

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,700

Open access books available

121,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Cyclostationary Approach to Signal Detection and Classification in Cognitive Radio Systems

Hao Hu

*Beijing University of Posts and Telecommunications  
P. R. China*

## 1. Introduction

Cognitive radio (CR) is a newly emerging technology (Mitola & Maguire, 1999, Mitola, 2001), which has been recently proposed to implement some kind of intelligence to automatically sense, recognize, and make wise use of any available radio frequency spectrum. With the increasing demand for wireless application, access to available spectrum is becoming increasingly difficult. On the other hand, most licensed spectrum go unused most of the time according to the FCC's Spectrum Policy Task Force Report (FCC, 2003). In order to solve these problems, cognitive radio is proposed for sharing the licensed spectrum to unlicensed users without harmful interference to licensed system.

Spectrum sensing is a key element in cognitive radio system which enables the cognitive radio to share the spectrum in licensed bands by detecting temporarily unused spectral resources (Haykin, 2005, Ghasemi & Sousa, 2005). Recently, several spectrum sensing techniques have been explored for cognitive radios, such as matched filter detection, energy detection and cyclostationary feature detection (Akyildiz et al., 2006). Many signals used in communication systems exhibit periodicities of their second order statistical parameters due to the operations such as sampling, modulating, multiplexing and coding. These cyclostationary properties, which are named as spectral correlation features, can be used for spectrum sensing. Moreover, spectrum sensing can not be restricted to simply monitor the power in some frequency bands of interest but must include detection and identification in order to avoid interference (Fehske et al., 2005). Therefore, cyclostationary feature detection is undoubtedly a good solution for primary user signal detection and recognition. In this chapter, several well-known spectrum sensing techniques are reviewed first. A survey of signal detection and classification for cognitive radios combining the spectral correlation analysis and support vector machine (SVM) is given in Section 3. Several spectral coherence characteristic parameters which are sensitive with modulation types and insensitive with SNR variation are chosen via spectral correlation analysis. In order to give better performance of the SVM, an alignment based kernel selection method is proposed in Section 4, which is used to choose the best kernel function for the SVM with spectral coherence characteristic training samples. A simple cross-validation method is also introduced to

choose the most appropriate kernel parameters and penalty parameters for the SVM. The performance analysis of the proposed approach is given over both Gaussian channel and IEEE 802.22 WRAN channel in Section 5. Compared to the existing methods including the classifiers based on binary decision tree (BDT) and multilayer linear perceptron network (MLPN), the proposed approach is more effective in the case of low SNR and limited training numbers.

## 2. Overview of Spectrum Sensing in Cognitive Radio System

Interference due to a cognitive radio network is deemed harmful if it causes the signal-to-interference ratio (SIR) at any primary receiver to fall below a certain threshold, which supplied by the regulatory bodies. This threshold depends on the receiver's robustness toward interference and varies from one primary band or service to another. In order to achieve spectrum sharing, the sensitivity of cognitive radio should be higher than that of primary receiver. For example, in order to share the radio spectrum between cognitive users and TV users in IEEE 802.22 standard, the sensitivity of cognitive receiver should exceed the TV receiver without hidden terminal problem. To improve the sensing accuracy, an additional margin of 30dB - 40dB should be added to the detection threshold. Moreover, due to the dynamic characteristic of the radio environment and difference between primary users as well as unknown influence of interference, spectrum sensing has become a challenging problem in cognitive radio. Generally, the spectrum sensing techniques can be classified as transmitter detection, cooperative detection, and interference-based detection, as shown in Fig. 1. In this chapter, we focus on the transmitter detection which is commonly used in the practical system. Transmitter detection approach is based on the detection of the weak signal from a primary transmitter. To achieve dynamic spectrum sharing, the cognitive radio transmitter should have capability to determine if a signal from primary user is locally present in a certain spectrum. Basic hypothesis model for transmitter detection can be defined as follows

$$x(t) = \begin{cases} n(t) & H_0 \\ hs(t) + n(t) & H_1 \end{cases} \quad (1)$$

where  $x(t)$  is the signal received by the cognitive user,  $s(t)$  is the transmitted signal of the primary user,  $n(t)$  is the additive white Gaussian noise and  $h$  is the amplitude gain of the channel.  $H_0$  is a null hypothesis, which states that there is no licensed user signal in a certain spectrum band. On the other hand,  $H_a$  is an alternative hypothesis, which indicates that there exists some licensed user signals.

Three approaches exist for transmitter detection, based on the sensing users' knowledge on the transmitted signals, which are illustrated in Fig. 1.

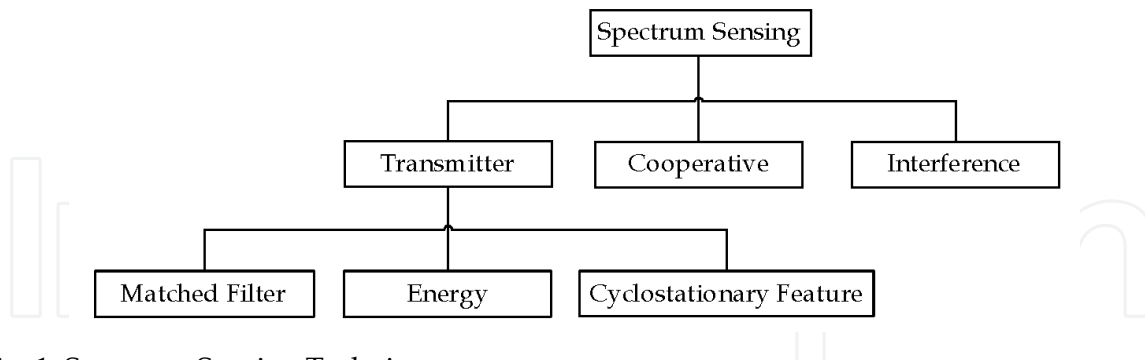


Fig. 1. Spectrum Sensing Techniques.

### 2.1 Matched Filter Detection

When the information of the primary user signal is well-known to the cognitive radio, the matched filter is the optimal linear filter for maximizing the signal to noise ratio (SNR) in the presence of additive Gaussian noise (Sahai et al., 2004, Akyildiz et al., 2006). In cognitive radio system, the matched filter is obtained by correlating a known signal, or primary user signal template, with an unknown signal to detect the presence of primary user signal in the unknown signal. This is equivalent to convolving the unknown signal with a time-reversed version of the template. According to the hypothesis model for transmitter detection, the received signal  $x(n)$  can be expressed as

$$x(n) = \begin{cases} w(n) & H_0 \\ \sqrt{1-\eta}s(n) + \sqrt{\eta}s_c(n) + w(n) & H_1 \end{cases} \quad (2)$$

where  $s(n)$  and  $s_c(n)$  are original signal and pilot signal transmitted by primary user, which are orthogonal.  $\eta$  is the ratio of power allocated to pilot signal. The correlation function of received signal  $x(n)$  and unit vector of pilot signal  $\hat{s}_c(n)$  is given by

$$T(x) = \frac{1}{N} \sum_{n=1}^N x(n)\hat{s}_c(n) \quad (3)$$

The Equation (3) can also be expressed as

$$T(x) = \begin{cases} \frac{\sum_{n=1}^N w(n)\hat{s}_c(n)}{N} & \sim N(0, \sigma^2/N) \quad H_0 \\ \frac{\sum_{n=1}^N \{\sqrt{1-\eta}s(n) + \sqrt{\eta}s_c(n) + w(n)\}\hat{s}_c(n)}{N} & \sim N(\sqrt{\theta P}, \sigma^2/N) \quad H_1 \end{cases} \quad (4)$$

where  $P$  is the power of the pilot signal, and  $\sigma^2$  is noise variance. Then the decision function can be given by

$$T(x) \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \gamma \quad (5)$$

where the threshold  $\gamma$  is determined by the probability of false alarm  $P_{FA}$ . The probability of false alarm  $P_{FA}$  and detection  $P_D$  are given as follows

$$P_{FA} = P(T(x) > \gamma | H_0) = Q\left(\gamma / \sqrt{\frac{\sigma^2}{N}}\right) \quad (6)$$

$$P_D = P(T(x) > \gamma | H_1) = Q\left(\gamma - \sqrt{\theta P} / \sqrt{\frac{\sigma^2}{N}}\right) \quad (7)$$

where  $Q(\cdot)$  is the generalized Marcum Q-function, which can be defined by

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (8)$$

Combining Equation (6) and Equation (7), we can obtain  $N$  by eliminating  $\gamma$ .

$$N = \left[ Q^{-1}(P_D) - Q^{-1}(P_{FA}) \right]^2 \theta^{-1} \text{SNR}^{-1} \quad (9)$$

According to the Equation (9),  $O(1/\text{SNR})$  samples are required to meet a probability of error constraint, thus the main advantage of the matched filter is that it requires less time to achieve high processing gain. However, as a coherent detection method, matched filter detection requires a priori knowledge of the primary user signal such as the modulation type and order, the pulse shape, and the packet format. Although most of the priori knowledge can be obtained from pilot, preambles, synchronization word or spreading codes of the primary user network systems, an obvious shortcoming is that the cognitive radio user requires specific receivers for different types of primary user signal.

## 2.2 Energy Detection

If the receiver cannot gather sufficient information about the primary user signal, for example, if the power of the random Gaussian noise is only known to the receiver, the optimal detector is an energy detector (Akyildiz et al., 2006, Digham et al., 2003). In order to measure the energy of the received signal, the output signal of bandpass filter with bandwidth  $W$  is squared and integrated over the observation interval  $T$ . Finally, the output of the integrator,  $x(t)$ , is compared with a threshold,  $\gamma$ , to decide whether a licensed user is present or not.

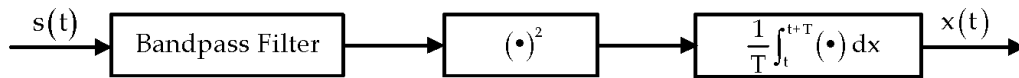


Fig. 2. The principal of energy detection.

According to the basic hypothesis model for transmitter detection, the probability of false alarm  $P_{FA}$  and detection  $P_D$  are given as follows

$$P_{FA} = P\{X > \gamma | H_1\} = Q(\sqrt{2\delta}, \sqrt{\gamma}) \quad (10)$$

$$P_D = P\{X > \gamma | H_0\} = \frac{\Gamma(m, 2/\gamma)}{\Gamma(m)} \quad (11)$$

Since it is easy to implement, the recent work on detection of the primary user has generally adopted the energy detection. Although this method can be implemented without any prior knowledge of the primary user signal, it also has some drawbacks. Since energy detection is non-coherent,  $O(1/\text{SNR})$  samples are required to meet a probability of error constraint. Moreover, the threshold selection for energy detection is highly susceptible to uncertainty in background noise and interference, and it can only determine the presence of the signal without differentiating signal types. However, the largest advantage of energy detection is simple and low complexity.

### 2.3 Cyclostationary Feature Detection

The cyclostationary feature is intentionally embedded in the physical properties of a communication signal, which may be easily generated, manipulated, detected and analyzed using low complexity transceiver architectures. This feature is present in all transmitted signals, requires little signalling overhead and may be detected using short signal observation times, and thus it can be used for primary user signal detection and recognition. Recent research efforts exploit the cyclostationary features of signals via spectral correlation analysis as a method for spectrum sensing (Digham et al., 2003, Sahai et al., 2004, Ghasemi & Sousa, 2005), which has been found to be superior to simple energy detection and matched filtering. Energy detection can only detect whether or not a signal is present and utilizing a matched filter system requires extensive knowledge about the channel and signals that are to be identified. The method, which is not susceptible to in-band interference, can be used to detect and classify different types of signal. Conventional signal classification approaches are mainly based on decision theory (Polydoros & Kim, 1990, Sapiano & Martin, 1996, Sills, 1999, Wei & Mendel, 2000, Hang et al., 2001) and statistical pattern recognition (Nandi & Azzouz, 1995, Azzouz & Nandi, 1995, Azzouz & Nandi, 1996, Nandi & Azzouz, 1998). In the work by Hang (Hang et al., 2001), a binary decision tree is designed to signal classification. However, it's difficult to obtain the decision thresholds and rules, which needs a large amount of calculation. For more efficient and reliable performance, a novel approach based on multilayer linear perceptron network for signal classification in cognitive radio is studied by Fehske (Fehske et al., 2005). Support vector machine (SVM) is a new statistical pattern recognition approach, which is based on structural risk minimization principle (Vapnik,

1995). Compared with the conventional methods based on empirical risk minimization like artificial neural network (ANN), it has been found to give better generalization and better performance for small training examples. In the next section, a novel approach of signal classification for cognitive radios combining the cyclostationary features and SVM is proposed.

### 3. Spectrum Sensing based on spectral correlation analysis and SVM

#### 3.1 System Framework

As our signal classification scheme combining spectral correlation analysis and SVM is based on statistical pattern recognition, which mainly consists of three modules (Han, 2003): feature extraction, classifier design and classification decision. Feature extraction is typically the first stage in any classification system in general, and in our spectrum sensing systems in particular. Given signal set to be classified, the feature parameters of different classes of signal and rules for classifier should be determined first. In order to achieve better classification performance, selected feature parameter should be insensitive with the SNR variation, and then a proper classifier is designed for specific classification problem using training data with known signal types. When the error probability of the classifier achieves a specific threshold, the classifier can be used for signal classification and recognition. In our scheme, there are three procedures adopted for primary user signal recognition:

1) Pro-processing procedure:

Several feature parameters are extracted via spectral correlation analysis first. The feature parameters insensitive with the SNR variation are selected as feature vector  $x_1, x_2, \dots, x_M$ .

After that, the feature vector and signal type  $y_k$  are used to form the training set  $(x_1, x_2, \dots, x_M, y_k)$  for classifier.

2) Training and learning procedure:

The SVM classifier is trained using selected feature parameters in the training set. By utilizing a nonlinear SVM, an amount of calculation for training is performed offline, thus the computational complexity is reduced. The optimal classification plane for SVM is obtained in this procedure via training and learning.

3) Test procedure

Selected feature parameters extracted for received signals are inputted well-trained SVM classifier for primary user signal detection and recognition.

The framework of our scheme combining spectral correlation analysis and SVM is shown in Fig. 3.



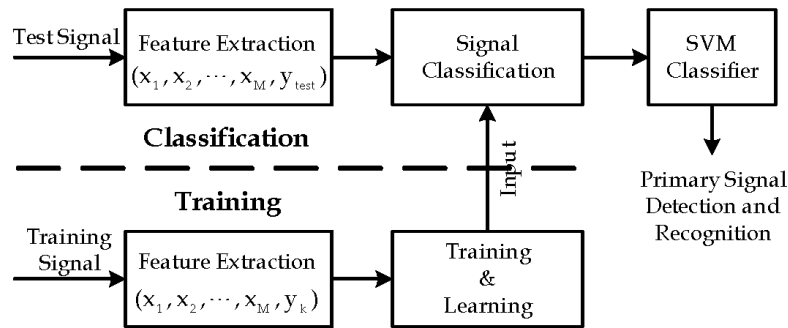


Fig. 3. System framework.

Although SVM is a better choice for the classifier, the selection of feature parameters has direct impact on the performance of the classification algorithm. In the next section, we will discuss the first step, how to choose spectral coherence characteristic parameters for our scheme.

### 3.2 Spectral Correlation Analysis

Many signals used in communication systems exhibit periodicities of their second order statistical parameters due to the operations such as sampling, modulating, multiplexing and coding. These cyclostationary properties, which are named as spectral correlation features, can be used for signal detection and recognition (Gardner, 1987).

In order to analyze the cyclostationary features of the signal  $x(t)$ , two key functions are typically utilized. The cyclic autocorrelation function (CAF) is used for time domain analysis, which can be expressed as

$$R_x^\alpha(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\alpha t} dt$$

$$= \left\langle x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\alpha t} \right\rangle \quad (12)$$

The spectral correlation function (SCF), which exhibits the spectral correlation of the signal  $x(t)$ , is obtained from the Fourier transform of the cyclic autocorrelation in Equation (12) (Gardner & Franks, 1975).

$$S_x^\alpha(f) = \int_{-\infty}^{+\infty} R_x^\alpha(\tau) e^{-j2\pi f\tau} d\tau$$

$$= \int_{-\infty}^{+\infty} \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-j2\pi\alpha t} e^{-j2\pi f\tau} dt d\tau \quad (13)$$

$$= \lim_{T \rightarrow +\infty} S_{xT}^\alpha(t, f)$$



Where  $\alpha$  is the cycle frequency,  $f$  is the spectral frequency and  $S_{xT}^{\alpha}(t, f)$  is the cyclic periodogram of  $S_x^{\alpha}(f)$

$$S_{xT}^{\alpha}(t, f) = [X_T(t, f + \alpha/2) \cdot X_T^*(t, f - \alpha/2)] / T \quad (14)$$

where the Fourier transform of the function  $x(u)$  on the bounded time interval  $[t - T/2, t + T/2]$  is defined as

$$X_T(t, f) = \int_{t-T/2}^{t+T/2} x(u) e^{-j2\pi fu} du \quad (15)$$

The correlation coefficient for the SCF between frequency components  $f \pm \alpha/2$ , which is known as spectral coherence coefficient (SCC), can be calculated by

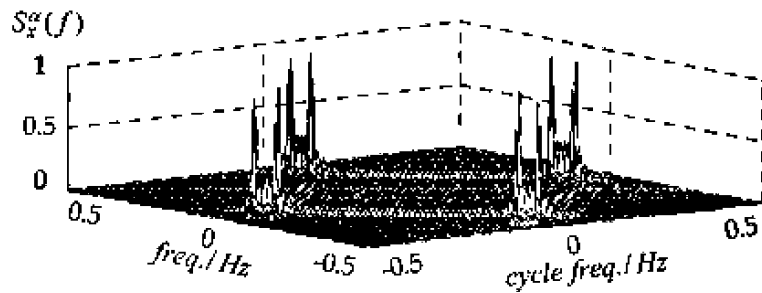
$$C_x^{\alpha}(f) = \frac{S_x^{\alpha}(f)}{\sqrt{S_x^0(f + \alpha/2) \cdot S_x^0(f - \alpha/2)}} \quad (16)$$

The magnitude of the SCC ranges from 0 to 1 with  $a = 0$  for all  $f$ . Different signal classes (*i.e.* AM, ASK, FSK, PSK, MSK, QPSK) can be distinguished based on several characteristic parameters of SCF and SCC.

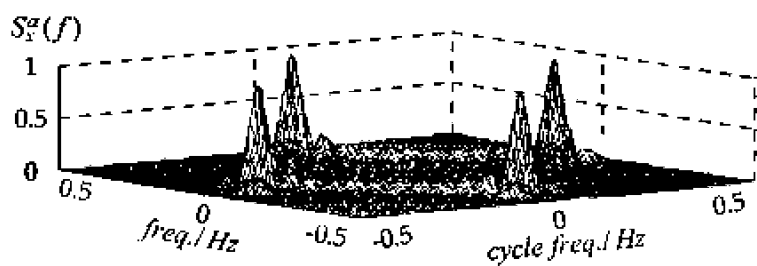
In practical situations, however, the number of observation samples at the sensor is limited. Therefore, the spectral correlation function needs to be estimated from a finite set of samples. In general, two methods are used for spectral correlation estimation including time-domain averaging and frequency-domain smoothing (Gardner & Spooner, 1988). In this section, the frequency-smoothing method is used for spectral correlation estimation, which can be expressed as follows.

$$\left\{ \begin{array}{l} S_x^{\alpha}(f)_{\Delta f} = \frac{1}{M} \sum_{n=(M-1)/2}^{(M+1)/2} \frac{1}{\Delta t} X_{\Delta t}(t, f + \frac{\alpha}{2} + nF_s) \cdot X_{\Delta t}^*(t, f - \frac{\alpha}{2} + nF_s) \\ X_{\Delta t}(t, f) = \sum_{k=0}^{N-1} W_{\Delta t}(kT_s) \cdot X(t - kT_s) e^{-j2\pi f(t - kT_s)} \end{array} \right. \quad (17)$$

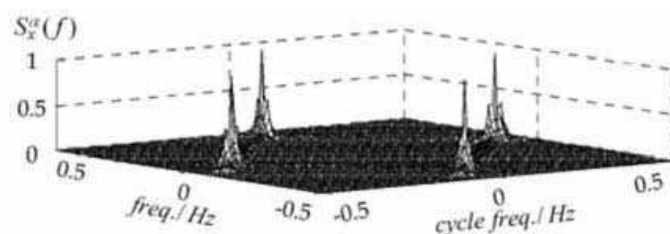
where  $F_s = 1/(N - 1)T_s$ ,  $T_s$  is the frequency increment,  $T_s$  is the cyclic sampling interval and the length of the sample is  $N = \Delta t / T_s$ . Thus, the data-tapering window  $W_{\Delta t}(kT_s)$  is of the width  $\Delta f = M \cdot F_s$ . The Spectral correlation functions of some typical signals can be obtained by frequency-smoothing method, which are shown in Fig. 4 (Gurman et al., 2008).



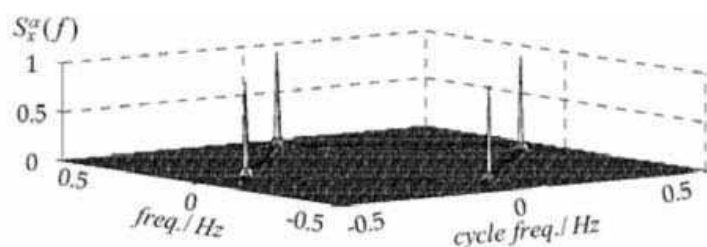
(a) FSK



(b) PSK



(c) FM



(d) AM

Fig. 4. Spectral correlation functions of some typical signal classes.

Even visually in the above figures, the spectral correlation functions of the different modulation types possess distinct characteristics. It is this fact that allows the successful application of the pattern recognition techniques to achieve primary user signal detection and recognition. In order to obtain better robustness of the proposed algorithm, some features less sensitive with SNR should be chosen for the classifier. Assumed that the

received signal  $s = s + n$ , where  $s$ ,  $n$  are the transmitted signal and additional white Gaussian noise. The feature parameter of the received signal has a better classification where performance which is insensitive with SNR variation, if it satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \leq \|\mathbf{n}\|_2 \quad (18)$$

$\mathbf{X}$ ,  $\mathbf{x}$ ,  $\mathbf{n}$  is the feature vectors of  $s$ ,  $s$ ,  $n$ .

Based on the calculation of the spectral correlation function, we can obtain the spectral correlation magnitude surface of different types of signal. According to above analysis, several spectral correlation features can be extracted for distinction of different modulation types. Typically, four key features  $x_1, x_2, x_3, x_4$  which are sensitive with modulation types and insensitive with SNR variation are listed in Table. 1.

Feature	Signal Type					
	AM	ASK	FSK	PSK	MSK	QPSK
$x_1$	2	2	4	0	0	0
$x_2$	1	$\geq 3$	2	$\geq 3$	$\geq 2$	$\geq 2$
$x_3$	1	1	/	1	0	0
$x_4$	/	/	/	$\geq 6$	$< 6$	$< 6$

Table 1. Typical value of spectral correlation features. Four key features  $x_1, x_2, x_3, x_4$  are described as follows.

- $x_1$ : Number of 5 pulse on  $f$ -domain of SCF  
Let  $a = 0$  in Equation (17), the SCF is transformed into  $S''(f)$ , thus  $x_1$  can be obtained from the ichnography of  $S_x^{(0)}(f)$ .
- $x_2$ : Number of cyclic spectral line on  $a$ -domain of SCF  
Let  $f = 0$  in Equation (17), the SCF is transformed into  $S^a(0)$ , and  $x_2$  can be obtained from the ichnography of  $S^a(0)$ .
- $x_3$ : Average energy of cyclic spectral line on  $a$ -domain of SCF  
The Average energy of cyclic spectral line on  $a$ -domain of SCF can be computed by the equation as follows:

$$x_3 = \int_0^{+\infty} |S_x^a(f)|^2 d\alpha \quad (19)$$

- $x_4$ : Maximum value of SCC  
The spectral coherence coefficient can be obtained via Equation (16) and Equation (17). Then, the maximum value of SCC is computed as key feature  $x_4$ .

In the simulation experiments, 4000 features are extracted from the signal for every trial. In order to prevent numerical computational errors, the features need to be normalized by

subtracting mean of each feature from the original feature and dividing the result by the standard deviation of the same feature.

$$x'_i = (x_i - \bar{x}_i) / \delta_{x_i} \quad i = 1, 2, 3, 4 \quad (20)$$

After normalization, the feature vector  $x = (x'_1, x'_2, x'_3, x'_4)$  can be generated as the input of the signal classifier.

### 3.3 Support Vector Machine

The traditional statistical theory is primarily based on the asymptotic principle, which provides conclusion only for the situation where the sample size is tending to infinity. However, in most practical applications, the samples are usually limited so that it is difficult to achieve the desired results via existing methods. Statistical Learning Theory is a novel statistical theory based on small sample statistics by Vapnik (Vapnik, 1995). Compared to the conventional statistical theory, statistical learning theory mainly concerns the statistic principles when samples are limited, especially the properties of learning procedure in such cases. Statistical learning theory provides us a new framework for the general learning problem, which not only considers the asymptotic performance but obtains the optimal results under the condition of limited information. In order to study the generalization performance and the speed of uniform convergence, a series of indicators used to evaluate the learning performance of function sets are defined in statistical learning theory. One of the most important concepts is Vapnik-Chervonenkis (VC) dimension which was originally defined by Vladimir Vapnik and Alexey Chervonenkis in 1971. VC-dimension is a measure of learning machine complexity or the capacity of a statistical learning algorithm, which is the cardinality of the largest set of points that the algorithm can shatter. The learning machine is more complex with a greater VC-dimension. Statistical learning theory provides a novel strategy that balances the empirical risk and confidence interval. A nested subset sequence is chosen from the given set of functions according to the size of the VC dimension. For a given subset, the minimal value of the empirical risk can be obtained as the minimal true risk, which is illustrated in Fig. 5. This method is named as Structural risk minimization (SRM) which was also coined by Vapnik and Chervonekis in 1974. The principle of SRM is to provide a method to reach the trade-off between hypothesis space complexity (the VC dimension of approximating functions) and the quality of fitting the training data (Wang, 2007). The procedure is described in detail as follows.

Assumed a function set  $Q(x, \alpha), \alpha \in \Lambda$  has a set  $S$ , which has a structure shown in Fig. 5. The structure is defined by a nested subset sequence  $S_1, S_2, \dots, S_n$ , which satisfies

$$S_1 \subset S_2 \subset \dots \subset S_n \subset \dots \quad (21)$$

The elements of the above structure have two properties as follows.

- (1) The VC dimension of each subset  $h_k$  is limited and satisfies

$$h_1 \leq h_2 \leq \dots h_n \leq \dots \quad (22)$$

(2) Any element in the structure  $S_k$  contains a set of totally bounded function

$$0 \leq Q(z, \alpha) \leq B_k, \quad \alpha \in \Lambda_k \quad (23)$$

or contains a function set which satisfies the following inequality for some  $(p, T_k)$

$$\sup_{\alpha \in \Lambda_k} \frac{\left( \int Q^p(z, \alpha) dF(z) \right)^{1/p}}{\int Q(z, \alpha) dF(z)} \leq \tau_k, \quad p > 2 \quad (24)$$

For a given set of observation set  $z_1, z_2, \dots, z_l$ , SRM aims to choose the proper function  $Q(z, \alpha_1^k)$  who makes the empirical risk minimal in subset  $S_k$  with smallest guaranteed risk. After performing empirical risk minimization on each subset, the subset whose sum of empirical risk and VC confidence is selected. The optimal function  $Q(z, \alpha_1^k)$  that makes the empirical risk minimal in selected subset is the optimal solution.

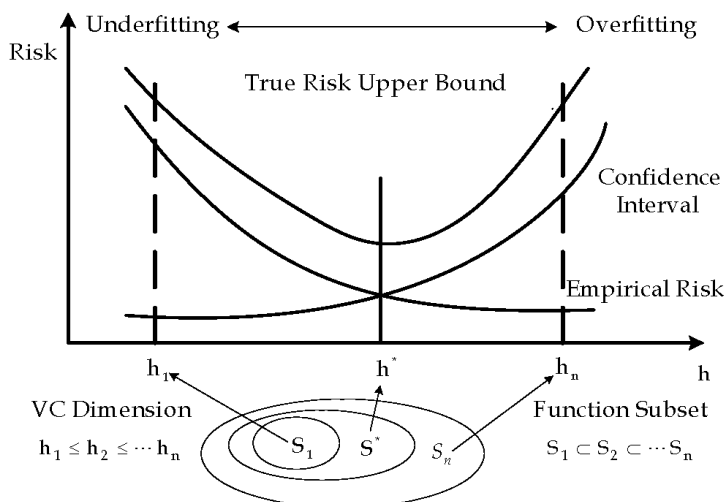


Fig. 5. The principle of Structural Risk Minimization.

Statistical learning theory also gives the required conditions for reasonable structure of function subset and the convergence property of actual risk in SRM principle. The actual risk is the sum of empirical risk and confidence interval. As the index of the elements in the structure increases, the empirical risk will be reduced with extended confidence interval. The smallest upper bound of the actual risk can be derived from a certain element in the structure. Support vector machine is a novel universal learning machine, which is widely used in the fields of pattern recognition, regression estimation and probability density. It is based on VC-dimension theory and SRM principle, which has a better generalization

performance by reaching a trade-off between model complexity with limited sample data and capacity of the learning algorithm. The support vector machine was coined by Vapnik in the late 1960s on the foundation of statistical learning theory. It was originally developed for binary classification problem. The optimal solution of SVM for a linearly separable case was introduced by Vapnik. Later this was extended to non-separable cases. In the previous research, a common solution to classification problem of communication signals is artificial neural network (ANN), such as multilayer linear perceptron network. After the first preliminary studies, SVM have shown a remarkable efficiency, especially when compared with traditional artificial neural networks. The main advantage of SVM, with respect to ANN, consists in the structure of the learning algorithm, characterized by the resolution of a constrained quadratic programming problem (CQP), where the drawback of local minima is completely avoided (Boser et al., 1992, Cortes & Vapnik, 1995, Scholkopf, 1995). Since the classification of communication signals is obvious to be a linearly non-separable problem, we will only discuss the computation of this optimization problem in this chapter. Given linearly non-separable classification problem, we suppose a training set is  $\{(x_i, y_i)\}$ , where  $x \in R^N$ ,  $y \in \{+1, -1\}$ ,  $i = 1, \dots, l$ . Assumed that the signal group can be classified by a hyperplane which is defined by

$$\mathbf{w}^T \Phi(\mathbf{x}) + b = 0 \quad (25)$$

where the vector  $\mathbf{w}$  defines the boundary of different classes of data,  $b$  is a scalar threshold and  $\Phi(\mathbf{x}) = \|\mathbf{x}\|^2 / 2$ .

In order to solve this non-separable problem, the non-minus slack variables  $\xi_i$  are introduced, and then the training vectors must satisfy

$$y_i (\mathbf{w}^T \mathbf{x}_i + b_i) - 1 \geq 1 - \xi_i, \quad i = 1, 2, \dots, l \quad (26)$$

The hyperplane, which makes  $O(\mathbf{w}) = \|\mathbf{w}\| / 2$  to be minimum, is named as the optimal hyperplane. All of the training vectors are correctly classified by it and the vectors of each class are separated with a maximum margin (Burges, 1997).

Usually, structuring a hyperplane is solved as a quadratic optimization problem that can be formulated as

$$\begin{aligned} \min_{\mathbf{w}, b, \xi_i} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \\ \text{s. t.} \quad & y_i (\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned} \quad (27)$$

where  $C$  is the penalty parameter, which is used to control the training error rate by different values.

Using a Lagrange multiplier technique, the optimization problem can be converted into

$$F = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l \xi_i - \sum_{i=1}^l \alpha_i [y_i (\mathbf{w}^T \Phi(\mathbf{x}_i) + b) - 1 + \xi_i] - \sum_{i=1}^l \beta_i \xi_i \quad (28)$$

where  $\alpha_i, \beta_i (> 0)$  are Lagrange multiplier factors.

Given linearly non-separable classification problem, we can map the input data into a high dimensional feature space through some non-linear transformation which makes the data linearly separable (Devroye, 1996). Noted the above solution to linearly non-separable classification problem, only the inner product operation of the training samples is involved in the decision function. While structuring high dimensional feature space, the algorithm only use the inner product  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})$  in the space without separated  $\Phi(\mathbf{x})$  or  $\Phi(\mathbf{x}_i)$ . If we can find a function  $K$  satisfying Mercer condition (Daniel & James, 2000), which can be denoted as

$$K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}) \quad (29)$$

where  $K(\mathbf{x}_i, \mathbf{x})$  is the kernel function, which is utilized for mapping the input data to higher dimensional space in order to reduce the computational load. There are different kernel functions like polynomial, sigmoid and radial basis function (RBF) used in SVM, which are defined as follows.

#### 1. Polynomial Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = [(\mathbf{x} \cdot \mathbf{x}_i) + 1]^k \quad (30)$$

A  $k$ -order polynomial classifier can be defined by Equation (30).

#### 2. Radial Basis Function (RBF) Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\{-\|\mathbf{x} - \mathbf{x}_i\|^2 / \sigma^2\} \quad (31)$$

The width of the RBF kernel parameter  $\sigma$  can be determined in general by an iterative process selecting an optimum value based on the full feature set. The main difference between RBF classifier and traditional RBF method is that each basis function in the RBF classifier corresponds to a support vector, which is automatically identified by the algorithm where the drawback of local minima is completely avoided.

#### 3. Sigmoid Kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \tanh(v(\mathbf{x} \cdot \mathbf{x}_i) + c) \quad (32)$$

This kernel uses sigmoid function as inner product, which is equivalent to a multilayer perceptron with only one hidden layer. The number of node in hidden layer is automatically determined by algorithm.

Till now, the choice of the kernel functions was often used empirically, and this also became a theoretical drawback of SVM. A proper kernel function for a specific problem is dependent on the specific training sample data. In the practical applications, how to choose the proper model according to training sample set with better generalization ability is currently a research direction in the field of SVM. For a signal classification problem using cyclostationary features, we use an improved method of model selection based on kernel alignment, which will be described in Section 4.1 in detail. The choice of the kernel functions is studied via computer simulations and optimal results are achieved using radial-basis



function (RBF) kernel function. A typical classification experiment using RBF kernel function based SVM is illustrated as follows

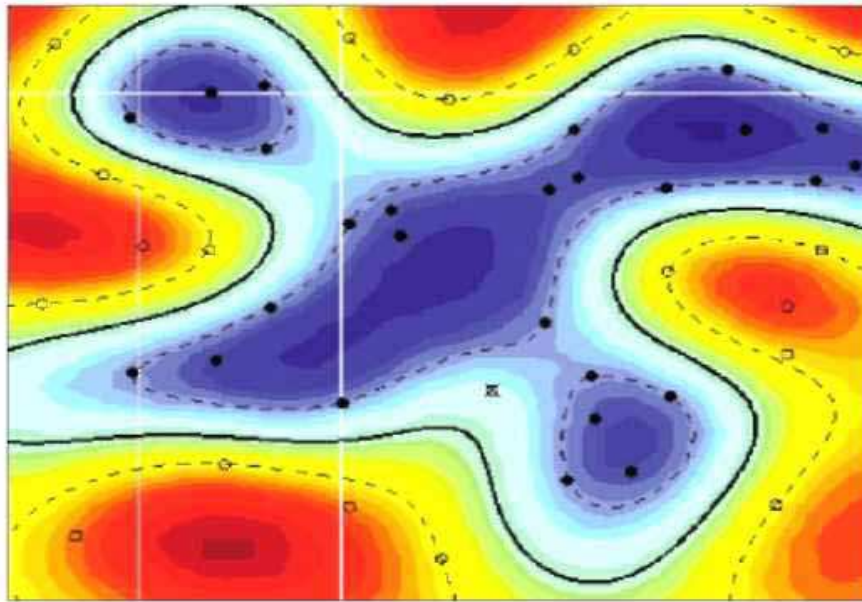


Fig. 6. A typical classification experiment using RBF kernel function (Sherrod, 2008).

After choosing the best kernel function, the dual representation of the optimization problem can be obtained by computing the derivatives with respect to  $w$ ,  $b$ ,  $\xi_i$ , which is described as

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s. t.} \quad & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l \\ & \mathbf{y}^T \alpha = 0 \end{aligned} \quad (33)$$

The resulting decision function is obtained as follows:

$$f(\mathbf{x}) = \text{sgn} \left\{ \sum_{i=1}^l \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right\} \quad (34)$$

The architecture of the SVM classifier combining spectral correlation analysis is shown in Fig. 7.

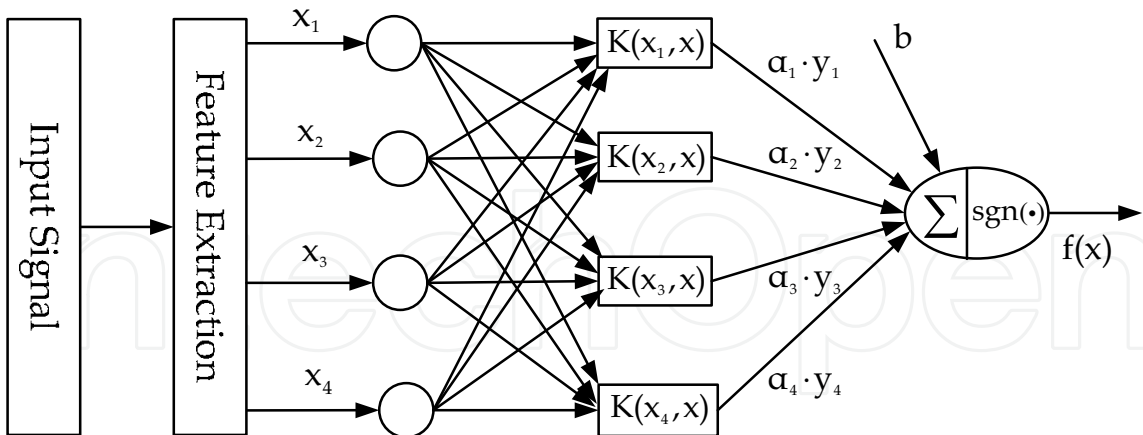


Fig. 7. Architecture of the SVM classifier combining spectral correlation analysis.

#### 4. Performance Evaluation and Analysis 4.1 Kernel Function and Parameters Selection

According to the definition of kernel function in the previous section, the kernel matrix can be defined as follows (Lanckriet et al., 2002)

$$\mathbf{K} = \left( k(x_i, x_j) \right)_{i,j=1}^n = \begin{bmatrix} K_{1,1} & K_{1,2} & \cdots & K_{1,l-1} & K_{1,l} \\ K_{2,1} & K_{2,2} & \cdots & K_{2,l-1} & K_{2,l} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ K_{l-1,1} & K_{l-1,2} & \cdots & K_{l-1,l-1} & K_{l-1,l} \\ K_{l,1} & K_{l,2} & \cdots & K_{l,l-1} & K_{l,l} \end{bmatrix} \quad (35)$$

where  $n$  is the number of the samples. It is a symmetric positive definite matrix, and since it specifies the inner products between all pairs of input elements, it completely determines the relative positions between those points in the embedding space.

In this section, we use an improved method of kernel selection based on kernel alignment to choose proper kernel function for our scheme (Cristianini et al., 2002). Assumed that  $\mathbf{K}_1$  and  $\mathbf{K}_2$  are kernel matrix of the kernel function  $k_1$  and  $k_2$ , respectively. The (empirical) alignment of a kernel  $k_1$  with a kernel  $k_2$  with respect to the sample  $S$  is the quantity, which can be defined by

$$\hat{A}(S, k_1, k_2) = \frac{\langle \mathbf{K}_1, \mathbf{K}_2 \rangle}{\sqrt{\langle \mathbf{K}_1, \mathbf{K}_1 \rangle \times \langle \mathbf{K}_2, \mathbf{K}_2 \rangle}} \quad (36)$$

Given a sample set  $S = \{x_i | x_i \in S, i = 1, 2, \dots, l\}$  for a specific signal classification problem, the sample can be divided into two types which are identified by +1 and -1. The signal type

set is denoted as  $Y = \{y_i | y_i \in \pm 1, i = 1, 2, \dots, L\}$  with the vector form  $Y = \{y_1, y_1, \dots, y_1\}^T$ , and then alignment matrix can be defined by

$$\mathbf{K}_{ad} = \mathbf{Y}\mathbf{Y}^T = \begin{bmatrix} y_1 y_1 & y_1 y_2 & \cdots & y_1 y_{l-1} & y_1 y_l \\ y_2 y_1 & y_2 y_2 & \cdots & y_2 y_{l-1} & y_2 y_l \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ y_{l-1} y_1 & y_{l-1} y_2 & \cdots & y_{l-1} y_{l-1} & y_{l-1} y_l \\ y_l y_1 & y_l y_2 & \cdots & y_l y_{l-1} & y_l y_l \end{bmatrix} \quad (37)$$

This kernel selection method using kernel alignment is based on an important assumption that the kernel function has better performance if the kernel alignment of the kernel matrix and the alignment matrix is higher. Thus, if we consider  $\mathbf{K} = \mathbf{K}_1$ ,  $\mathbf{K}_{ad} = \mathbf{Y}\mathbf{Y}^T = \mathbf{K}_2$ , then

$$\hat{A}(S, \mathbf{K}, \mathbf{Y}\mathbf{Y}^T) = \frac{\langle \mathbf{K}, \mathbf{Y}\mathbf{Y}^T \rangle}{\sqrt{\langle \mathbf{K}, \mathbf{K} \rangle \times \langle \mathbf{Y}\mathbf{Y}^T, \mathbf{Y}\mathbf{Y}^T \rangle}} = \frac{\langle \mathbf{K}, \mathbf{Y}\mathbf{Y}^T \rangle}{\sqrt{\langle \mathbf{K}, \mathbf{K} \rangle}} \quad (38)$$

According to the above derivation, the optimal kernel function problem can be transformed into kernel alignment maximizing problem. In this section, the kernel alignment values of different kernel functions are compared via computer simulation by MATLAB 7.0. For the simulations, we define a signal set as {AM, ASK, FSK, PSK, MSK, QPSK}. To obtain the kernel alignment at different SNR, simulations are carried out with 1024 samples at SNR ranging from 0 dB to 20 dB. Simulation results show that the kernel alignment of the RBF kernel is the greater than that of other kernels, which is shown in Fig. 8. According to the simulation results, we choose RBF kernel as the kernel function of the SVM in our scheme. After the kernel function is selected, two key parameters of the SVW should be considered next. The first parameter, penalty parameter  $C$  of the SVM, is used for adjusting the range of the confidence interval to control the training error rate by different values. The second one, the width of the RBF kernel parameter  $\sigma$ , can control the classification error by changing the largest VC dimension of linear classification plane. Therefore, these two parameters have a great impact on the classification performance (Chapelle et al., 2002). In this section, we use a simple cross-validation method to search the best parameters  $(C, \sigma)$ .

In  $n$ -fold cross-validation, we first divide the training set into  $n$  subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining  $n-1$  subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data that are correctly classified. The process of cross-validation is described as follows.

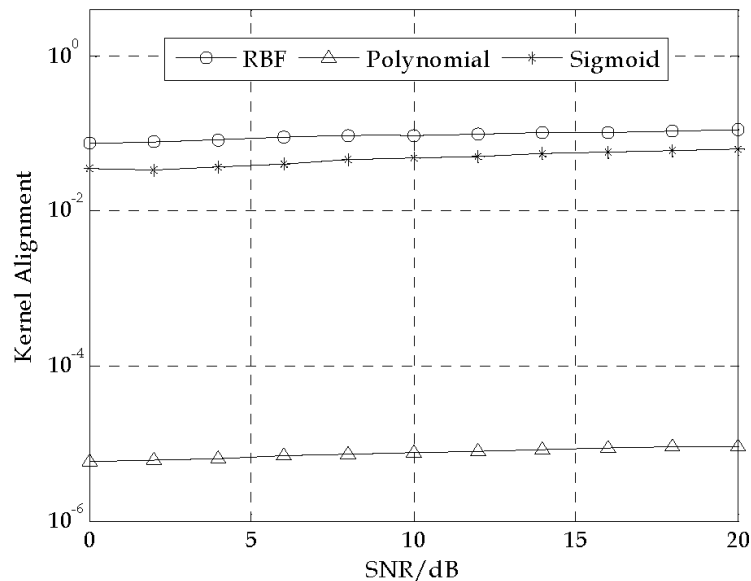


Fig. 8. Performance comparison of different kernel functions.

1. The training set is divided into  $n$  subsets  $(\Gamma_0, \Lambda_0)$  of equal size, and the elements of the subset are  $\Gamma_0 = \{C_{0,1}, C_{0,2}, \dots, C_{0,n}\}$ ,  $\Lambda_0 = \{\sigma_{0,1}, \sigma_{0,2}, \dots, \sigma_{0,n}\}$   $n$  is typically from 4 to 10.
2. After training each  $(C_{i,j}, \sigma_{i,j})$  using the classifier, the classification performance is tested.
3. The  $(C_{i,j}, \sigma_{i,j})$  with best classification performance is used to form the parameter set  $\Gamma_1 = \{C_{1,1}, C_{1,2}, \dots, C_{1,n}\}$  and  $\Lambda_1 = \{\sigma_{1,1}, \sigma_{1,2}, \dots, \sigma_{1,n}\}$ ,  $\sigma_{1,n} - \sigma_{1,1} \ll \sigma_{0,n} - \sigma_{0,1}$
4. Repeat step 3 until the number of  $\Gamma_i$  and  $\Lambda_i$  reach to a preconcerted threshold. The  $(C_{i,j}, \sigma_{i,j})$  will be the optimal choice for the specific SVM after iterations.

After performing cross-validation method for our scheme by MATLAB, we can obtain the best kernel parameters  $\sigma^2 = 0.476$  and  $C = 0.212$ . After the best  $(C, \sigma)$  is found, the whole training set is trained again to generate the final SVM classifier.

## 4.2 Classifier Design

In order to compare the performance of different classifiers, two approaches based on existing methods, such as decision theory and artificial neural network, are introduced with spectral correlation features as training data.

### 4.2.1 Binary Decision Tree

After decades of research, decision theory has been widely studied in mathematics, statistics and communication concerned with identifying the values, uncertainties and other issues

relevant in a given decision and the resulting optimal decision. In the conventional decision theory, the binary decision tree (BDT) is a decision support tool that uses a graph or model of decisions mapping from observations to target value. Since it is simple and easy to understand, binary decision tree is widely used in signal detection and recognition. After observation of the value range the different features in Table. 1, it's easy to find that feature  $x$  can be used to classify the signals into three groups, which are {PSK, MSK, QPSK}, {FSK} and {AM, ASK}. Furthermore, feature  $x_2$  and  $x_4$  can be used to distinguish {ASK, AM} and {MSK, QPSK}, respectively. Thus, a binary decision tree is designed based on spectral correlation features for primary user signal recognition. In the decision algorithm given in Fig. 9, we make use of feature  $x$  to recognize the FSK signal in the first layer, and then feature  $x_2$  and  $x_3$  are used for the classification of ASK and AM signals and recognition of PSK signal in the second layer. In the third level, feature  $x_4$  is utilized to distinguish MSK and QPSK signal.

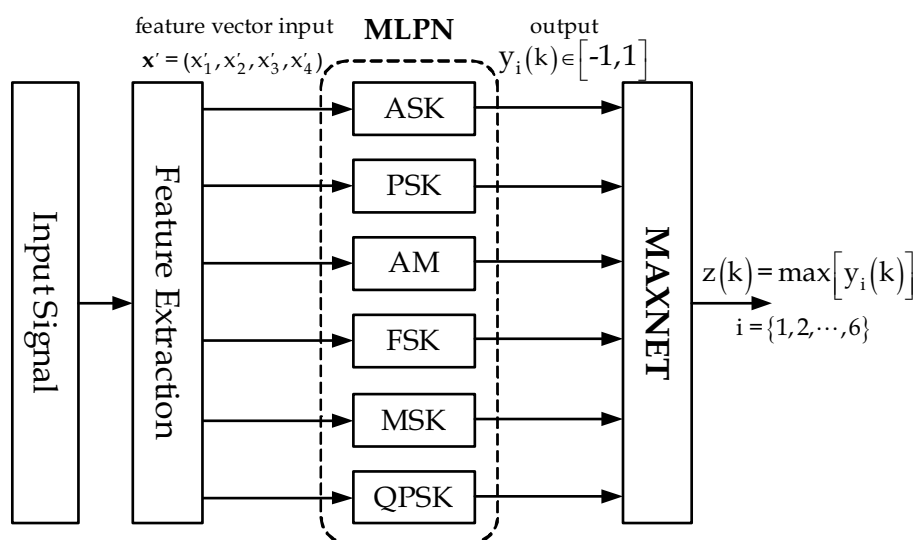


Fig. 9. Architecture of the BDT classifier based on spectral correlation analysis.

#### 4.2.2 Multilayer Linear Perceptron Network

Artificial Neural networks have long been considered for pattern recognition and modulation classification and have proven to be robust to a variety of conditions such as interfering signals and noise. In order to compare the classifier performance of artificial neural network and support vector machine, a signal classification approach using spectral correlation and neural networks, which was proposed by A. Fehske (Fehske et al., 2005), is introduced below. Due to its simplicity, a multilayer linear perceptron network (MLPN) with 4 neurons in the hidden layer was used for each signal class, and each input layer uses the normalized spectral correlation feature vector  $\mathbf{x}' = (x'_1, x'_2, x'_3, x'_4)$  as input. Each MLPN was trained with a back propagation algorithm (Gupta, 2003) with an initial learning rate  $\eta = 0.05$  decreasing with each epoch, a momentum constant  $\alpha = 0.7$ , and an activation function  $\tanh(x)$ . The output of each MLPN is a continuous value in the range  $(-1, 1)$ . The MAXNET structure shown in Fig. 10 simply chooses the signal whose MLPN outputs the largest value. A typical gradient descent algorithm can be used to solve the linearly non-separable signal

classification problem, which can achieve minimal mean square error of expected output and actual output. The training results of all the MLPN are inputted into a simple MAXNET for final decision. The decision function of the MAXNET is defined by

$$z(k) = \operatorname{argmax} [y_i(k)] \quad (39)$$

The signal classification approach using spectral correlation and MLPN is shown in Fig. 10.

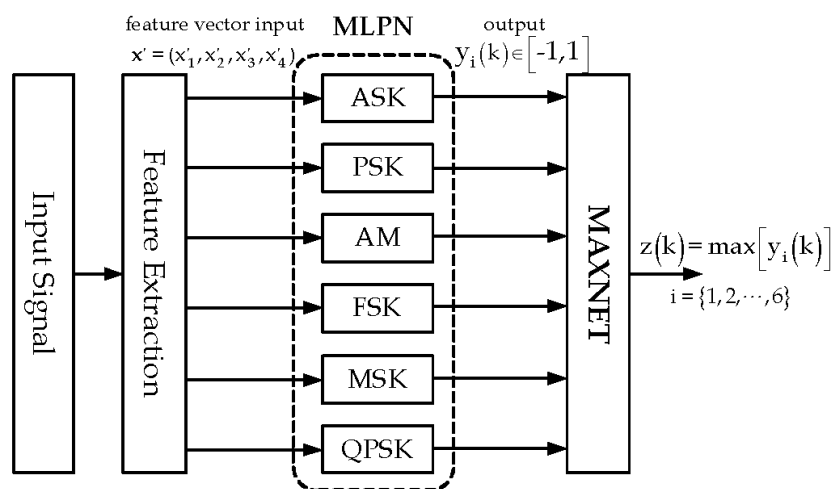


Fig. 10. Architecture of the MLPN classifier based on spectral correlation analysis.

## 5. Simulation Results

In this section, a variety of Monte Carlo simulations are presented to illustrate the performance of the algorithm. In the simulations, we define a signal set as {AM, ASK, FSK, PSK, MSK, QPSK}. For each type of signal,  $N$  signal samples constitute one frame which is used as an observation window to compare the performance of the algorithm with different data samples. To distinguish 6 modulation classes, simulation are carried out with 1100 frames at SNR ranging from 0 dB to 20 dB using three classifiers developed. 100 frames are used for training samples, and the remaining 1000 frames are used to calculate the probability of correct classification of different classifiers. The probability of correct classification ( $P_{cc}$ ) can be defined by

$$P_{cc} = \frac{1}{N} \sum_{i=1}^N P_r [H_i | H_i] \quad (40)$$

where  $N$  is the number of simulations,  $P_r [H_i | H_i]$  is the probability that the algorithm determine the signal class  $i$  correctly.

The radio channel models considered in the simulations include Gaussian channel and cognitive radio channel. In order to simulate the wireless environment of cognitive radios, WRAN channel model B recommended in IEEE 802.22 standard is used as cognitive radio channel (Sofer, E. & Chouinard, 2005). The WRAN channel reference model B determined

by the IEEE 802.22 standard group has a multi-path (6-path) delay profile, which is summarized in Table. 2.

In the IEEE 802.22 WRAN system, cognitive radio technology is considered to share the licensed spectrum of Digital TV, the typical service coverage is from 33 kilometres to 100 kilometres. The reference channel model B for the IEEE 802.22 WRAN are derived from a scenario that transmits the signals between the fixed BS and CPEs in wireless broadband environments. In such a dynamic channel environment, the delay extension is high with lower Doppler frequency. An Example channel responses for the nominal WRAN channel B is illustrated in Fig. 11.

Path	Delay( $\mu$ s)	Power(dB)	Doppler Freq.(Hz)
1	0	0	0
2	3	-7	0.10
3	8	-15	2.50
4	11	-22	0.13
5	13	-24	0.17
6	21	-19	0.37

Table 2. Multipath profiles of the IEEE 802.22 WRAN reference channel model B.

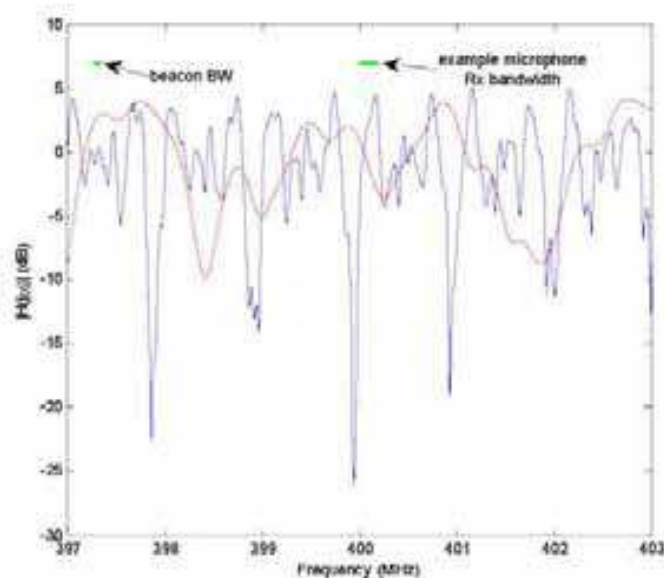


Fig. 11. An Example channel responses for the nominal WRAN channel B.

Table.3 and Table.4 indicate the probability of correct classification ( $P_{cc}$ ) for each modulation type with the training data length of 1000 over Gauss channel and cognitive radio channel, respectively. Results show the overall correct rate is above 92.83% for a SNR of 4dB, and 97.32% for a SNR of 8dB. These good results for signals with low SNR in the cognitive radio environment show the proposed approach is insensitive with SNR variation, which come from the effects of the robustness of SVM classifier. According to Table.3, the proposed approach has better performance in both channel conditions.



Signal Type	SNR/dB					
	0dB	4dB	8dB	12dB	16dB	20dB
AM	96.92	100	100	100	100	100
ASK	95.23	100	100	100	100	100
FSK	92.07	97.23	100	100	100	100
PSK	86.34	93.02	98.24	99.91	100	100
MSK	88.58	95.54	99.29	100	100	100
QPSK	85.14	92.83	97.32	99.72	99.93	100

Table 3. Pec of SVM classifier over Gaussian channel.

Signal Type	SNR/dB					
	0dB	4dB	8dB	12dB	16dB	20dB
AM	94.32	96.61	99.71	99.92	100	100
ASK	91.33	96.14	99.88	100	100	100
FSK	89.87	95.22	99.91	100	100	100
PSK	87.29	92.36	98.29	99.91	100	100
MSK	85.61	94.19	97.90	99.89	100	100
QPSK	82.20	90.13	96.19	98.43	99.92	100

Table 4.  $P_{cc}$  of SVM classifier over WRAN channel B.

Fig. 12 and Fig. 13 show the performance of SVM classifier with data length as parameter over different channel models. When the data length is 100 and for a SNR of 4dB, the  $P_{cc}$  is up to 80.62% and with data length 200 and for a SNR of 6dB, the  $P_{cc}$  increases to 90%. When the data length is 1000 and for a SNR of 10dB, the  $P_{cc}$  is close to 100%. Above results show that the performance of the SVM classifier is high for small training data in both channel models. Fig. 14 and Fig. 15 are the performance comparison between BDT classifier, MLPN classifier and SVM classifier with the spectral correlation features over Gaussian channel and WRAN channel, which are calculated for different signals using data length of 200 and 1000, respectively. It is shown that when the SNR is lower, the MLPN classifier shows poor performance. While the SNR is higher, the probability of correct classification is increased. In lower SNR, the variation of the spectral correlation features (SCFs) is drastically due to the effect of the noise. Thus, the construction of neural network is not complete with small training data, which results in the performance degradation. The decision tree based classifier only use the partial information of the spectral correlation features (SCFs), therefore, the correct probability is lower than SVM classifier in the whole SNR range. All the results show the high performance of SVM classifier based on spectral correlation features (SCFs).

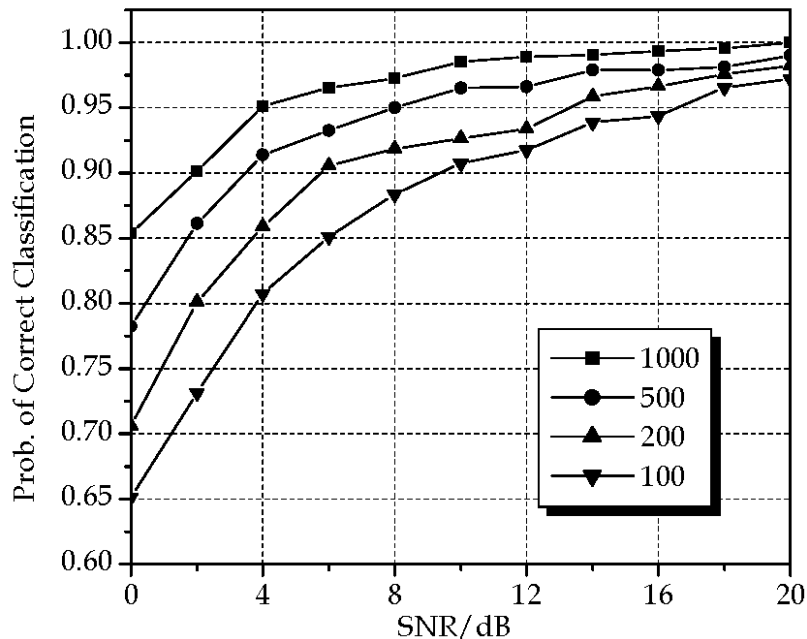


Fig. 12. The performance of SVM classifier with different data length over Gaussian channel.

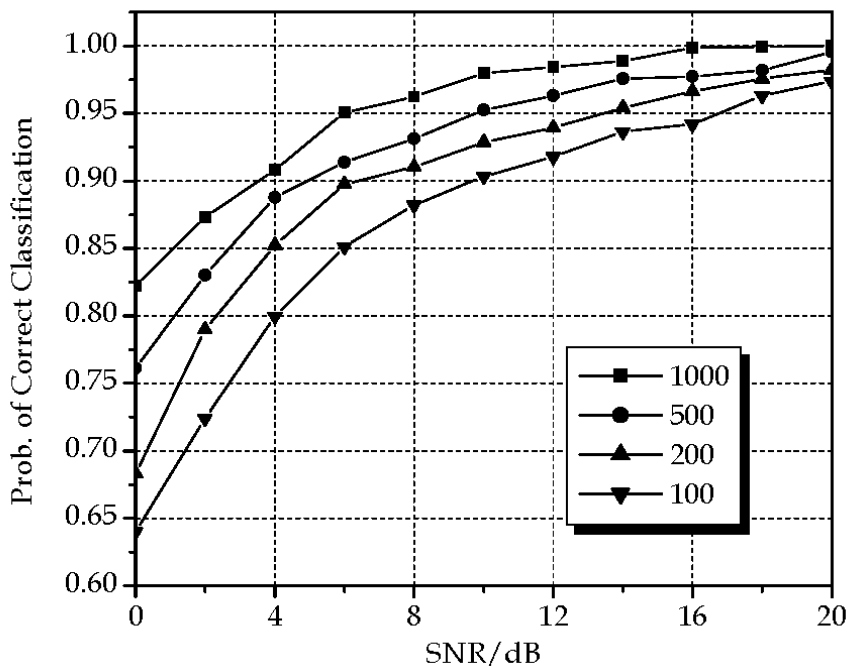


Fig. 13. The performance of SVM classifier with different data length over WRAN channel.

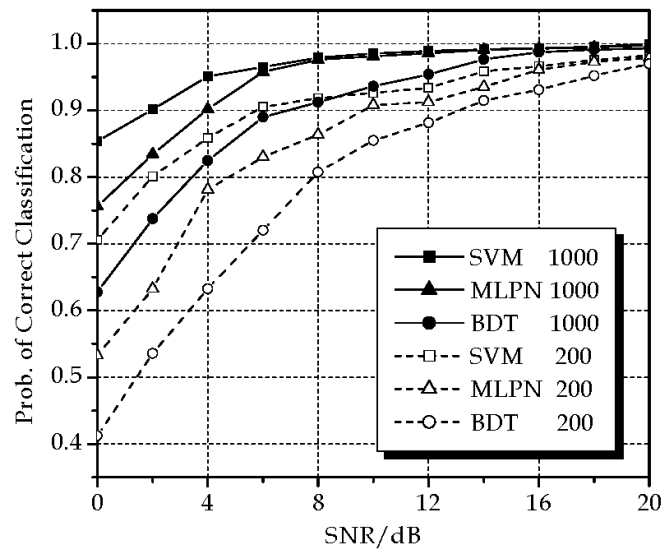


Fig. 14. Comparison between three classifiers with different data length over Gaussian channel.

## 6. Conclusion and Future Work

In this chapter, we proposed a novel approach combining the spectral correlation features and SVM for signal classification in cognitive radio environment. Four spectral correlation characteristic parameters were chosen as feature vector of SVM classifier. Simulation results show that the overall success rate is above 92.83% with data length of 1000 when SNR is equal to 4dB. Compared to existing methods, the proposed approach is more effective in the case of low SNR and limited training numbers. Future work in the area of signal classification for cognitive radio systems will involve the analysis of higher order spectrum correlation features of more communication signals. Based on these features, a multi-class SVM classifier can be used to improve the accuracy of classification and reduces the computational complexity. In addition, the classifier performance will be tested via simulations using several different channel models.

## 7. Acknowledgements

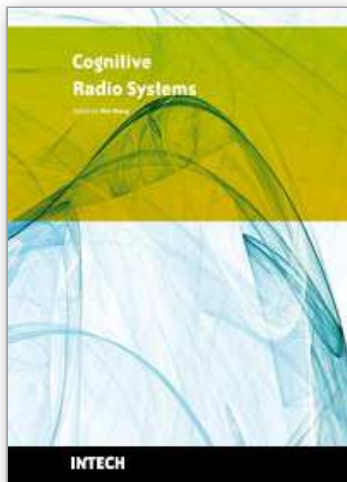
This research work is supported by Chinese Top University PhD Studying Abroad Program from China Scholarship Council. The authors wish to thank Yujing Wang for her helpful suggestions and support.

## 8. References

- Mitola, J. & Maguire, G. Q. (1999). Cognitive radio: making software radios more personal, *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13-18
- Mitola, J. (2001). Cognitive Radio - An Integrated Agent Architecture for Software Defined Radio, *PhD Thesis*, Royal Institute of Technology (KTH), Kista, Sweden
- FCC. (2003). *Establishment of Interference Temperature Metric to Quantify and Manage Interference and to Expand Available Unlicensed Operation in Certain Fixed Mobile and*

- Satellite Frequency Bands*, ET Docket 03-289, Notice of Inquiry and Proposed Rulemaking.
- Haykin, S. (2005). Cognitive Radio: Brain-Empowered Wireless Communications, *IEEE Journal on Selected Area in Communications*, vol.23, no 2, pp. 201-220
- Akyildiz, I. F.; Lee, W. Y.; Vuran, M. C. & Mohanty, S. (2006). NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey, *Computer Networks Journal(Elsevier)*, vol. 50, no. 13, pp. 2127-2159
- Ghasemi, A. & Sousa, E. S. (2005). Collaborative spectrum sensing for opportunistic access in fading environment, *Proc. of IEEE DySPAN 2005*, pp. 131-136
- Sahai, A.; Hoven, N. & Tandra, R. (2004). Some fundamental limits in cognitive radio, *Proc. of Allerton Conf. on Commun., Control and Computing*, pp. 121-127
- Digham, F. F.; Alouini, M. S. & Simon, M. K. (2003). On the energy detection of unknown signals over fading channels, *Proc. of IEEE ICC, Anchorage, USA*, pp. 3575-3579
- Tang, H. (2005). Some physical layer issues of wide-band cognitive radio system, *Proc. of IEEE DySPAN 2005*, pp. 151-159
- Cabric, D.; Mishra, S. M. & Brodersen, R. W. (2004). Implementation issues in spectrum sensing for cognitive radios, *Proc. of the 38th. Asilomar Conference on Signals, Systems, and Computers*, pp. 772 - 776
- Fehske, A.; Gaeddert, J. D. & Reed, J. H. (2005). A new approach to signal classification using spectral correlation and neural networks, *Proc. of IEEE DySPAN 2005*, pp. 144-150
- Bennett, W.R. (1958). *Statistics of Regenerative Digital Transmission*, Bell Syst. Tech. J, 37: 1501-1542
- Gardner, W. & Franks, L. (1975). Characterization of Cyclostationary Random Signal Processes, *IEEE Trans on Information Theory*, vol. 21, pp. 4-14
- Polydoros, A. & Kim, K. (1990). On the Detection and Classification of Quadrature Digital Modulations in Broadband Noise, *IEEE Trans. Commun*, vol. 38, pp. 1199-1211
- Sapiano, P.C. & Martin, J. D. (1996). Maximum Likelihood PSK Classifier, *Proc. of IEEE Military Communications Conference*, vol. 3, pp.1010-1014
- Sills, J. A. (1999). Maximum-likelihood Modulation Classification for PSK/QAM, *Proc. of IEEE Military Communications Conference*, vol. 1, pp. 217-220
- Wei, W. & Mendel, J. M. (2000). Maximum-likelihood Classification for Digital Amplitude Phase Modulations, *IEEE Transactions on Communications*, vol. 48, no. 2, pp. 189-193
- Hang, G. D.; Cai, B. & Wu, J. X. (2001). Spectral Correlation Method for Modulation Recognition, *Journal of Systems Engineering and Electronics*, vol. 23, no.3, pp. 34-37
- Nandi, A. K. & Azzouz, E. E. (1995). Automatic Analogue Modulation Recognition, *Signal Processing*, vol. 46, pp. 211-222
- Azzouz, E.E. & Nandi, A. K. (1995). Automatic Identification of Digital Modulation Types, *Signal Processing*, vol. 47, pp. 55-69
- Azzouz, E.E. & Nandi, A. K. (1996). Procedure for Automatic Recognition of Analogue and Digital Modulations, *IEEE Proceedings on Communications*, vol. 143, no. 5, pp. 259-266
- Azzouz, E.E. & Nandi, A. K. (1996). *Automatic Modulation Recognition*, Elsevier Science Ltd, pp. 241-273
- Nandi, A. K. & Azzouz, E. E. (1998). Algorithms for Automatic Modulation Recognition of Communication Signals, *IEEE Trans. on Communications*, vol. 46, no. 4, pp. 431-436

- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*, Springer-Verlag, New York
- Han, G. (2003). Research on Blind Signal Detection Techniques in Adaptive Single- and Multi-carrier Modulation, *PhD Thesis*, Xidian University, Xi'an, China
- Gardner, W.A. (1987). Spectral Correlation of Modulated Signals: PART I – Analog Modulation, *IEEE Trans on Commun.*, vol. 35, no. 6, pp. 584-594
- Gardner, W.A. (1987). Spectral Correlation of Modulated Signals: PART II – Digital Modulation, *IEEE Trans on Commun.*, vol. 35, no. 6, pp. 595-601
- Gardner, W. & Franks, L. (1975). Characterization of Cyclostationary Random Signal Processes, *IEEE Trans. on Information Theory*, vol. 21, pp.4-14
- Gao, Y. (2007). Modulation recognition and high dynamic synchronization based on cyclic spectral density, *PhD Thesis*, Harbin Institute of Technology, Mar 2007.
- Gardner, W. A. & Spooner, C. M. (1988). Cyclic Spectral Analysis for Signal Detection and Modulation Recognition, *Proceedings of the Conference on Military Communications (MILCOM)*, San Diego, CA, pp. 419-423
- Sofer, E. & Chouinard, G. (2005). *WRAN Channel Modeling*, IEEE802.22-05/0055r7
- Wang, Y. (2007). Research on Identification of Modulation Signals, *Master Thesis*, Xidian University, Xi'an
- Gurman, D. B. Hwang, E. S. & Vizcaino, J. M. (2008). *Modulation Identification Using Classic Pattern Recognition Techniques*, [www.andrew.cmu.edu/user/danielgu/Pat\\_Rec\\_Final\\_Paper\\_Gurman\\_Vizcaino\\_Hwang.pdf](http://www.andrew.cmu.edu/user/danielgu/Pat_Rec_Final_Paper_Gurman_Vizcaino_Hwang.pdf)
- Boser, B.; Guyon, I. & Vapnik, V. (1992). A training algorithm for optimal margin classifiers, *Proc. of Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh: ACM Press
- Cortes, C. & Vapnik, V. (1995). Support vector networks. *Machine Learning*, vol. 20, pp. 273 - 297
- Scholkopf, B. & Burges, C. & Vapnik, V. (1995). Extracting support data for a given task, *Proc. of First Intl. Conf. on Knowledge Discovery & Data Mining*, AAA I Press, pp. 262-267
- Burges, J. C. (1997). *A Tutorial on Support Vector Machines for Pattern Recognition*. Bell Laboratories, Lucent Technologies
- Devroye, L.; Gyrofi, L. & Lugosi, L. (1996). *A Probabistic Theory of Pattern Recognition*, Springer, New York:
- Daniel, J. S. & James, A. B. (2000). Support vector machine techniques for nonlinear equalization, *IEEE Trans. on Signal Processing*, vol. 11, no. 48, pp. 3271-3227
- Cristianini, N.; Shawe-Taylor, J.; Elisseeff, A. & Kandola, J. (2002). On Kernel Target Alignment, *Proc. of Neural Information Processing Systems*, Cambridge, MA: MIT Press, pp. 367-373
- Sherrod, P. H. (2008). *DTREG Predictive Modeling Software Manual*, <http://www.dtrege.com/DTREG.pdf>
- Chapelle, O.; Vapnik, V. N. & Bacsquest, O. et al. (2002). Choosing multiple parameters for support vector machines, *Machine Learning*, vol. 46, pp. 131-159
- Lanckriet, G.; Cristianini, N.; Bartlett, P. L. & Jordan, M. (2002). *Learning the Kernel Matrix with Semi-Definite Programming*, Technical Report No. UCB/CSD-02-1206, University of California, Berkeley
- Gupta, M. M. (2003). *Static and Dynamic Neural Networks: from Fundamentals to Advanced Theory*, Wiley, New York



## **Cognitive Radio Systems**

Edited by Wei Wang

ISBN 978-953-307-021-6

Hard cover, 340 pages

**Publisher** InTech

**Published online** 01, November, 2009

**Published in print edition** November, 2009

Cognitive radio is a hot research area for future wireless communications in the recent years. In order to increase the spectrum utilization, cognitive radio makes it possible for unlicensed users to access the spectrum unoccupied by licensed users. Cognitive radio let the equipments more intelligent to communicate with each other in a spectrum-aware manner and provide a new approach for the co-existence of multiple wireless systems. The goal of this book is to provide highlights of the current research topics in the field of cognitive radio systems. The book consists of 17 chapters, addressing various problems in cognitive radio systems.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hao Hu (2009). Cyclostationary Approach to Signal Detection and Classification in Cognitive Radio Systems, Cognitive Radio Systems, Wei Wang (Ed.), ISBN: 978-953-307-021-6, InTech, Available from: <http://www.intechopen.com/books/cognitive-radio-systems/cyclostationary-approach-to-signal-detection-and-classification-in-cognitiveradio-systems>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen