Digital Modulation Classification by Support Vector Machines and Hilbert–Huang Transformation

ZHIJIN ZHAO, WEIGUO HU, DIANWU GUO

School of Telecommunication
Hangzhou Dianzi University
Hangzhou, Zhe Jiang 310018
CHINA

Abstract:-Support Vector Machines (SVMs) map inputs vectors nonlinearly into a high dimensional feature space and construct the optimum separating hyperplane in space to realize signal classification. Automatic classification of digital modulation signals plays an important role in communication applications such as an intelligent demodulator, interference identification and monitoring, so many investigations have been carried out in the past. Hilbert-Huang transformation (HHT) is a novel method of time frequency analysis for nonlinear and non-stationary data. In this paper, a new method based on SVM and HHT for classifying BFSK, BPSK and 16QAM is proposed. The method can classify these signals well, and the correct classification rates are above 88%.

Key-Words:- Support Vector Machines (SVMs), modulation identification, modulation classification, intelligent demodulator, Hilbert-Huang transformation (HHT), time frequency analysis

1 Introduction

Many studies on modulation type classification using a decision-theoretical or a statistical pattern recognition framework have been carried out [1,2,3]. SVM is a pattern recognition method. It has been used in speech recognition [4], digital recognition and etc. Being different from other learning machines, SVM [5-7] uses a structural risk minimization (SRM) principle, while others use an empirical risk minimization principle. It uses a kernel function for efficiently performing computations in high dimensional spaces and constructs nonlinear decision function to perform an optimal separating hyperplane in feature space.

Hilbert-Huang transformation is a novel method of time frequency analysis for nonlinear and non-stationary data, which was developed by Huang et al in 1998 [8]. This technique is expected to decompose time-dependent data series into its individual characteristic oscillations with the so-called empirical mode decomposition (EMD). This procedure is capable of empirically disintegrating any complex set of data into a finite number of hidden embedded intrinsic mode functions (IMFs). It has been used in other fields of geophysics, e.g. to examine earthquake processes as well as for the determination of the dispersion curves of seismic surface waves [9,10]. It has been used in tsunami research to detect earthquake generated water waves from data series recorded from bottom pressure transducers in the Northern Pacific [11].

In this paper a new method based on support vector machines (SVMs) and Hilbert-Huang transformation for classifying BFSK, BPSK and 16QAM is proposed.

2 Support Vector Machines

Support vector machines are based on the structural risk minimization principle and Vapnik-Chervonenkis (VC) dimension from statistical learning theory developed by Vapnik, et al.[5] Traditional techniques for pattern recognition are based on the minimization of empirical risk, that is, on the attempt to optimize performance on the training set, SVMs minimize the structural risk to reach a better performance [4,5].

We can suppose that S is a set that is made up of points $x_i (i = 1,2\cdots, N)$, which belong to R^n, These points are divided into two classes by an objective function y_i,

$$y_i = \begin{cases} 1 & x_i \in S_1 \\ -1 & x_i \in S_2 \end{cases} \tag{1}$$

where S_1 and S_2 belong to different classes. We try to find a hyperplane to separate the two classes, and sort the same class in same side of the hyperplane as much as possible, and make the margin as far as possible. If S can be separated linearly, there may be $w \in R^n$, $b \in R$ to satisfy

$$\left.\begin{array}{ll} w \cdot x_i + b \geq 1 & y_i = 1 \\ w \cdot x_i + b \leq -1 & y_i = -1 \end{array}\right\} \tag{2}$$

Eqn (2) also can be represented by

$$y_i (w \cdot x_i + b) \geq 1 \tag{3}$$

Parameters (w, b) have determined a hyperplane,

$$w \cdot x_i + b = 0 \tag{4}$$

This plane is called the separating hyperplane. The problem of finding the optimal separating hyperplane is converted to an optimal problem as follows.

$$\min \frac{1}{2} \| w \|^2 \tag{5}$$

with constraints,

$$y_i (w \cdot x_i + b) \geq 1$$

$$(i = 1,2\cdots, N)$$

It is then converted to a dual problem by using Lagrange multiplies,

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{6}$$

with constraints,

$$\sum_{i=1}^{N} y_i \alpha_i = 0 \qquad \alpha \geq 0$$

When S cannot be separated linearly, introducing a nonnegative relax factor $\xi = (\xi_1, \cdots \xi_N)$, Eqn (3) can be rewritten as

$$y_i (w \cdot x_i + b) \geq 1 - \xi_i \tag{7}$$

The optimal problem can be described as

$$\max \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{8}$$

with constraints,

$$\sum_{i=1}^{N} y_i \alpha_i = 0 \quad 0 \leq \alpha_i \leq C$$

Formula (8) is a general form of SVM. When C tends to infinite, formula (8) degenerates into a linear separating problem as formula (6). Replacing $y_i y_j x_i \cdot x_j$ by D_{ij}, the optimal objective function turns to be the maximum $\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i D_{ij} \alpha_j$. Obviously, this is a quadratic program. We can solve it by using the sequential minimal optimization (SMO) proposed by Platt [7]. When parameters

α_i and b are obtained, the different classes can be distinguished by objective function

$$y = \mathrm{sgn}(w^* \cdot x + b^*)$$
$$= \mathrm{sgn}(\sum_{i=1}^{N} \alpha_i^* y_i x_i \cdot x + b^*) \quad (9)$$

In most cases, discrimination is not linear in input space. A higher order function is introduced for mapping a nonlinearly separating problem to a linearly separating problem. Because the optimal problem mentioned above deals with inner product only, a kernel function $K(x_i, x_j)$ can be constructed to substitute the inner product.

Two typical kernel functions are the polynomial kernel function as follows and the Gauss (Radial Basis Function) kernel function, defined by

$$K(x_i, x_j) = [(x_i \cdot x_j) + 1]^d \quad (10)$$

$$K(x_i, x_j) = \exp(-\frac{|x_i - x_j|^2}{2\sigma^2}) \quad (11)$$

A kernel function exists when the Mercer condition is satisfied.

When SVM is used for classification, it works like a neural network, which classifies the different classes by inner product between the input vectors and support vectors. Inner product is substituted by kernel function operation.

3 Feature Parameters

Hilbert-Huang Transformation consists of two parts. Its key part is the so-called empirical mode decomposition (EMD), by which any complicated data set can be decomposed into finite number of intrinsic mode functions(IMFs) . With Hilbert transform, the IMFs yield instantaneous frequencies as functions of time. The final presentation of the results is a time-frequency-energy distribution, designated as the Hilbert spectrum. Being different from Fourier decomposition and wavelet decomposition, EMD has no specified ”basis”. Its ”basis” is adaptively produced depending on the signal itself, which brings not only high decomposition efficiency but also sharp frequency and time localization. EMD is capable of adaptively decomposing signals into oscillating intrinsic components. An IMF is defined as a function that satisfies the following two conditions:

(1) The number of extrema and thus the number of zero-crossings in the whole data series must be equal or differ at the most by one.

(2) At any instant in time, the mean value of the envelope defined by the local maxima and the envelope of the local minima is zero.

The first condition is similar to the narrow-band requirement for a stationary Gaussian process. It ensures that the local maxima of the data series are always positive and the local minima are negative, respectively. The second condition modifies a global requirement to a local one, and is necessary to ensure that the instantaneous frequency will not have unwanted fluctuations as induced by asymmetric waveforms. Regarding an arbitrary data series $x(t)$, the IMFs are obtained, using the following algorithm:

(1) Initialize: $r_0(t)=x(t),\ i=1$

(2) Extract the ith IMF:

(a) Initialize: $h_0(t)=r_i(t),\ k=1$

(b) Extract the local maxima and minima of $h_{k-1}(t)$

(c) Interpolate the local maxima and the local minima by a cubic spline to form upper and lower envelopes of $h_{k-1}(t)$

(d) Calculate the mean $m_{k-1}(t)$ of the upper and lower envelopes of $h_{k-1}(t)$

(e) Define: $h_k(t)=h_{k-1}(t)-m_{k-1}(t)$

(f) If IMF criteria are satisfied, then set $IMF_i(t)=h_k(t)$ else go to (b) with k=k+1

(3) Define: $r_i(t)=r_{i-1}(t)-IMF_i(t)$

(4) If $r_i(t)$ still has at least two extrema, then go to (2) with i=i+1; else the decomposition is completed and $r_i(t)$ is the "residue" of $x(t)$.

At the beginning, the original data set $x(t)$ is initialized as $r_0(t)$. This initialization can be characterized as the introduction to the outer loop to decompose the input signal into successive IMFs. The second inner loop is started to find every single IMF. Again this loop is initiated by an introductory process, setting $r_i(t)$ as the starting array for the inner loop. The first run of the inner loop, array $h_{k-1}(t)$ with k=1, corresponds to the original data series $x(t)$. Extrema of the signal are revealed next. The minima and maxima are linked by a cubic spline to form an upper and lower envelope of $x(t)$. Then the corresponding mean $m_{k-1}(t)$ is defined as the difference of upper and lower envelopes and subtracted from the initial data series $h_{k-1}(t)$ to represent a tentative first IMF $h_k(t)$. The conditions of defining an IMF are subsequently approved. Usually, after the first run, the criteria are not satisfied; so the inner loop is restarted from (b) by using $h_k(t)$ to initialize $h_{k-1}(t)$ with k=k+1. This so-called "sifting" process is repeated until the stopping criteria are fulfilled. Then the first IMF is disintegrated and the whole procedure is redone to sift additional IMF from the data series $x(t)$ provided that the stopping criteria of the outer loop fail. The iterative decomposition process ends when the stopping criteria of the outer loop are satisfied so that $r(t)$ is taken as the residue of $x(t)$ which can also be interpreted as the trend of the signal. The original signal $x(t)$ is then represented through the sum of a specified number of IMFs so that

$$x(t) = \sum_{i=1}^{n} IMF_i + r(t) \tag{12}$$

where n is the total number of IMFs and r(t) is the residue of the sifting process. Due to this iterative procedure, none of these sifted IMFs is derived in the closed analytical form.

In this paper three kinds of commonly used digital modulation signals 16QAM, BFSK and BPSK are classified. Being different from FSK and PSK signals, the amplitude of QAM signals is modulated, so its amplitude and power fluctuate largely from one symbol to another one. The amplitude of ideal FSK and PSK signals is a constant. The energy deviation of main component of IMFs of QAM signals is much larger than zero, while the counterpart of PSK and FSK signals is near to zero. So the characteristic parameter used for distinguishing QAM , FSK and PSK can be chosen as the energy variance of IMFs' main component, denoted δ_{IMF}^2 .

The signal is decomposed into a series of IMFs from the high-frequency components to the low frequency components, the first two IMFs have reflected the basic character. Calculating the Hilbert spectrum of first two components, we find that the spectrum of FSK signal has two main frequency components, while the spectrum of PSK and QAM signal is nearly a spectrum thread. So we chose the relative frequency spectrum width as the characteristic parameter, denoted B_{IMF} .

4 Classification Results

For samples of BPSK, BFSK and 16QAM, experiments have been done by SVM using Gauss kernel functions, linear kernel functions and their parallel combination and using the feature parameters δ_{IMF}^2 and B_{IMF} . The correct classification rates are given in Table 1 and Table 2 at SNR 20dB and 25dB.

Table 1 Correct classification rates at SNR 20dB

Signal Type	Gauss kernel	Linear kernel	Parallel combination
BFSK	92.31%	92.31%	100%
BPSK	61.54%	90.39%	88.46%
16QAM	100%	100%	100%

Table 2 Correct classification rates at SNR 25dB

Signal Type	Gauss kernel	Linear kernel	Parallel combination
BFSK	100%	100%	100%
BPSK	86.54%	88.46%	88.46%
16QAM	100%	100%	100%

From Tables we can see that using the two characteristic parameters and the SVMs, we successfully identify these three kinds of digital modulated signals. The correct classification rates are above 88% at SNR 20dB.

5 Conclusion

In this paper, we proposed the method using SVMs and Hilbert-Huang Transformation for classifying modulation. Results show that better results can be obtained by using parallel combination of SVM using linear kernel and SVM using Gauss kernel, and correct classification rates is much better than 88% at SNR 20dB.

References

[1] A.K. Nandi, E.E. Azzouz, Modulation Recognition Using Artificial Neural Networks, *Signal Processing* 56 (1997) 165-175

[2] S. Samir, Soliman, S.-Z. Hsue, Signal Classification Using Statistical Moments, *IEEE Transactions on Communications*, Vol. 40, No.5 May (1992)

[3] M. Leonardo, Reyneri, Unification of Neural and Wavelet Networks and Fuzzy Systems, *IEEE Transactions on Neural Network*, Vol. 10, No.4 July (1999)

[4] Z.P. Wang, L. Zhao, C. Zou, Support Vector Machines for Emotion Recognition, *Chinese Speech Journal of Southwest University* (English Edition) Vol. 19, No.4 Dec (2003) 307-310

[5] V.N. Vapnik, *Statistical Learning Theory*, New York Wiley (1998)

[6] Christopher, C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge*, Vol. 2, No.2 (1998) 121-167

[7] J. Platt, Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines, Microsoft Research Technical Report MSR-TR-98-4 (1998)

[8] N.E. Huang, Z. Shen, S. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, and H.H. Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of Royal Society London* A **454**, (1998) 903–995.

[9] N.E. Huang, Z. Shen, and S. Long, A new view of nonlinear water waves: the Hilbert spectrum, *Annual Review of Fluid Mechanics* **31**, (1999) 417–457.

[10] Chen, C.-H., Li, C.-P. and Teng, L.-T.,. Surface-wave dispersion measurements using the Hilbert–Huang transformation. *Journal of Terrestrial, Atmospheric and Ocean Science* (TAO) **13** 2, (2002) 171–184.

[11] Schlurmann, T.,. Spectral frequency analysis of nonlinear water waves derived from the Hilbert–Huang transformation. *Journal of Offshore Mechanics and Arctic Engineering (JOMAE), American Society of Mechanical Engineers (ASME)* **124** 1, (2002) 22–27.