

Article

Modulation Classification Using Compressed Sensing and Decision Tree–Support Vector Machine in Cognitive Radio System

Xiaoyong Sun, Shaojing Su, Zhen Zuo *, Xiaojun Guo and Xiaopeng Tan

College of Intelligent Science and Technology, National University of Defense Technology, Changsha 410073, China; sunxiaoyong14@nudt.edu.cn (X.S.); susj-5@163.com (S.S.); jeanakin@nudt.edu.cn (X.G.); tanxiaopeng14@nudt.edu.cn (X.T.)

* Correspondence: z.zuo@nudt.edu.cn; Tel.: +86-731-84573353

Received: 15 January 2020; Accepted: 3 March 2020; Published: 6 March 2020



Abstract: In this paper, a blind modulation classification method based on compressed sensing using a high-order cumulant and cyclic spectrum combined with the decision tree–support vector machine classifier is proposed to solve the problem of low identification accuracy under single-feature parameters and reduce the performance requirements of the sampling system. Through calculating the fourth-order, eighth-order cumulant and cyclic spectrum feature parameters by breaking through the traditional Nyquist sampling law in the compressed sensing framework, six different cognitive radio signals are effectively classified. Moreover, the influences of symbol length and compression ratio on the classification accuracy are simulated and the classification performance is improved, which achieves the purpose of identifying more signals when fewer feature parameters are used. The results indicate that accurate and effective modulation classification can be achieved, which provides the theoretical basis and technical accumulation for the field of optical-fiber signal detection.

Keywords: modulation classification; high-order cumulant; cyclic spectrum; compressed sensing; decision tree–support vector machine

1. Introduction

As one of the booming communication technologies in the information era, modulation classification (MC) technology [1] has a very important application value in the field of wireless communication. For example, it can play an important role in communication investigation, electronic countermeasures, signal authentication, interference identification, spectrum management, etc. At present, the wireless communication network has maintained a steady and rapid development trend, the network construction is increasingly integrated, and network applications are everywhere. At the same time, the inherent contradiction between the centralized static network and the dynamic change of the environment also causes serious problems like the low utilization of spectrum resources in the wireless communication network. Therefore, cognitive radio (CR) technology [2–5] is proposed and considered as a promising technology to solve these problems. MC plays an important role in CR based on spectrum sensing and feature analysis.

Cognitive radio has been widely accepted as a new technology in the field of wireless communication in the new era. In the cognitive radio network (CRN), in order to avoid interference with the transmission of the primary users, it is essential to accurately sense the presence for any contemporaneous transmission of the primary users in the observed spectrum [6]. The primary user signal error detection will cause the secondary user to waste the spectrum opportunity. Noise, shadow and multipath fading lead to a serious degradation of signal characteristics in conventional wireless

communication scenarios. This makes signal detection very difficult in a low signal-to-noise ratio (SNR) environment [7,8]. In addition, because the primary user (authorized user) and the cognitive user (unauthorized user) cannot communicate with each other, accurate MC can not only avoid mutual interference between them, but also provide the multi-dimensional spectrum information of the surrounding wireless environment, which helps to improve the inefficient use of spectrum resources in the CRN. With the different modulation parameters and methods used in the wide-band communication signal, MC has gradually been studied in depth and has become one of the main methods of signal recognition and classification.

MC has been playing an important role in the field of wireless communication for a long time, especially in dynamic spectrum management and interference recognition. A variety of methods and classifiers have been proposed in the literature, but most of them only identify a few modulation formats, such as low-order modulation format, or require some knowledge of parameters of the signal. MCs of a CR system are roughly divided into 4 categories: (a) Multiple quadrature amplitude modulation (MQAM) and multiple phase shift keying (MPSK) signals are classified based on signal envelope variance and wavelet transform, but the recognition rate is low at low SNR [9,10]; (b) artificial neural networks (ANN) based on machine-learning algorithms for automatic signal type recognition, which requires the most appropriate ANN and will lead to an increase in calculation time and risk of over-fitting [11,12]; (c) identification from higher-order cumulant (HOC) using fourth-order cumulant, which cannot identify some signals with the same fourth-order cumulant [13,14]; (d) feature parameters are extracted from the time domain, frequency domain and power spectrum of signals to classify and identify a modulation signal, but some feature parameter extraction processes are complex and easily interfered with by noise [15–17]. The proposed method mainly focuses on the recognition of single-feature parameters, and most classifiers adopted an increase the complexity of the system.

In this paper, we propose a new modulation classification method that combines high-order cumulants and cyclic spectrum feature extraction methods with a decision tree–support vector machine (DT–SVM) classifier. In the feature extraction phase, the compressed sensing (CS) method is used to obtain the compressed sample size of the feature parameters, and the influence of key factors on the classification accuracy in the modulation classification process is analyzed. CS is a signal processing technique called “sampling compression combo”. The CS method can map signals from high-dimensional space to low-dimensional space through a small number of observations (non-adaptive linear projection) of sparse signals, and maintain the original structure of the signal [18]. The sparse signal reconstruction is actually reconstructing the original signal from the signal observations with high probability by solving the non-linear optimization problem, which breaks through the limitations of the traditional Shannon–Nyquist sampling theorem and solves the performance requirements of a sampling system when processing cognitive radio signal. It also relieves the pressure of storage, transmission and processing for large amounts of the traditional sampled data. The combination of HOC and the cyclic spectrum can distinguish the same cumulant of different signals, and achieve the MC through the DT–SVM classifier. Combining the advantages of HOC and cyclic spectrum features, the algorithm directly obtains the compressed values of feature parameters through the CS theory, and analyzes the influence of symbol length and compression ratio on the recognition accuracy. The simulation results show that the algorithm has a better classification performance in low SNR and the validity of the method is verified.

The rest of this paper is structured as follows. Section 2 introduces the feature extraction method and its characteristics in detail. Section 3 introduces the compression sampling values of feature parameters obtained by combining the compression sensing theory. Section 4 describes the structure of the decision tree–support vector machine classifier. In Section 5, some simulation results are presented. Finally, Section 6 sums up the conclusions.

2. Feature Extraction

2.1. Feature Extraction Based on Higher-Order Cumulant (HOC)

For wireless channel model, we studied the property of HOC and the insensitivity of its second-order terms to Gaussian noise, the k th-order cumulant $C_{k,n}(m_1, m_2, \dots, m_{k-1})$ of a complex-valued stationary random process $x(t)$, can be defined as:

$$C_{k,n}(m_1, m_2, \dots, m_{k-1}) = cum(x(t), x(t + m_1), \dots, x(t + m_{k-1})) \tag{1}$$

where $x(t + m_k)$ denotes a function of different time delays and regardless of t , $cum(\bullet)$ means taking the cumulant. Therefore, its fourth-order cumulant is:

$$C_{4,n}(m_1, m_2, \dots, m_3) = E[x(n)x(n + m_1)x(n + m_2)x(n + m_3)] - C_{2,n}(m_1)C_{2,n}(m_2 - m_3) - C_{2,n}(m_2)C_{2,n}(m_3 - m_1) - C_{2,n}(m_3)C_{2,n}(m_1 - m_2) \tag{2}$$

Based on the above theory, the fourth-order, sixth-order and eight-order cumulants of the zero-mean $x(t)$, are shown as:

$$\begin{aligned} C_{4,0} &= cum(x, x, x, x) = M_{4,0} - 3M_{2,0}^2 \\ C_{4,1} &= cum(x, x, x, x^*) = M_{4,1} - 3M_{2,1}M_{2,0} \\ C_{4,2} &= cum(x, x, x^*, x^*) = M_{4,2} - |M_{2,0}|^2 - 2M_{2,1}^2 \\ C_{6,0} &= cum(x, x, x, x, x, x) = M_{6,0} - 15M_{4,0}M_{2,0} + 30M_{2,0}^3 \\ C_{6,3} &= cum(x, x, x, x^*, x^*, x^*) = M_{6,3} - 9C_{4,2}C_{2,1} - 6C_{2,1}^3 \\ C_{8,0} &= cum(x, x, x, x, x, x, x, x) = M_{8,0} - 28M_{6,0}C_{2,0} - 35M_{4,0}^2 + 420M_{4,0}M_{2,0}^2 - 630M_{2,0}^4 \end{aligned} \tag{3}$$

where $M_{pq} = E[x(t)^p x^*(t)^q]$ denotes the p th-order mixing moment [19].

In the practical application of MC, we need to estimate the HOC value of the signal from the received symbol sequence in the shortest possible time. Sample estimations of the correlations are given by:

$$\begin{aligned} C_{4,0} &= \frac{1}{N} \sum_{n=1}^N (x(t))^4 - 3C_{2,0}^2 \\ &\dots\dots \\ C_{8,0} &= \frac{1}{N} \sum_{n=1}^N (x(t))^8 - 28C_{2,0} \frac{1}{N} \sum_{n=1}^N (x(t))^6 - 35M_{4,0}^2 + 420M_{4,0}M_{2,0}^2 - 630M_{2,0}^2 \end{aligned} \tag{4}$$

Substituting the estimated values into Equation (4), we can obtain all of the features for the considered six wireless signal types. Table 1 shows some of these features for a number of these signals. These values are computed under the constraint of unit variance in noise free conditions. It can be seen that by computing of these values, we can classify the wireless signal types.

Table 1. Theoretical values of higher-order cumulant (HOC) for six wireless signal modulations.

	$ C_{4,0} $	$ C_{4,1} $	$ C_{4,2} $	$ C_{6,0} $	$ C_{6,3} $	$ C_{8,0} $
OOK	2	2	2	16	13	272
DPSK	2	2	2	16	13	272
QPSK	1	0	1	0	4	34
OQPSK	1	0	1	0	4	34
16QAM	0.68	0	0.68	0	2.08	13.9808
64QAM	0.619	0	0.619	0	1.7972	11.5022

Table 1 shows that OOK (on-off keying), DPSK (differential phase shift keying), QPSK (quadrature phase shift keying), OQPSK (offset quadrature phase shift keying) have the same theoretical values

of HOC. In addition, 16QAM (16 quadrature amplitude modulation) and 64QAM (64 quadrature amplitude modulation) have similar HOC values. Therefore, we can define a feature parameter $T1 = |C_{8,0}|/|C_{4,0}|$ that is calculated in Table 2 and divides signals into three categories including (OOK, DPSK), (QPSK, OQPSK) and (16QAM, 64QAM). It is worth noting that the absolute value and ratio form are used to eliminate the effect of phase jitter and amplitude [20].

Table 2. Theoretical values of T1 for six wireless signal modulations.

	OOK,DPSK	QPSK,OQPSK	16QAM	64QAM
T1	136	34	20.56	18.5819

Owing to the difference between the phase jump rules of QPSK and OQPSK, the sampling sequence of both can be performed with a differential operation, i.e.,

$$\Delta x(t) = x(t+1) - x(t) = (a_{t+1} - a_t) \exp[j(2\pi f_c + \Delta\theta_c)] \quad (5)$$

where $x(t)$ denotes the signals of QPSK and OQPSK, a_k is the transmitted symbol sequences, f_c denotes the carrier frequency and θ_c denotes the phase jitter. For the sake of discussion, we assume that f_c and θ_c have been completed timing synchronization. The values of HOC under difference operation are calculated in Table 3. Then we define another feature parameter $T2 = |C_{d8,0}|/|C_{d4,0}|^2$ is calculated in Table 4 to classify QPSK and OQPSK, where $C_{d8,0}$ and $C_{d4,0}$ represent the cumulants after differential operation.

Table 3. Theoretical values of HOC after difference between QPSK and OQPSK.

	$ C_{d4,0} $	$ C_{d4,1} $	$ C_{d4,2} $	$ C_{d6,3} $	$ C_{d8,0} $
QPSK	2	0	2	8	68
OQPSK	2	0	0.89	2	131.4

Table 4. Theoretical values of T2 for QPSK and OQPSK.

	QPSK	OQPSK
T2	17	32.85

2.2. Feature Extraction Based on Cyclic Spectrum

Since the T1 of (OOK, DPSK) and (16QAM, 64QAM) are the same or similar, a cyclic spectral density function for noise suppression is proposed for identification. Assuming $x(t)$ is the cyclostationary signal, and its mean value and autocorrelation function are periodic with T_0 shown as:

$$m_x(t + T_0) = m_x(t) \quad (6)$$

$$\mathbf{R}_x(t + T_0 + \frac{\tau}{2}, t + T_0 - \frac{\tau}{2}) = \mathbf{R}_x(t + \frac{\tau}{2}, t - \frac{\tau}{2}) \quad (7)$$

where τ is the delay variable. Because the autocorrelation function has periodicity, its Fourier series can be written as:

$$\mathbf{R}_{x\alpha}(t + \frac{\tau}{2}, t - \frac{\tau}{2}) = \sum_{\alpha} \mathbf{R}_{x\alpha}(\tau) e^{j2\pi\alpha t} \quad (8)$$

where α stands for the frequency corresponding to the instantaneous autocorrelation and is often called the cyclic frequency. In addition, $\mathbf{R}_{x\alpha}$ is the coefficient of the Fourier series which is given by:

$$\mathbf{R}_{x\alpha}(\tau) = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} \mathbf{R}_{x\alpha}(t + \frac{\tau}{2}, t - \frac{\tau}{2}) e^{-j2\pi\alpha t} dt \quad (9)$$

The Fourier transform of the cyclic autocorrelation function can be written as:

$$\mathbf{S}_{x\alpha}(f) \triangleq \int_{-\infty}^{\infty} \mathbf{R}_{x\alpha}(\tau) e^{-j2\pi f t} d\tau \quad (10)$$

where $\mathbf{S}_{x\alpha}(f)$ is called power spectral density and f is the spectral frequency.

The $\mathbf{R}_{x\alpha}(\tau)$ can be seen as the cross-correlation of two complex frequency shift components $u(t)$ and $v(t)$ of $x(t)$, i.e.,

$$\mathbf{R}_{x\alpha}(\tau) = \mathbf{R}_{uv\alpha}(\tau) = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} u(t + \frac{\tau}{2}) v^*(t - \frac{\tau}{2}) dt. \quad (11)$$

where $u(t) = x(t)e^{j2\pi\alpha t}$, $v(t) = x(t)e^{-j2\pi\alpha t}$.

From Equation (10) we can obtain $\mathbf{S}_{x\alpha}(f) = \mathbf{S}_{uv\alpha}(f)$. Through the cross-spectrum analysis, we can obtain:

$$\mathbf{S}_{x\alpha}(f) \triangleq \lim_{T_0 \rightarrow \infty} \lim_{\Delta t \rightarrow \infty} \mathbf{S}_{uvT_0}(f)_{\Delta t} = \lim_{T_0 \rightarrow \infty} \lim_{\Delta t \rightarrow \infty} \frac{1}{\Delta t} \int_{-\frac{\Delta t}{2}}^{\frac{\Delta t}{2}} \mathbf{S}_{X_{T_0}\alpha}(t, f) dt \quad (12)$$

$$\mathbf{S}_{X_{T_0}\alpha}(t, f) = \frac{1}{T_0} \mathbf{X}_{T_0}(t, f + \frac{\alpha}{2}) \mathbf{X}_T^*(t, f + \frac{\alpha}{2}) \quad (13)$$

$$\mathbf{X}_{T_0}(t, f + \frac{\alpha}{2}) = \int_{t-\frac{T_0}{2}}^{t+\frac{T_0}{2}} x(u) e^{-j2\pi f u} du \quad (14)$$

where Equation (12) is used to estimate the cyclic spectral density, Equation (13) is the cyclic periodic diagram, Equation (14) is the short-time Fourier transform (STFT) formula, Δt is the length of received data, T_0 is the window length for the STFT, and $(\bullet)^*$ is the complex conjugate.

According to the insensitivity of the cyclic spectrum to noise and the above theory, the characteristic parameter $\mathbf{T3} = \max(\mathbf{S}_{x\alpha})$ is defined to distinguish the signal set of (OOK, DPSK) and (16QAM, 64QAM).

3. Compressed Values of Feature Parameters Based on Compressed Sensing

3.1. Compressed Value of HOC

CS techniques perform successfully whenever applied to so-called compressible and/or K -sparse signals, i.e., signals that can be represented by $K \ll N$ significant coefficients over an N -dimensional basis [21,22]. The K -sparse signal $s(t) \in \mathbb{R}^N$ of dimension N is accomplished by computing a measurement vector $y(t) \in \mathbb{R}^M$ that consists of $M \ll N$ linear projections of the vector $s(t)$. The compression sampling rate $f_{cs} = (M/N)f_s$, where f_s is traditional sampling rate and $\delta = M/N \in (0, 1)$ is called the compression ratio. The linear compression sampling process can be described as:

$$y = \Phi s + n \quad (15)$$

where Φ represents a $M \times N$ matrix, usually over the field of real numbers. It is noted that the measurement matrix Φ is a random matrix satisfying the restricted isometry property (RIP) [23], and its form is various, such as Gaussian matrix [24] and local Hadamard matrix [25].

According to the theory of feature extraction and compression sensing, the linear square compression sampling process can be defined as:

$$\llbracket x \rrbracket^2 = \Phi \llbracket s \rrbracket^2 \quad (16)$$

where $\llbracket \bullet \rrbracket$ represents the product operation of the corresponding elements between the vectors. From Equation (15), we can through CS to simplify the reconstruction.

First, we define a relationship between the autocorrelation matrix $\mathbf{R}_{\llbracket x \rrbracket^2}$ and $\mathbf{R}_{\llbracket s \rrbracket^2}$:

$$\mathbf{R}_{\llbracket x \rrbracket^2} = \llbracket x \rrbracket^2 (\llbracket x \rrbracket^2)^T = (\Phi \llbracket s \rrbracket^2) (\Phi \llbracket s \rrbracket^2)^T = \Phi \mathbf{R}_{\llbracket s \rrbracket^2} \Phi^T \tag{17}$$

Second, we obtain a relationship between $\llbracket x \rrbracket^2$ and $\mathbf{R}_{\llbracket x \rrbracket^2}$:

$$\llbracket x \rrbracket^2 = \mathbf{P}_{\llbracket x \rrbracket^2} \text{vec}(\mathbf{R}_{\llbracket x \rrbracket^2}) \tag{18}$$

where $\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$ (\otimes denotes Kronecker product) and $\mathbf{P}_{\llbracket x \rrbracket^2} \in \{0, 1\}^{n \times n^2}$ that maps the linearly products to the vectorized counterparts $\llbracket x \rrbracket^2$ and $\mathbf{R}_{\llbracket x \rrbracket^2}$ [26].

According to Equation (4), the compressed value of fourth-order cumulant can be defined as:

$$\mathbf{C}_{(4,0)\alpha} = F \llbracket s \rrbracket^4 - 3 \llbracket F \llbracket s \rrbracket^2 \rrbracket^2 \tag{19}$$

where $F = \frac{1}{N} [\exp(-j2\pi\alpha n/N)]_{(\alpha,n)} \in \mathbb{R}^{M_\alpha \times N}$ is the discrete Fourier transform (DFT) matrix.

Therefore, the linear representation process of the vector-form $\mathbf{C}_{(4,0)\delta}$ is shown as:

$$\begin{aligned} \mathbf{C}_{(4,0)\delta} &= FP_s \text{vec}(R_{\llbracket s \rrbracket^2}) - 3P_{F_s} \text{vec}(\mathbf{R}_{F \llbracket s \rrbracket^2}) = FP_s \text{vec}(\mathbf{R}_{\llbracket s \rrbracket^2}) - 3P_{F_s} \text{vec}(F \mathbf{R}_{\llbracket s \rrbracket^2} F^T) \\ &= [FP_s - 3P_{F_s}(F \otimes F)] \text{vec}(\mathbf{R}_{\llbracket s \rrbracket^2}) = [FP_s - 3P_{F_s}(F \otimes F)] [P_{\llbracket x \rrbracket^2} (\Phi \otimes \Phi)]^\dagger \llbracket x \rrbracket^2 \end{aligned} \tag{20}$$

where $[\bullet]^\dagger$ stands for the pseudo-inverse operation.

Finally, by deriving the linear compressed sampling process of the fourth-order cumulant, we can obtain the compressed value of the eighth-order as follows:

$$\mathbf{C}_{(8,0)\delta} = \{FP_s[(P_s - 28P_s^{1/2}P_{F_s}^{1/2}(F \otimes F)^{1/2} - 35FP_s + 420P_{F_s}(F \otimes F)] - 630P_{F_s}^2(F \otimes F)^2\} \bullet [P_{\llbracket x \rrbracket^2} (\Phi \otimes \Phi)]^\dagger \llbracket x \rrbracket^2 \tag{21}$$

Therefore, the first and second characteristic parameters after CS is $T1 = |\mathbf{C}_{(8,0)\delta}|/|\mathbf{C}_{(4,0)\delta}|$ and $T2 = |C_{d(8,0)\delta}|/|C_{d(4,0)\delta}|^2$.

3.2. Compressed Value of Cyclic Spectrum

It can be seen from Equations (10) and (16) that there is no direct linear relationship between the compressed sampled value x in the time domain and the cyclic spectrum $S_{x\alpha}$, so the existing reconstruction algorithm cannot be used to implement the cyclic spectrum estimation. It is necessary to use some explicit linear relations between the second-order statistic to derive its transformation, and indirectly establish the linear relationship between them, so as to use the existing reconstruction algorithm to complete the estimation of the cyclic spectrum [27].

In order to obtain the compressed values of the cyclic spectrum, we first need to obtain the relationship between the cyclic spectrum matrix $\mathbf{S}_{s\delta}$ and the cyclic autocorrelation matrix $\mathbf{R}_{s\delta}(u, v)$. Let $\mathbf{R}_{s\delta}(u, v)$ denote the form of the time-varying covariance matrix \mathbf{R} . When x is a real value, \mathbf{R} is a symmetric semi-positive definite matrix. For the convenience of calculation, we convert it into an auxiliary covariance matrix:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{s\delta}(0,0) & \mathbf{R}_{s\delta}(0,1) & \mathbf{R}_{s\delta}(0,2) & \cdots & \mathbf{R}_{s\delta}(0,N-1) \\ \mathbf{R}_{s\delta}(1,0) & \mathbf{R}_{s\delta}(1,1) & \mathbf{R}_{s\delta}(1,2) & \cdots & 0 \\ \mathbf{R}_{s\delta}(2,0) & \mathbf{R}_{s\delta}(2,1) & \mathbf{R}_{s\delta}(2,2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{R}_{s\delta}(N-1,0) & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{22}$$

where N is the sampling point.

The matrix \mathbf{R} contains all the elements in the $\mathbf{R}_{s\delta}$ vector except for the zero elements. The relationship between $\mathbf{R}_{s\delta}$ and \mathbf{R} can be expressed as:

$$vec\{\mathbf{R}\} = \mathbf{H}_N \mathbf{R}_{s\delta} \tag{23}$$

where $\mathbf{H}_N \in \{0, 1\}^{N^2 \times (N(N+1)/2)}$, $vec\{\bullet\}$ means the vectorization operation, and the $\mathbf{R}_{s\delta}$ vector is mapped to $vec\{\mathbf{R}\}$.

So the relationship between the cyclic spectrum matrix and the cyclic autocorrelation matrix is shown as:

$$\begin{aligned} \mathbf{R}_{s\delta} &= \sum_{v=0}^{N-1} \mathbf{G}_v \mathbf{R} \mathbf{D}_v \\ \mathbf{S}_{s\delta} &= \mathbf{R}_{s\delta} \mathbf{F} \end{aligned} \tag{24}$$

where $\mathbf{G}_n = [\frac{1}{N} \exp(-j\frac{2\pi}{N}a(n + \frac{v}{2}))]_{(a,n)} \in \mathbb{R}^{N \times N}$, $\mathbf{F} = [\exp(-j2\pi vb/N)]_{(v,b)}$ is the N -point DFT matrix and \mathbf{D}_v is an $N \times N$ matrix with only its (v, v) th diagonal element being 1 and all other elements being 0. In addition, n and v are time delay, $a, b \in [0, N - 1]$ denotes digital cyclic frequency and $\alpha = fa/N$ stands for the cyclic frequency.

The time-varying covariance matrix $\mathbf{R}_x = E\{x_m x_m^T\}$ of the compressed value x is also a symmetric semi-definite matrix, which can be rearranged into a vector $\mathbf{R}_{x\delta}$ of length $M(M + 1)/2$ to represent as:

$$\begin{aligned} \mathbf{R}_{x\delta} &= [\mathbf{R}_{x\delta}(0, 0), \mathbf{R}_{x\delta}(1, 0), \dots, \mathbf{R}_{x\delta}(M - 1, 0) \\ &\quad \mathbf{R}_{x\delta}(0, 1), \mathbf{R}_{x\delta}(1, 1), \dots, \mathbf{R}_{x\delta}(M - 2, 1) \\ &\quad \mathbf{R}_{x\delta}(0, M - 1)]^T \end{aligned} \tag{25}$$

Through the linear formula conversion, we can define two projection matrices $\mathbf{P}_m \in \{0, 1\}^{N^2 \times (N(N+1)/2)}$ and $\mathbf{Q}_m \in \{0, 1/2, 1\}^{(M(M+1)/2) \times M^2}$ map the entries of x , s to those in $vec\{\mathbf{R}_x\}$ and $vec\{\mathbf{R}_s\}$, it can be shown that:

$$\begin{aligned} vec\{\mathbf{R}_x\} &= \mathbf{P}_m x \\ s &= \mathbf{Q}_m vec\{\mathbf{R}_s\} \end{aligned} \tag{26}$$

where \mathbf{P}_m and \mathbf{Q}_m are special mapping matrices.

Because of $x(t) = \Phi s(t)$, we can obtain:

$$x = \mathbf{Q}_m vec\{\mathbf{R}_x\} = \mathbf{Q}_m (\Phi \otimes \Phi) vec\{\mathbf{R}_s\} = \mathbf{Q}_m (\Phi \otimes \Phi) \mathbf{P}_m s = \Theta s \tag{27}$$

where $\Theta = \mathbf{Q}_m (\Phi \otimes \Phi) \mathbf{P}_m \in \mathbb{R}^{\frac{M(M+1)}{2} \times \frac{N(N+1)}{2}}$.

Following the equation (24), we can obtain $vec\{\mathbf{R}_{s\delta}\}$ is:

$$vec\{\mathbf{R}_{s\delta}\} = \sum_{v=0}^{N-1} (\mathbf{G}_v^T \otimes \mathbf{D}_v) vec\{\mathbf{R}_s\} = \Omega s \tag{28}$$

where $\Omega = \sum_{v=0}^{N-1} (\mathbf{G}_v^T \otimes \mathbf{D}_v) \mathbf{P}_m \in \mathbb{R}^{N^2 \times (N(N+1)/2)}$.

Through the Equations (24), (27) and (28), we can derive the measurement vector x as a linear function of the vector-form cyclic spectrum $\mathbf{S}_{s\delta}$ as:

$$\mathbf{S}_{s\delta} = \Xi \Omega \Theta^\dagger x \tag{29}$$

where $\Xi = (\mathbf{F}^{-T} \otimes \mathbf{I}_N)^{-1}$, \mathbf{I}_N is the N dimension unit matrix. Therefore, the third characteristic parameter after CS is $T3 = \max(\mathbf{S}_{s\delta})$.

4. The Structural Process of Decision Tree–Support Vector Machine Classifier

4.1. The Principle of Support Vector Machine

Support vector machine (SVM) is based on the principle of structural risk minimization [28–30]. Its final solution can be transformed into a quadratic convex programming problem with linear constraints. There is no local minimum problem. By introducing the kernel function, the linear SVM can be simply extended to the non-linear SVM, and there is almost no additional computation for high-dimensional samples.

The main idea can be seen from Figure 1 that for a linear separable case, the idea of maximizing the classification boundary is used to seek the optimal hyperplane H , while $H1$ and $H2$ are hyperplanes passing through the closest sample to the H and parallel to h , respectively, and the distance between them is called the classification interval. In the case of linear indivisibility, the linear indivisible samples in the low-dimensional input space are transformed into the high-dimensional feature space by the non-linear mapping algorithm, so that they can be linearly separable, in order that the high-dimensional feature space can be solved by the linear analysis method.

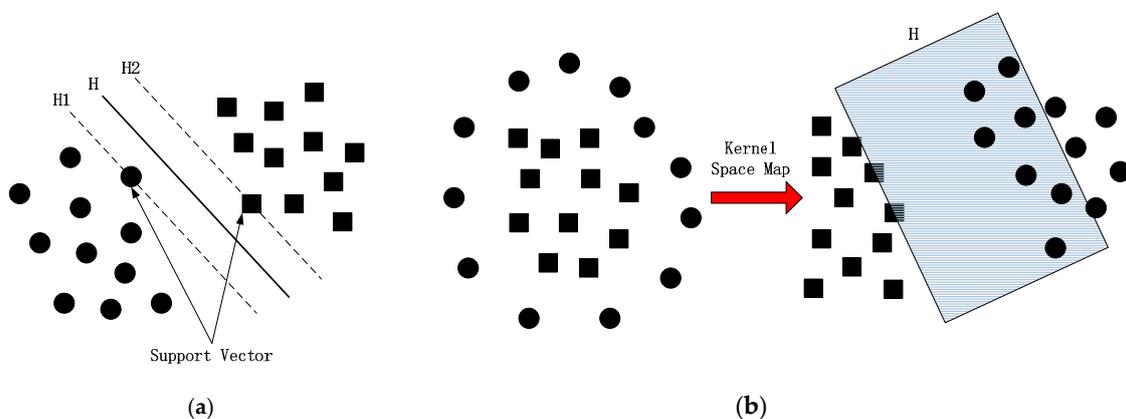


Figure 1. Linear (a) and non-linear (b) classification of support vector machine (SVM).

Suppose that the training set is $\{(x_i, y_i), i = 1, 2, \dots, L\}$ and the expected output is $y_i \in \{+1, -1\}$, where $+1$ and -1 represent two kinds of class representation respectively. If $x_i \in R^n$ belongs to the first category, the corresponding output is $y_i = +1$; if it belongs to the second category, the corresponding output is $y_i = -1$. The linear separability of the problem shows that there is a hyper-plane $(w \cdot x) + b = 0$, which makes the positive and negative inputs of the training points located on both sides of the hyper-plane, respectively. When the training sets are not completely linearly separable, we can introduce the relaxation variable $\xi_i \geq 0 \quad i = 1, 2, \dots, L$, then the objective function is transformed into:

$$\begin{aligned} \min \quad & \phi(w) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L \xi_i \\ \text{s.t.} \quad & y_i [(w \cdot x_i) + b] \geq 1 - \xi_i \\ & \xi_i \geq 0 \quad i = 1, 2, \dots, L \end{aligned} \quad (30)$$

where $C \geq 0$ is the penalty parameter. A larger C indicates a larger penalty for misclassification and is the only parameter that can be adjusted in the algorithm. By choosing the proper kernel function $K(x, x^T)$ and using the Lagrange multiplier method to solve Equation (30), the corresponding dual problem can be obtained as follows:

$$\begin{aligned} \min_{\alpha} & \frac{1}{2} \sum_{i=1}^L \sum_{j=1}^L y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^L \alpha_j \\ \text{s.t.} & \sum_{i=1}^L y_i \alpha_i = 0 \\ & 0 \leq \alpha_i \leq C \quad i = 1, 2, \dots, L \end{aligned} \quad (31)$$

Equation (31) obtains the optimal solution $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_L^*)$ and selects a positive component $0 \leq \alpha_j^* \leq C$ of α^* , and calculates $b^* = y_j - \sum_{i=1}^L y_i \alpha_i^* K(x_i, x_j)$ accordingly. Finally, the policy function $f(x) = \text{sgn}(\sum_{i=1}^L y_i \alpha_i^* K(x_i, x_j) + b^*)$ is obtained.

4.2. The Structure of Decision Tree–Support Vector Machine Classifier

Through the above calculation and analysis of the compressed values of the feature parameters based on CS, the feature parameters are input into the decision tree–support vector machine (DT–SVM) structure with efficient calculation to realize signal classification. Adding support vector machine (SVM) to every node of the decision tree can comprehensively utilize the efficient computing power of the decision tree structure and the high classification performance of the SVM to achieve the high-precision classification of MC. In particular, for the K-class classification problem, the method of DT–SVM only needs to construct K-1 SVM sub classifier. The classification process of decision tree is shown in Figure 2.

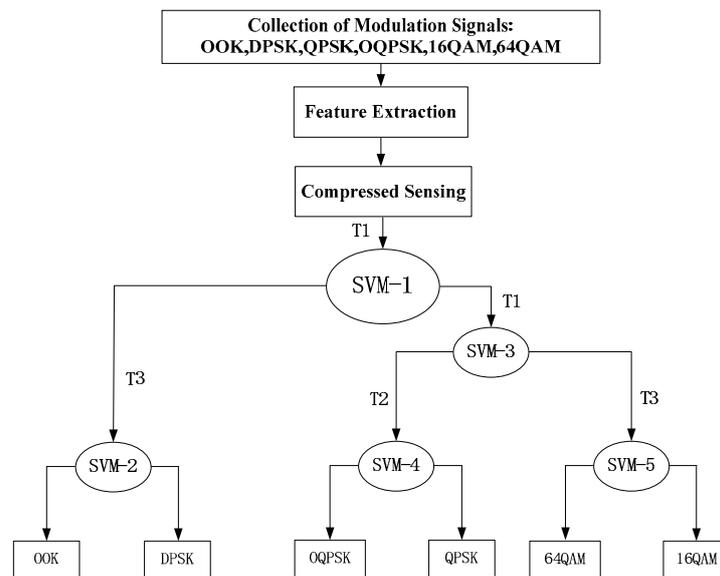
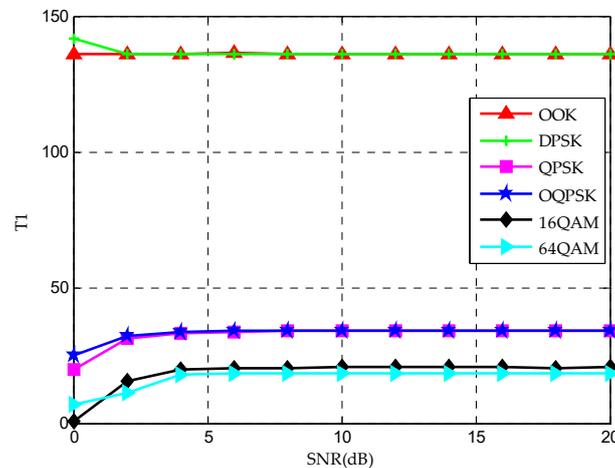


Figure 2. The identification process of decision-tree (DT) classifier.

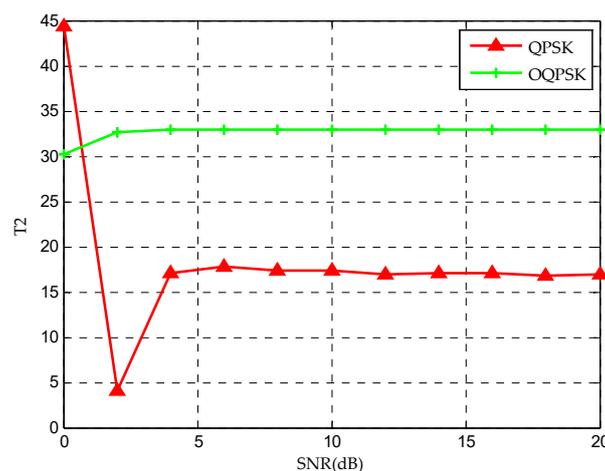
Using the three features in Figure 1 to complete the MC, the specific steps of the process are as follows:

- (1) Three feature vectors are obtained from six kinds of wireless modulation signal data through feature extraction module;
- (2) The feature vector is input into the compression-sensing module to obtain the respective compression sampling values, as shown in Figures 3 and 4;
- (3) Six kinds of wireless signals are roughly classified by T1. The (OOK, DPSK) signals can be separated by SVM-1, and the remaining signals are classified into one class;
- (4) For (OOK, DPSK) signals, SVM-2 and T3 are used to realize classification;

- (5) By SVM-3 and T1, the residual signals can be divided into two categories: (QPSK, OQPSK) and (16QAM, 64QAM);
- (6) The T2 after differential operation and SVM-4 are used to classify QPSK and OQPSK;
- (7) Finally, the classification of 16QAM and 64QAM signals is realized by the T3 and SVM-5.



(a)



(b)

Figure 3. The values of feature parameters (a) T1 and (b) T2 under different signal-to-noise ratios (SNRs).

It can be seen from Figure 3a that the compressed value of the feature parameter T1 tends to be stable with the increase of SNR and conforms to the theoretical value. In addition, it is obvious from the figure that T1 can classify the six signals into three categories includes (OOK, DPSK), (QPSK, OQPSK) and (16QAM, 64QAM). Similarly, it can be seen from Figure 3b that the feature parameter T2 after difference can distinguish QPSK from OQPSK. The circular spectrum and the cross-sectional diagrams of cycle frequency of OOK, DPSK, 16QAM and 64QAM are shown in Figure 4. It can be seen from Figure 4 that the maximum value of the cyclic spectrum of different signals is different, so the remaining signals can be distinguished by T3.

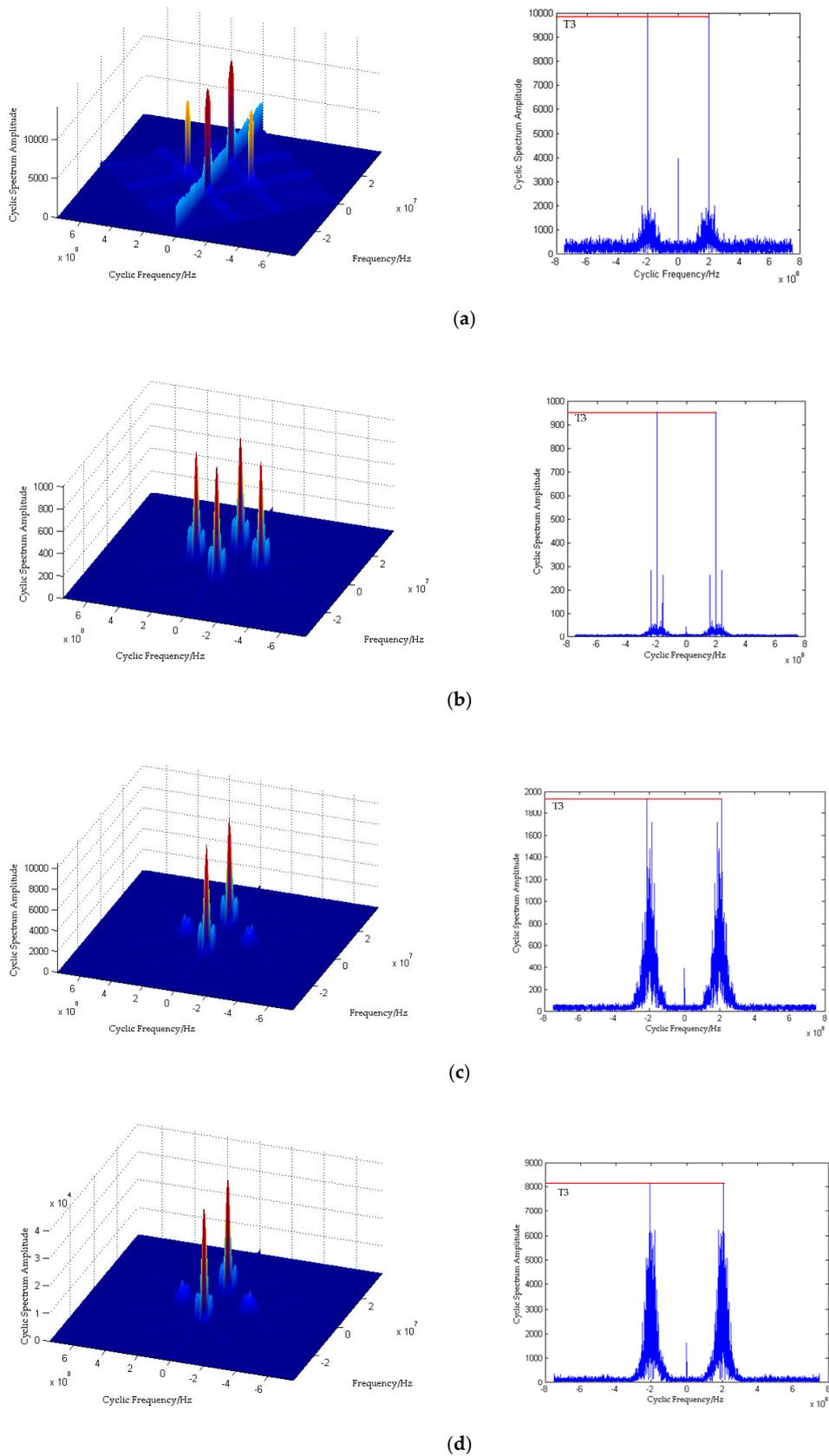


Figure 4. Cyclic spectrum and cross-sectional diagrams of (a) OOK (on-off keying), (b) DPSK (differential phase shift keying), (c) 16QAM (16 quadrature amplitude modulation) and (d) 64QAM (64 quadrature amplitude modulation).

5. Simulation Results and Discussion

For signal modulation classification task, the modulation set is {OOK, DPSK, QPSK, OQPSK, 16QAM, 64QAM}. In this simulation process, all modulation signals adopt the same modulation parameters, that is, the carrier frequency is 100 kHz, the symbol rate is 40 kbps, and the sampling frequency is 800kHz. For each kind of modulation signal, the simulation generates 500 characteristic samples under different SNR (SNR from -5 dB to 15 dB, interval 5 dB). The K-fold cross validation (K-CV) method is used to evaluate the generalization ability of the model, which can not only improve the data utilization, but also solve the over fitting problem to a certain extent, so as to select the model. In this paper, K is chosen as 10. The basic principle of K-CV is to divide the training data set into K equal subsets. Each time, K-1 data are used as training data, and other data are used as test data. In this way, we repeat K times, estimate the expected generalization error according to the mean square error (MSE) average value after K times iteration, and finally select a group of optimal parameters.

To evaluate the influence of different symbol lengths on the classification performance, we select the symbol length $N = 512, 1024, 2048, 4096$ for the OOK, QPSK and 16QAM signals (as examples), respectively, to analyze the classification accuracy in Figure 5. As can be seen from Figure 5, with the increase of symbol length, the trend of classification accuracy is gradually increasing and finally tends to 100%. However, the increase of symbol length affects the classification accuracy mainly in the case of low SNR. Considering the influence of the computation cost in the classification process, 2048 is chosen as the symbol length in this paper.

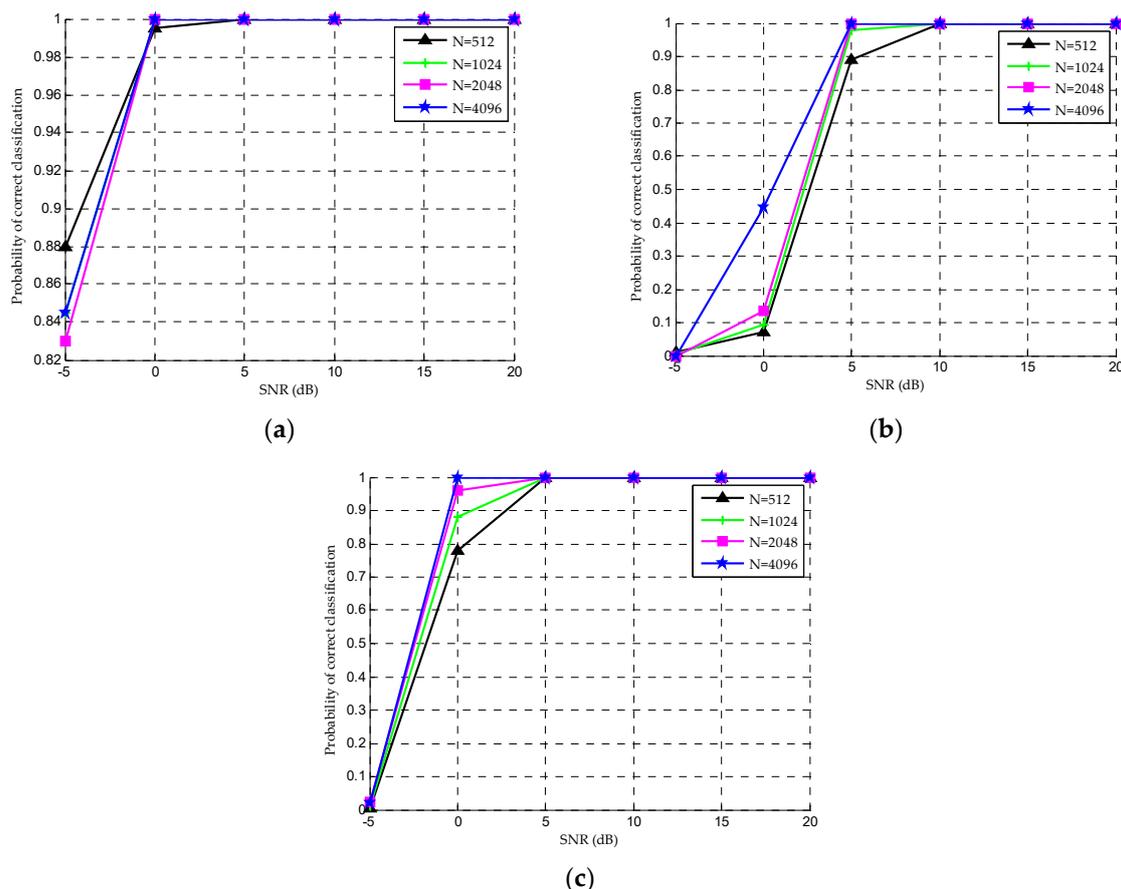


Figure 5. Correct classification rate with different symbol length for (a) OOK, (b) QPSK and (c) 16QAM under different SNRs.

In order to evaluate the impact of different compression ratios on the classification performance, we analyzed the classification accuracy in Figure 6 with the compression ratios of 25%, 37.5%, 50% and

75% for OOK, QPSK and 16QAM signals (as examples) when the symbol length is 2048. As can be seen from Figure 6, with the increase of compression ratio, the classification accuracy increases slightly and finally tends to 100%. The increase of symbol rate has little effect on recognition rate. Therefore, in order to reduce the sampling rate and system complexity as much as possible, 25% is selected as the compression rate value.

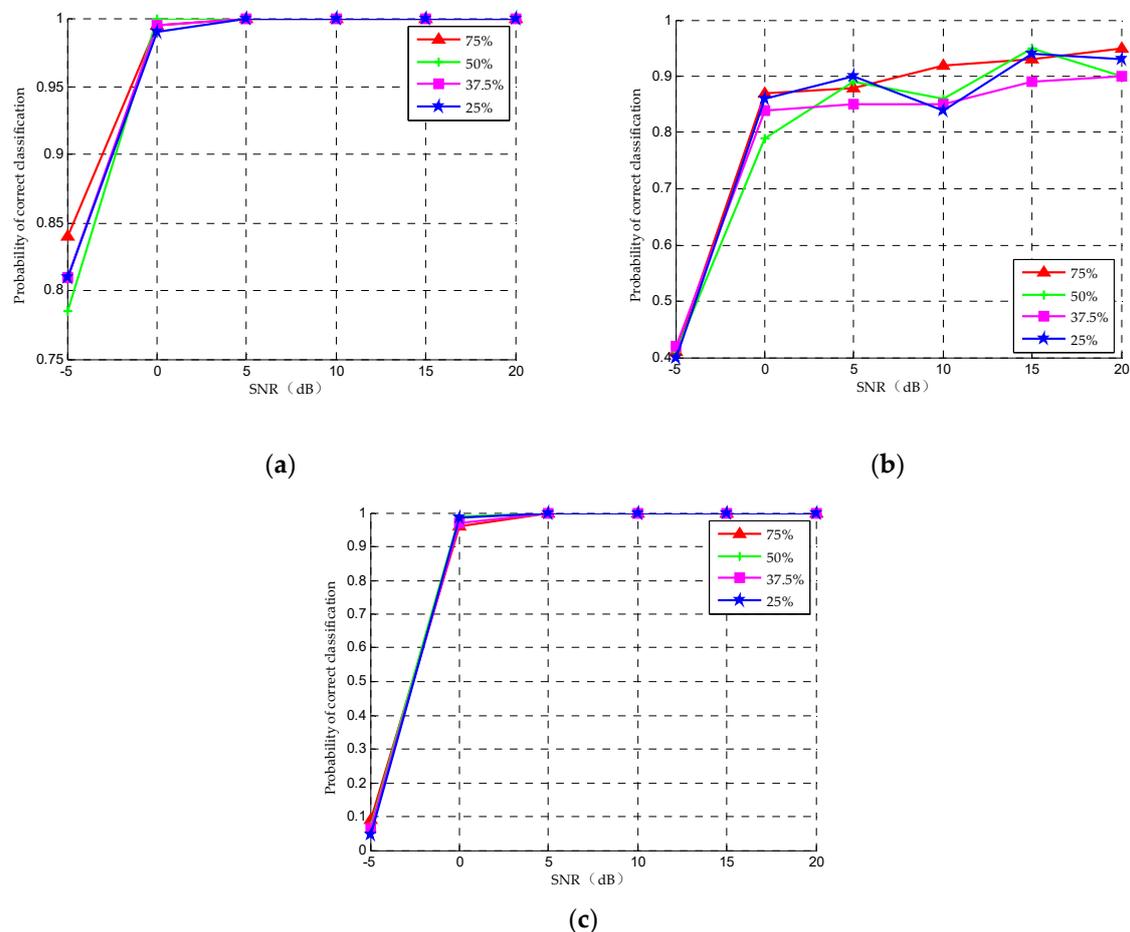


Figure 6. Correct classification rate with different compression rate for (a) OOK, (b) QPSK and (c) 16QAM under different SNRs.

In order to optimize the parameters of kernel function and improve the classification accuracy, the grid search method is used. Table 5 shows the optimization results of penalty parameter and kernel parameter of each node of DT-SVM and the classification accuracy of sub-SVM under the RBF kernel function. It can be seen from the table that under different SNR conditions, with the improvement of SNR, the classification accuracy of sub nodes has improved. When the SNR is 0 dB or above, the average classification accuracy has reached 100%.

Table 5. The optimization results and identification accuracy of the sub-SVM of each node in DT-SVM.

SNR	SVM-1	SVM-2	SVM-3	SVM-4	SVM-5	AVERAGE
	Acc/% (c, γ)	Acc/% (c, γ)	Acc/% (c, γ)	Acc/% (c, γ)	Acc/% (c, γ)	Acc/%
−5 dB	88.33 ($2^{11.2}, 2^{13.8}$)	100 ($2^0, 2^0$)	81.25 ($2^{0.5}, 2^3$)	95 ($2^{-2.5}, 2^{15}$)	100 ($2^0, 2^0$)	92.92
0 dB	100 ($2^{-8}, 2^2$)	100 ($2^0, 2^0$)	100 ($2^{-5}, 2^{8.5}$)	100 ($2^{-5}, 2^{7.5}$)	100 ($2^0, 2^0$)	100
5 dB	100 ($2^{-8}, 2^{-2}$)	100 ($2^0, 2^0$)	100 ($2^0, 2^0$)	100 ($2^0, 2^0$)	100 ($2^0, 2^0$)	100

In order to prove the superiority of this method in recognition accuracy, Table 6 shows the classification accuracy of six different cognitive radio signals using multidimensional HOC, cyclic spectrum and DT-SVM classifier when the kernel function is RBF. In addition, the sizes of training and testing subsets are selected as 80% and 20% of the whole set of eigenvectors. The results show that with the increase of SNR, the classification accuracy of six kinds of signal is improved. When the SNR is 0 dB, the classification accuracy is 100%. It is proved that this method still has high recognition accuracy under low SNR. In addition, a new modulation signal can be introduced to expand the flexibility of the method and shows better compatibility, which will certainly increase the complexity of the algorithm and the classification time of the whole classification system.

Table 6. The classification accuracy of the cognitive radio signals using multidimensional HOC, cyclic spectrum and DT-SVM classifier.

SNR	Classification Accuracy of Cognitive Radio Signals (%)						
	OOK	DPSK	QPSK	OQPSK	16QAM	64QAM	AVERAGE
−5 dB	72.5	72.5	74.69	74.69	83.25	83.25	76.81
0 dB	100	100	100	100	100	100	100
5 dB	100	100	100	100	100	100	100
10 dB	100	100	100	100	100	100	100
15 dB	100	100	100	100	100	100	100

6. Conclusions

We have proposed a method through simulation to identify the cognitive radio signals based on compressed sensing combined with HOC and cyclic spectrum, which has been proved to perform well in noisy situations. It successfully achieves reconstructing the feature parameters of HOC and cyclic spectrum directly from the sub-Nyquist rate rather than reconstructing the original signal. The simulation results indicate that this method can effectively achieve modulation classification for six kinds of cognitive radio signals with three feature parameters. In this paper, the proposed method is relatively simple and the feature parameters used are few, and they are also less affected by noise. By analyzing the effect of symbol length and compression rate on the classification rate, the classification performance is improved and the classification accuracy can reach 100% when the SNR is 0 dB. This technique utilizes a sampling rate of CS much lower than the Nyquist sampling rate and noise-insensitive feature extraction algorithm, realizes blind classification without any prior information from the transmitter, and has low computational complexity.

Author Contributions: X.S. and Z.Z. conceived and designed the experiments; X.S. and X.T. performed the experiments; X.S. and X.G. analyzed the data; X.S., Z.Z. and X.G. contributed the simulation software and experimental facilities; X.S. and Z.Z. wrote the paper; S.S. participated in the funding acquisition and investigation. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Frontier Science and Technology Innovation Project in the National Key Research and Development Program under Grant No. 2016QY11W2003, Natural Science Foundation of

Hunan Province under Grant No. 2018JJ3607, Natural Science Foundation of China under Grant No. 51575517 and National Technology Foundation Project under Grant No. 181GF22006.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhu, X.; Fujii, T. A modulation classification method in cognitive radios system using stacked denoising sparse autoencoder. In Proceedings of the 2017 IEEE Radio and Wireless Symposium (RWS), Phoenix, AZ, USA, 15–18 January 2017; pp. 218–220.
2. Chen, S.; Shen, B.; Wang, X.; Yoo, S. A strong machine learning classifier and decision stumps based hybrid adaBoost classification algorithm for cognitive radios. *Sensors* **2019**, *19*, 5077. [[CrossRef](#)] [[PubMed](#)]
3. Sutton, P.; Nolan, K.E.; Doyle, L. Cyclostationary signatures in practical cognitive radio applications. *IEEE J. Sel. Area. Comm.* **2008**, *26*, 13–24. [[CrossRef](#)]
4. Hu, H. Cyclostationary Approach to Signal Detection and Classification in Cognitive Radio Systems. In *Cognitive Radio Systems*; Wang, W., Ed.; Beijing University of Posts and Telecommunications: Beijing, China, 2009.
5. Ganesan, G.; Li, Y.G. Cooperative spectrum sensing in cognitive radio Part I: Two users networks. *IEEE Trans. Wirel. Commun.* **2007**, *6*, 2204–2213. [[CrossRef](#)]
6. Ma, J.; Li, G.Y.; Juang, B.H. Signal Processing in Cognitive Radio. *Proc. IEEE* **2009**, *97*, 805–823.
7. Zhao, Q.; Sadler, B.M. A Survey of Dynamic Spectrum Access. *IEEE Signal Proc. Mag.* **2007**, *24*, 79–89. [[CrossRef](#)]
8. Tandra, R.; Sahai, A. SNR walls for signal detection. *IEEE J-STSP* **2008**, *2*, 4–17. [[CrossRef](#)]
9. Li, C.; Xiao, J.; Xu, Q. A novel modulation classification for PSK and QAM signals in wireless communication. In Proceedings of the IET International Conference on Communication Technology and Application (ICCTA), Beijing, China, 14–16 October 2011.
10. Liu, J.; Luo, Q. A novel modulation classification algorithm based on daubechies5 wavelet and fractional fourier transform in cognitive radio. In Proceedings of the IEEE 14th International Conference on Communication Technology, Chengdu, China, 9–11 November 2012; pp. 115–120.
11. Liu, A.; Zhu, Q. Automatic modulation classification based on the combination of clustering and neural network. *J. China Univ. Posts Telecommun.* **2011**, *18*, 13–19. [[CrossRef](#)]
12. Xu, Y.; Li, D.; Wang, Z.; Liu, G.; Lv, H. A deep learning method based on convolutional neural network for automatic modulation classification of wireless signals. In Proceedings of the International Conference on Machine Learning and Intelligent Communications (MLICOM), Weihai, China, 5–6 August 2017; Volume 226, pp. 373–381.
13. Liu, L.; Xu, J. A novel modulation classification method based on high order cumulants. In Proceedings of the International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, China, 22–24 September 2006.
14. Chen, X.; Wang, H.; Cai, Q. Performance analysis and optimization of novel high-order statistic features in modulation classification. In Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing, Dalin, China, 12–14 October 2008.
15. Yuan, H.; Sun, X.; Li, H. The modulation recognition based on decision-making mechanism and neural network integrated classifier. *High Technol. Lett.* **2013**, *19*, 132–136.
16. Liu, N.; Liu, B.; Guo, S.; Luo, R. Investigation on signal modulation recognition in the low SNR. In Proceedings of the International Conference on Measuring Technology and Mechatronics Automation, Changsha, China, 13–14 March 2010.
17. Yoo, Y.; Baek, J. A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network. *Appl. Sci.* **2018**, *8*, 1102. [[CrossRef](#)]
18. Wang, S.; Sun, Z.; Liu, S.; Chen, X.; Wang, W. Modulation classification of linear digital signals based on compressive sensing using high-order moments. In Proceedings of the European Modeling Symposium, Pisa, Italy, 21–23 October 2014.
19. Zhang, X. *Modern Signal Processing*, 3rd ed.; Tsinghua University: Beijing, China, 2015; pp. 219–221.
20. Hui, B.; Tang, X.; Gao, N.; Zhang, W.; Zhang, X. High order modulation format identification based on compressed sensing in optical fiber communication system. *Chin. Opt. Lett.* **2016**, *14*, 14–18.

21. Candes, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **2006**, *52*, 489–509. [[CrossRef](#)]
22. Baraniuk, R.G. Compressive Sensing. *IEEE Signal Proc. Mag.* **2007**, *24*, 118–121. [[CrossRef](#)]
23. Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inform. Theory* **2005**, *51*, 4203–4215. [[CrossRef](#)]
24. Candes, E.J.; Romberg, J.K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurement. *Commun. Pur. Appl. Math.* **2006**, *59*, 1207–1223. [[CrossRef](#)]
25. Tsaig, Y.; Donoho, D.L. Extensions of compressed sensing. *Signal Process.* **2006**, *86*, 549–571. [[CrossRef](#)]
26. Tian, Z.; Tafesse, Y.; Sadler, B.M. Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing. *IEEE J-STSP* **2012**, *6*, 58–69. [[CrossRef](#)]
27. Kirolos, S.; Laska, J.; Wakin, M.; Duarte, M.; Baron, D.; Ragheb, T.; Massoud, Y.; Baraniuk, R. Analog-to-information conversion via random demodulation. In Proceedings of the IEEE Dallas/CAS Workshop on Design, Application, Integration and Software, Richardson, TX, USA, 29–30 October 2006.
28. Awe, O.P.; Deligiannis, A.; Lambbotharan, S. Spatio-temporal spectrum sensing in cognitive radio networks using beamformer-aided SVM algorithms. *IEEE Access* **2018**, *6*, 25377–25388. [[CrossRef](#)]
29. Yokota, S.; Endo, M.; Ohe, K. Establishing a classification system for high fall-risk among inpatients using support vector machines. *CIN Comput. Inform. Nurs.* **2017**, *35*, 408–416. [[CrossRef](#)] [[PubMed](#)]
30. Zhang, W. Automatic modulation classification based on statistical features and support vector machine. In Proceedings of the URSI General Assembly and Scientific Symposium (URSI GASS), Beijing, China, 16–23 August 2014.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).