and so on. Otherwise, MC technique, which is indispensable to choose the appropriate demodulator automatically, plays an important part in the multimode communication system in many novel applications in recent years.

MC is a pattern discriminate problem. There are different suggestions for this problem. Considering the independent and identical distribution, Maximum-likelihood (ML) [1] methods have been proposed and well studied, but they suffer from very high computational complexity and are not reliable with model mismatch such as phase and frequency offsets, residual channel influence, timing errors [2]. On the other hand, artificial neural networks (ANN) can be used for this purpose was welcome in the 90s [3] of the last century, but there were problems, too, except that the problems involved were local minima, and so on, [4].

This paper tell us a new MC way based on the SVMs and ECOC technique. Digital modulation method thought over in the paper include 2ASK, 4ASK, 2FSK, 4FSK, BPSK, QPSK, 8PSK, MSK, 8QAM, 16QAM and 64QAM. Simulation results show that the innovative methods in this paper is more flexible, more reliable and more efficient than before.

II. PROBLEM STATEMENT

The objective of modulation classification is to identify the modulation type of a modulated signal corrupted by noise and other channel effects. A digitally moduated signal can be wrote as：

$$s(\mathrm{x}) = A_m(\mathrm{x})\cos(2\pi f_m(\mathrm{x}) + \phi_m(\mathrm{x})) \qquad (1)$$

here $A_m(x)$, $f_m(x)$ and $\Phi_m(x)$ are the signal ampitude, signal frequency and signal phse, rspectively, in accordance with approprate modultion technques.We have considered four major digital modulation techniques in our study, that is, including the 11 types of interrogations, PSK, FSK, and QAM.

Suppose we work in a same conditions with carrier, andsingle-tone signaling, timing, and waveform recovery have been achieved. After preprocesing, we get a series of base band complex signals that can be show:

$$y(\mathrm{a}) = x(\mathrm{a}) + w(\mathrm{a}) \qquad (2)$$

where x (a) represents a digitally modulated signal, w (a) represents the accessory food Gauss noise sequence. x(a) also w(a) are not dependently and dentically dstributed (i.i.d.) also jontly uncorrelted stochstic sequnces.

III. FEATURE EXTRACTION

In order to weaken the scale of the original data set, pattern recognition systems typically drawt some different features before classification. Although there are several feature extraction methods. But they are different, the search using different characteristics in MC. In the text, we use a caller feature set with the statstical feature subst and spctral feature subset for classification.

A. Statistical feature set

As we can see, in this setting, it is natural to use higher-order statistics because they describe the shape of the noise base band samples [4]. Given a received baseband signal y(a), its second cumulants, fourth cumulants and sixth cumulants could be defined as follows respectively. The statistical feature subset used in the study is an extension of the feature set proposed in [1].

$$C_{20} = E[y^2(\mathrm{a})]$$
$$C_{21} = E[|\,\mathrm{y(a)}\,|^2]$$
$$C_{40} = cum(y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}))$$
$$C_{41} = cum(y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}), y^*(\mathrm{a}))$$
$$C_{42} = cum(y(\mathrm{a}), y(\mathrm{a}), y^*(\mathrm{a}), y^*(\mathrm{a}))$$
$$C_{60} = cum(y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}), y(\mathrm{a}))$$
$$C_{63} = cum(y(a), y(a), y(a), y^*(a), y^*(a), y^*(a)) \qquad (3)$$

978-1-5090-6414-4/17/$31.00 ©2017 IEEE

The cumulants defined in (3) can be obtained from the sample pass through the formula given below.

①Delete the average of y(a).

②Compute sample estimates of C20、C21

$$\hat{C}_{20} = \frac{1}{N}\sum_{a=1}^{A} y^2(a)$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{a=1}^{A} |y(a)|^2 \quad (4)$$

where a is the distance of samples and ^ show sample average.

③Using the result of formula (4) and definition (3) we get the estimates of 4th and 6th cumulants by the subsequent formulas.

（Note: the variable n has the same meaning as the variable a. Instead, n is used instead of a）

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} y^4(n) - 3\hat{C}_{20}^2$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} y^3(n)y^*(n) - 3\hat{C}_{20}\hat{C}_{21}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2$$

$$\hat{C}_{60} = \frac{1}{N}\sum_{n=1}^{N} y^6(n) - 15\frac{1}{N}\sum_{n=1}^{N} y^4(n)\hat{C}_{20} + 30\hat{C}_{20}^3$$

$$\hat{C}_{63} = \frac{1}{N}\sum_{n=1}^{N} |y(n)|^6 - 9\frac{1}{N}\sum_{n=1}^{N} y^3(n)y^*(n)\hat{C}_{21} - 6\hat{C}_{21}^3 \quad (5)$$

④Compute the normalized cumulants.

$$\bar{C}_{4k} = \hat{C}_{4k}/\hat{C}_{21}^2 \text{, k=0、1、2}$$

$$\bar{C}_{6k} = \hat{C}_{6k}/\hat{C}_{21}^3 \text{, k=0、3} \quad (6)$$

B. Spectral feature set

Azzouz and Nandi[2] [4] [5]put forward a kind of suitable for modulation recognition of spectral feature set, it contains instantaneous amplitude, instantaneous frequency and instantaneous phase information hiding. In the analysis process, it serves as a feature subset of the classifier. It has five characteristics as follows:

①Maximum value of the power spectral density of the instantaneous amplitude:

$$\gamma_{\max} = \max |DFT(a_{cn}(i))|^2 / X_s \quad (7)$$

where X_s is the number of samples and $a_{cn}(i)$ is the ith normalized instantaneous amplitude.

②Standard deviation of entire integrity of instantaneous amplitude:

$$\sigma_{a\alpha} = \sqrt{\frac{1}{N_s}(\sum_{i=1}^{N_s} a_{cn}^2(i)) - (\frac{1}{N_s}(\sum_{i=1}^{N_s}|a_{cn}(i)|))^2} \quad (8)$$

③ Standard deviation of the entire complete of the normalized instantaneous frequency

$$\sigma_{af} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} f_N^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t}|f_N(\mathrm{i})|)^2} \quad (9)$$

here C is the samples quantity on{Φ_N(i)} if $a_n(i)>a_t$ and a_t at the beginning of a(i) under then the general idea of the instantaneous phase is can be done.

④Standard deviation of the absolute valuation of centered nonlinear components of the instantaneous phase

$$\sigma_{ap} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} \Phi_{NL}^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t}|\Phi_{NL}(\mathrm{i})|)^2} \; . \quad (10)$$

⑤Standard deviation of the exteriort value of the middle nonlinear component of the instantaneous phase

$$\sigma_{ap} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} \Phi_{NL}^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t}\Phi_{NL}(\mathrm{i}))^2} \; . \quad (11)$$

IV. Digital Modulation Classification Using ECOC-SVMs

The proposed MC method can be shown in figure 1 which includes four main parts. We will discuss the final part in detail in the following section and give an overview of two-class Support Vector Machines at first.

A. Binary classification using Support Vector Machines

Let m-dimensional training data be x_i (i=1,2,…,M) and $y_i \in \{-1,+1\}$ is the class labels of x_i, where y_i = 1 and y_i = -1 respectively denote x_i for Classes 1 and 2. If the input data are linear separable in feature space, will the decision function that we can get is:

$$\mathrm{D(t)} = \mathrm{w}^{\mathrm{X}}\mathrm{g(t)} + \mathrm{b} \quad (12)$$

where g(t) is a kernel function that maps t into the l-dimensional space, w is an l-dimensional vector , b is a scalar, and

$$\mathrm{y_i(w^T g(t_i) + b) \geq 1 \;\; for\; i = 1,...,M} \quad (13)$$

The kernel functions used in the study are as follows:

①Polynomial kernels

$$g(x, x') = (x^T x' + 1)^d \quad (14)$$

where d is an integer.

②RBF kernels

$$g(x, x') = \exp(-\gamma \| x - x' \|^2) \quad (15)$$

where γ is a positive slope control prameter.

In order to obtain the optimal separating hyperplane with the maximum margin, we must find w with the minimum ||w||. This leads to solving the following optimization problem. Namely, minimize

$$\frac{1}{2}\parallel w\parallel^2 \tag{16}$$

and subject to the constraints (13). When trainng data are not linearly separable, a slack variables $\xi_i(>0)$ can be introduced into (13) as follows:

$$y_i(w^T g(x_i)+b) \geq 1 - \xi_i \tag{17}$$

for I = 1,…,M.

The best hyperplane is firm which can be used the maximization of the margin and the min of the train g error are achieved. Namely, minimize

$$\frac{1}{2}\parallel w\parallel^2 + \frac{D}{2}\sum_{i=1}^{M}\xi_i^p \tag{18}$$

subject to (17), where D is a parmeter that detrmines the tradoff connect the max margin and the min class fication mistake rate and p is 1 or 2.

This is a quadratic problem and its optimal solution can be got by solving a series of linear equations. To derive the dual problem of (17) and (18) we simplify a Lagrange function as follows:

$$Q(w,b,a,\xi) = \frac{1}{2}\parallel w\parallel^2 + \frac{D}{2}\sum_{i=1}^{M}\xi_i^2 - \sum_{i=1}^{M} a_i \{y_i(w^T g(x_i)+b) - 1 + \xi_i\} \tag{19}$$

where $\alpha_i \geqslant 0$ is Lagrange multipliers. The condition of the optimal solution is obtained by separating the above equation from the w region, ξ_i, b and α_i and let's get the equation of zero and obtain the following equations:

$$w^* = \sum_{i=1}^{M} a_i y_i g(x_i), \sum_{i=1}^{M} a_i y_i = 0, a_i = D\xi_i \tag{20}$$

Then we get the classification decision function as:

$$f(t) = \mathrm{sgn}\{((w^*)^T g(t)) + b^*\} \tag{21}$$

B. Multiclass Support Vector Machines based on Error-correcting output codes

SVM is originally designed for binary classification. How to effectively extend it to multiclass classification is an ceaseless research topic. We construct a multiclass method combining several repeated SVMs classifier, on the basis of a mistake of correcting output coding plan [6-7].

ECOC has been presented raise the generalization ability of pattern classification. Class encoding is recommended to add the codeword error recovery capability by the decomposition method of DS, which makes the classifier is insensitive to noise. This idea is achieved through the redundacy encoding scheme, such as encoding theory shows. The coding scheme, we can covertly. Ompose k-polychotomy set f_1,... f_B, which B is a codeword that encodes a class of length.

First, a set of binary classifiers {f1,... And the structure and each classifier FB} FB is based on the decomposition of matrix G B column and peer training, and G is the Dj function I value J of the decision target (x) where:

$$g_{ij}(x) = \{\begin{matrix} 1 & if\ D_j(x) > 0\ for\ class\ i, \\ -1 & otherwise. \end{matrix} \tag{22}$$

The jth column vector $g_j=(g_{1j},...,g_{nj})^T$ is the target vector of the first j decision function, and n is the corresponding class label. Line I, row vector (gi1,... (giB) corresponds to a class I codeword.

Next,the clas $\varepsilon_{ij}(x) = \begin{cases} 0 & for \quad g_{ij} = 0, \\ \max(1 - g_{ij}D_j(x), 0) & otherwise. \end{cases}$ s tag of X can be find by computing the distance between per codeword. Bring the "No concern for output " shows 0 for its value, we could define an error $\varepsilon_{ij}(x)$ for

(23)

where $g_{ij}D_j(x) \geqslant 1$ represents X in a right aspect of the first j decision function, equal or exceeding to the max profitand and $g_{ij}D_j(x)<1$ indicates that the X is error, although it’s right, the margn is still more small than the max margn.

The level of x through class i is listed through

$$d_i(x) = \sum_{j=1}^{B}\varepsilon_{ij}(x). \tag{24}$$

So x isdivided into class:

$$\arg \min_{i=1,\ldots,n} d_i(x). \tag{25}$$

C. GA ECOC-SVMs modulation recognition algorithm

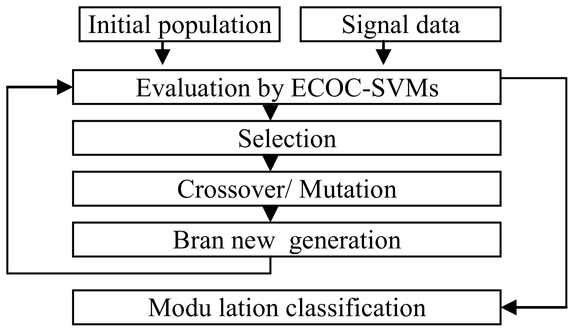

Fig. 1. ECOC-SVMs classification with GA subset selection

Typical pattern recognition methods decrease classification accuracy by employing feature selection to reduce import dimensionality and delete extra import data. In our AIMR application, GA is used in different SNR environment for feature selection of the most appropriate specific recognizers. An individual represents a feature subset by means of a binary string.Using the length L of each string, it is evaluated in the preprocessing phase, which takes full advantage of the total numbers input feature. In the genome bound, the bnary "1" represents apprpriate feature on the coresponding inde number,and the binay "0" indicates that it does not exist. The proposed AIMR algorithm is shown in Figure 1

V. Figures and Tables

We evaluate modulation classification performance of ECOC SVMs by using the 11 classes digitally modulated signal data described in section II. All simulation signals are digitally generated on the basis of equation (1) in MATLAB envronment. Random intiger M-level (M = 2，4，8， 16， 64) be produced through the unform conj numbers produce machine. These will be before the micro plastic baud rate to re-examine, through their regulators, basis different SNR (i.e., 5, 0, 5,10,20 dB) plus additive white Gauss noise. For the modulation parameters are shown in Table 1. The analog signal is band limited.

Table 1. Modulation Parameters

parameters	Baud rate	carrier frequency	samples frequency
values	100 baud	2 kHz	100 kHz

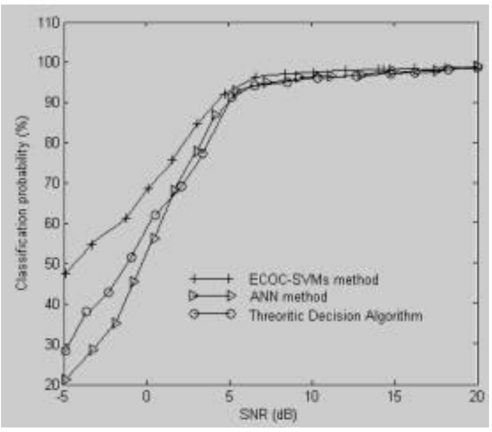

Fig. 2. Overall simulation results of ECOC-SVMs， the theoretic decision algorithm and the ANN method

Figure 2 is an overall function comparison among the decison- theoretic algorithm[8], an ANN way [2] and the advocated ECOC-SVMs way. As can be shown in the figure, the ANN method doesn't good at a lower SNR, for an instance, 0dB. It is because in a lower SNR and with a same training data length the ANN method couldn't construct a proper network for classification. On the other hand, the two methods can only identify the modulation types using less error correction encoding method of support vector machine. Error correction encoding the proposed SVM method is robust and efficient cause is not more than a good classifier structure, but also an adopts the combined spectrum and the statistical feature set.

VI. CONCLUSIONS

In the paper, a new feature set is proposed, which combines statistical and spectral feature sets, uses support vectors for modulation classification and automatic modulation of digital modulation signals. The machine and ECOC introduces the introduction to medicine. In contrast to the popular artificial neural networks (ANN) approach, the method proposed avoids local optimal problems and need fewer training data. Simulation results show that, especially in low signal-to-noise ratio (SNR), the proposed method is more useful and efficient than the existing methods

REFERENCES

[1] Bijan G. Mobasseri Digital modulation classification using constellation shape IEEE Trans. Signal Processing , 2000,80:251-277

[2] M.L.D. Wong, A.K. Nandi Automatic digital modulation recognition using artificial neural network and genetic algorithm, IEEE Trans. Signal Processing 2004, 84:351 – 365

[3] W.C. D, Z.G. Wang, Modulation recognition of MAPSK signals using template matching, Electronics Letters, 2014,1986-1988

[4] E.E. Azzouz, A.K. Nandi, Automatic identification of digital modulation types, IEEE Trans. Signal processing, 1995, 47(1):55-69.

[5] A.K. Nandi, E.E. Azzouz, Modulation recognition using artificial networks, IEEE Trans. Signal processing, 1997,56:165-175

[6] V.Vapnik, An Overview of Statistical Learning Theory, IEEE Trans. Neural Networks, 1999,10(5): 988-999

[7] Swami, B.M. Sadler, Hierachical digital modulation classification using cumulants, IEEE Trans. Commun. 2000, 48 (3):416 – 429.