International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.3 (2016), pp.133-142
http://dx.doi.org/10.14257/ijsip.2016.9.3.12

A Novel Signal Identification Method via Improved Random Forest in Cognitive Network

Xin WANG[1,2,*], Jin-kuan WANG[1], Zhi-gang LIU[1], Xin Song[1], Xi HU[1]

[1]College of Information Science and Engineering, Northeastern University, Shenyang 110004, China
[2]School of Information & Control Engineering , Shenyang Jianzhu University, Liaoning, Shenyang, 110168, China;
wangxin7988@163.com

Abstract

To increase the identification performance of primary user signal modulation types in cognitive network under low signal-to-noise ratio, a novel approach based on improved random forest (AL-RF) is proposed to identify the various modulation types of primary user signals. First and foremost, a set of cyclic spectrum features of the received radio signal are calculated via cyclic spectral correlation analysis. Then, the dynamic sample selection strategy is applied to construct the sample sets dynamically and to train the random forest (RF) classifier repeatedly. Eventually, the trained RF classifier is utilized to identify the modulation type of signals. The experimental results show that the proposed algorithm can effectively solve the problem of the low accuracy on modulation type recognition of the primary users especially in the case of low SNR environment.

Keywords*: Cognitive network, spectrum sensing, cyclic spectral correlation analysis; dynamic sample selection strategy; improved random forest*

1. Introduction

As one of the most challenging functions in cognitive radio system in cognitive networks, spectrum sensing can detect the availability of radio frequency bands for possible use by secondary user without interference to primary user[1,2]. The key technology can detect situation of spectrum utilization quickly, accurately, and effectively. Some traditional techniques proposed for spectrum sensing are energy detection, matched filter detection, cyclostationary feature detection[3,4]. In[5],the authors proposed spectrum sensing algorithm based on the maximum-minimum eigenvalue (MME). Authors in [6] took the maximum value along the spectral domain as a simplified cyclostationarity feature set for modulation classification, which was called Cyclic Domain Profile (CDP). Artificial intelligence (AI) algorithms, such as artificial neural networks (ANN) and some other machine learning methods, have been widely employed and promising recognition results have been reported [7]. In [8], an approach of signal classification for cognitive radios combining the spectral correlation analysis and support vector machine (SVM) is proposed. These approaches can embody good spectrum sensing and modulation type recognition of the primary users properties in the case of high signal-to-noise ratio. However, in low signal-to-noise ratio environments, various modulation signals are not easy to be found by utilizing the approaches.

Random forest is coined by Leo Breiman in 2001 on the foundation of machine learning theory[9,10]. It is a popular classification method which is an ensemble of a set of classification trees. One of the most popular forest construction procedures is to randomly select a subspace of features at each node to grow branches of a decision trees, then to use bagging method to generate training data subsets for building individual trees,

ISSN: 2005-4254 IJSIP

finally to combine all individual trees to form random forests model. Due to its algorithmic simplicity and prominent classification performance, random forest has become a promising method to solve many classification problems. In traditional random forest, a bagging method is used to generate training data subsets for building individual trees, finally to combine all individual trees to form random forests model. However, the degree of difference between some selected training samples via these methods is not significant, so the classification performance of generated random forest is restricted in a certain extent.

To solve the above problems, an approach to signal recognition based on improving random forests (AL-RF) is proposed to enhance the detection and modulation type recognition of the weak primary users in low signal-to-noise ratio. A novel dynamic sample selection strategy is proposed to select the training samples, and an unbiased splitting manner based on conditional probability is adopted to the decision of the trees which form random forest. The contributions of this paper are as follows:

1. A set of characteristic parameters (e.g., $E_0(S)$, $D_0(S)$, ε_0, N_α^0, ρ_M^0) of the received signal spectral correlation function (SCF) are calculated via cyclic spectral correlation analysis.

2. The training samples that are helpful to improve classification performance can be obtained through repetitious iterative sampling by utilizing the dynamic sample selection strategy.

3. Each tree that belongs to AL-RF classifier is trained by the training samples that are obtained via the above mentioned iterative procedure. Then, the testing samples are classified and identified by the trained AL-RF classifier.

The remainder of the paper is organized as follows. In Section 2, system model is described. The proposed algorithm is investigated in Section 3 and is well validated with computer simulation in Section 4. The concluding remarks are made in Section 5.

2. Problem Description

Given a cognitive radio network with one primary user and *W* secondary users, for any one of the secondary user, the presence of the primary user can be summarized as a hypothesis test model of two elements

$$\begin{cases} H_0: & y(t) = n(t) \\ H_1: & y(t) = s(t) + n(t) \end{cases} \tag{1}$$

where H_0 denotes that the primary user is not exist, H_1 denotes that the primary user exists, $0 \le t \le T$, T represents the sampling time of the received signal. A process $s(t)$ (the primary user signal) is said to be cyclostationary in wide sense if its mean and variance are periodic with a period T; $n(t)$ represents the additive white Gauss noise, the mean is zero, variance is σ_n^2 。

On the basis of this model, We have extracted the cyclic spectrum characteristic parameters of the received signal ($a_1^i, a_2^i, \cdots, a_N^i$), and obtain characteristic vectors $x_i = \left(a_1^i, a_2^i, \cdots, a_N^i\right)^T$, $i = 1, 2, \cdots, N$ as AL-RF classifier training samples, thus we generate improved random forest(AL-RF) classifier and the identification model based on AL-RF. As we see from Fig.1.

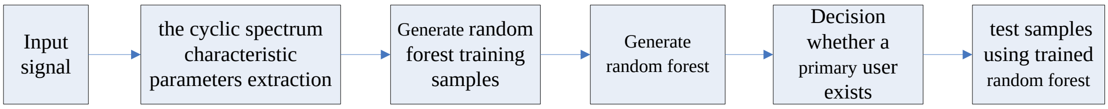

Figure 1. The Identification Model Diagram of the Modulation Types of Primary user Signals based on AL-RF

3. Proposed Algorithm

3.1 Cyclic Spectrum Characteristic Parameters Extraction

Given each secondary user receiving signal is $y(t)$, $R_y^{\alpha}(\tau)$ expresses the autocorrelation function of $y(t)$. The spectral correlation function (SCF) can be obtained from the Fourier transform of the cyclic autocorrelation $R_y^{\alpha}(\tau)$.

$$S_y^{\alpha}(f) = \int_{-\infty}^{\infty} R^{\alpha}(\tau) e^{-j2\pi f t} d\tau \tag{2}$$

We use the spectrally smoothed cyclic periodogram method to calculate the estimated the spectral correlation function (SCF) in Eq.2.

$$S_y^{\alpha}(k) = \frac{1}{TL} \sum_{l=-(L-1)/2}^{(L-1)/2} Y(k+l+\frac{\alpha}{2}) Y^*(k+l-\frac{\alpha}{2}) \quad , \quad k = 1, 2, \cdots, K \tag{3}$$

where $Y(k)$ is the discrete Fourier transform of $y(t)$, L is sample length that participate in the frequency domain smoothing , $Y^*(k)$ is the conjugate of $Y(k)$.

$$S(k) = S_y^{\alpha}(k)\big|_{\alpha = 1/T_0} \tag{4}$$

For the received signal with more than one circular frequency, it obtains the largest energy spectrum as $S(k)$.

According to above analysis, the key features of SCF which are sensitive with modulation types and insensitive with SNR variation can be extracted.

1. The mathematical expectation and variance of cyclic spectrum $S(k)$

The mathematical expectation $E(S)$ and variance $D(S)$ of cyclic spectrum $S(k)$ is calculated under the presence of primary user(H_1) and the absence of primary user(H_0) respectively. The mathematical expectation $E(S)$ and variance $D(S)$ of cyclic spectrum $S(k)$ are deduced in Appendix specifically.

2. The spectrum energy is expressed in Eq.5.

$$\varepsilon = \frac{1}{K} \sum_{k=0}^{K-1} |S(k)|^2 \tag{5}$$

3. Number of cyclic spectral line on α-domain of SCF: N_{α}

Let $f = 0$ in Eq.3, N_{α} can be obtained from the ichnography of $S_y^{\alpha}(0)$.

4. Maximum value of spectral coherence coefficient (SCC): ρ_M

The correlation coefficient for the SCF can be calculated as

$$\rho_y^\alpha(f)=\frac{S_y^\alpha(f)}{\sqrt{S_y^0(f+\alpha/2)S_y^0(f-\alpha/2)}}$$

(6)

3.2 Improving RF Strategy (AL-RF)

3.2.1 Random Forest.

Random forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the random forest. It denotes that the error of the random forest algorithm is more stable, and the algorithm can overcome the disadvantages of a single decision tree. Given a dataset containing N samples for training, $(\vec{X},\vec{Y})=\{(\vec{x}_1,y_1),\cdots,(\vec{x}_N,y_N)\}$, where $\vec{x}_i$ is the feature vector of m dimensions and y_i is the class label which value is between 1 and T. The building procedure of the random forest can be stated as follows:

Step1. For a given training dataset G, extract a new sample set by W time's repeated random sampling using bootstrap method.

Step2. The remaining samples are used to calculate the out-of-bag error (OOB-error). The proportion specifies the splitting.

Step3. Select the feature in m_1 dimensions($m_1 \prec m$) that the best split as the each node's classification properties to train the tree.

Step4. Completely grow the tree to the largest possible extension without pruning. The decision tree stops growing when the desired proportions of classification purity of each node has been achieved, or a given layer has been reached.

Step5. Repeat step1 to step4, a completed random forest consists of several classification trees has been built.

A classification result is assigned to each leaf node and the final decision is determined by taking the class having the most votes.

3.2.2 The procedure of Improved RF algorithm (AL-RF)

The proposed algorithm consists of the dynamic sample selection strategy and improved random forest classifier. Several samples are selected to train the AL-RF classifier through repetitious iteration test in the dynamic sample selection strategy. An unbiased splitting manner based on conditional probability is adopted while the AL-RF classifier is designed, and can be given by

$$MR(A_t)=\frac{(\sum_{j=1}^{m_t}p(v_{t,j})^2\sum_{t=1}^{h}p(y_t\left|v_{t,j}\right)^2)}{(\sum_{t=1}^{h}p(v_{t,j})^2(1-\sum_{t=1}^{h}p(y_t)^2))}-\frac{(\sum_{j=1}^{m_t}p(v_{t,j})^2\sum_{t=1}^{h}p(y_t)^2)}{(\sum_{t=1}^{h}p(v_{t,j})^2(1-\sum_{t=1}^{h}p(y_t)^2))}$$

(7)

Where m_t is the number of instances with the *j-th* value of the given attribute, h is the number of training samples classifications, $p(v_{t,j})$ is the probability of the attribute $A_t=v_i$, and $p(y_j\left|v_{t,j}\right.)$ denote the probability of the classmark $y_j=v_j$ under the attribute A_t.

According to the system model and the above procedure of cyclic spectrum characteristic parameters extraction, we assume that $E_0(S)$, $D_0(S)$, ε_0, N_α^0, ρ_M^0 is eigenvalue under the condition of H_0, the corresponding characteristic vector expresses as $x_0 = (E_0(S), D_0(S), \varepsilon_0, N_\alpha^0, \rho_M^0)^T$; $E_i(S)$, $D_i(S)$, ε_i, N_α^i, ρ_M^i is eigenvalue under the condition of H_1, the corresponding characteristic vector expresses as $x_i = (E_i(S), D_i(S), \varepsilon_i, N_\alpha^i, \rho_M^i)^T$, $i = 1,2,\cdots,I$, I is the number of the primary users signal modulation type.

The proposed algorithm is described as follows:

Step1. Calculate the characteristic parameters of the received signal spectral correlation function (SCF) $S(k)$ under the condition of H_1 and H_0, then form characteristic vectors and the corresponding samples.

Step2. Build a sampleset U consists of these samples, then select randomly a few samples to from sampleset U as the training samples to initialize the random forest.

Step3. Initialize the random forest with these selected samples through the building procedure of random forest.

Step4. Select m samples from remaining samples in sampleset U, then these samples are classified by the initialized random forest.

Step5. According to the classification outcome, the correct samples which are classified are abandoned, the incorrect samples are replaced in a training set U. All of the incorrect samples will participate in the next train of random forest.

Step6. Repeat step2 to step5, the process will not be stopped until the end of the random forest training standards is reached.

Step7. Generate the testing samples which are unmarked utilizing the characteristic parameters of the received signal spectral correlation function (SCF) $S(k)$ and build a sampleset G. Then the samples in sampleset G are selected randomly as the testing samples to complete detection and the recognition for the primary users signal modulation type.

4. Simulation Results

In this section the performance of the proposed approach is measured by conducting some experiments. We compare with SVM, RF algorithms and the proposed AL-RF algorithm at SNR ranging from -15 dB to 5 dB circumstance respectively in Matlab7.0. By using three modulation signal sets of BPSK, 2FSK and OFDM as input, the resolution of the carrier frequency is 1 MHZ, and input signals appear to uniform distribution between 3.1~4.8 GHz. The SVM, RF and AL-RF are compared under the signal-to-noise ratio (dB) in different circumstances, such as $-15,-10,-5,0$. We randomly select 600 samples from each set as the training samples respectively to establish the initial random forest classifer, and 2000 samples are selected from set U every time to be tested by the initial random forest classifer. Then the incorrect samples by classification are replaced in a training set U and be extracted for the next train of the random forest classifer. The DFT length of cyclic spectrum is 512, and the random forest is established by 100 trees. The channel is assumed to have a multi-path (3-path) delay profile with 150ns delay spread to transmit the three modulation signals. Simulations are carried out 10^4 orders of magnitude times.

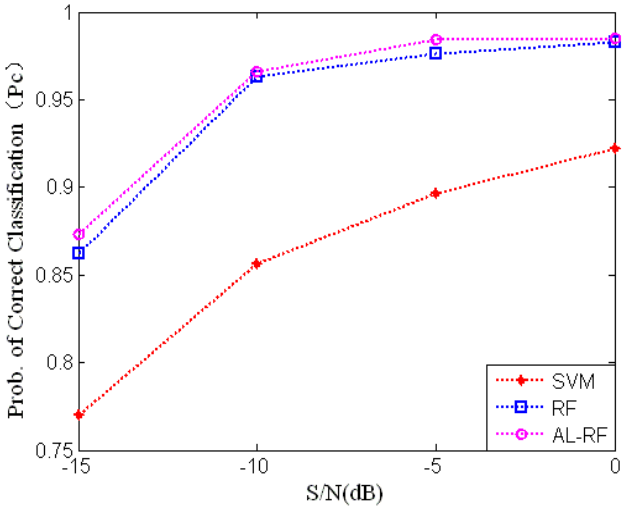

Figure 2. The Recognition Rate of the Proposed Algorithm versus SVM and RF Algorithms for BPSK

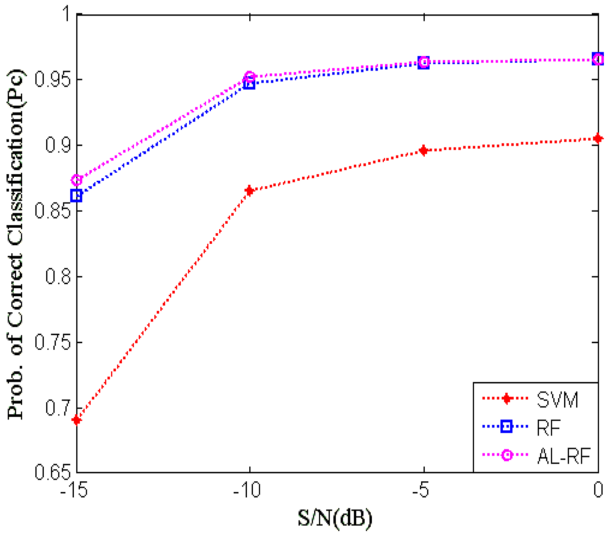

Figure 3. The Recognition Rate of the Proposed Algorithm versus SVM and RF Algorithms for 2FSK

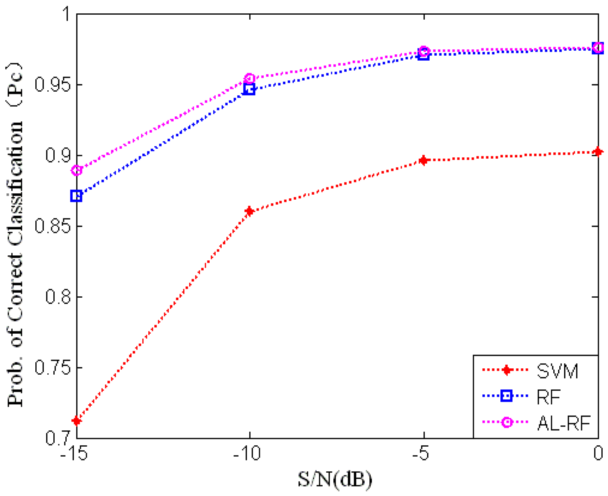

Figure 4. The Recognition Rate of the Proposed Algorithm versus SVM and

The comparison of recognition rate between simulated and theoretical recognition rate of the SVM and RF algorithms and AL-RF are shown in Fig.2, Fig.3 and Fig.4. In Fig.2, it is observed that the recognition rate of various algorithms increases with increases of signal-to-noise ratio in the low SNR environment. The recognition rate of RF increases from 0.874 to 0.965, and SVM algorithm increases from 0.690 to 0.905 when SNR vary between -15 dB to 0dB with interval of 5dB. Under the same varying SNR case, the recognition rate of AL-RF increases from 0.874 to 0.966, higher than RF and SVM algorithms obviously. In Fig.3 and Fig.4, the same comparison between the above algorithms for 2FSK and OFDM are repeated under the same SNR case, the proposed AL-RF algorithm still outperforms the other two algorithms.

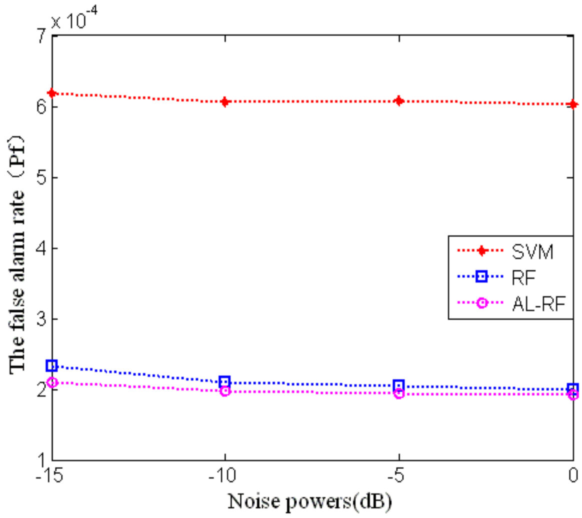

Figure 5. The Overall Average False Alarm Rate of the Proposed Algorithm versus SVM and RF Algorithms for all Modulation Signals

The overall average false alarm rate of the AL-RF algorithm compare with SVM, RF for RF algorithms obviously. It can be observed that our proposed method (AL-RF) different modulated signals are shown in Fig.5. Under different noise power, it is observed that the false alarm rate range of SVM, RF and the proposed AL-RF algorithm are 10^{-4} orders of magnitude respectively. The false alarm rate of the three algorithms increases with the decreases of noise power, but the proposed AL-RF algorithm is lower than SVM and can reduce the errors that are caused by sample selection mechanism of traditional random forest classifier, and the signal identification performance of the proposed is superior to SVM and RF algorithms in the range of lower signal-to-noise ratio.

5. Conclusions

In this paper, we propose an innovative signal identification method based on improved random forest in low SNR. A set of cyclic spectrum features of the received radio signal are calculated via cyclic spectral correlation analysis. The dynamic sample selection strategy is proposed to select the training samples for building random forest. An unbiased splitting manner based on conditional probability is applied to the decision of the trees in random forest. The experimental results show that the proposed algorithm can effectively enhance the recognition performance on signal types in the low SNR environment.

Future works in the area of signal identification for spectrum sensing systems will involve the improvement of detection and recognition performance: (1) In order to reduce the classification errors, the effective selected strategy on trees of random forest will be explored; (2) The signal identification method will be developed to improve the accuracy of recognition in the case of cooperative spectrum sensing.

Acknowledgment

Project supported by the National Natural Science Foundation of China under Grant (No.61374097; No.61473066), Hebei Province Natural science fund under Grant (No.F2011501021; No. F2014501082) and the Science and Technology Research Project of Higher Education of Hebei Province under Grant (QN20132010) respectively.

Appendix A

In the absence of the primary user, that is H_0, mathematical expectation of cyclic spectrum is expressed as follow:

$$E_0(S)=\frac{1}{TL}\sum_{l=-(L-1)/2}^{(L-1)/2}E[N(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2})] \tag{A1}$$

Where $N(k)$ is the discrete Fourier transform of Gauss white noise $n(t)$. Because of the cyclic frequency $a\neq 0$, we first prove that Equation (A2) has real solution under H_0.

$$E_0(S)=0 \tag{A2}$$

The variance of cyclic spectrum $S(k)$ is

$$D_0(S)=\frac{1}{T^2L^2}\sum_{l=-(L-1)/2}^{(L-1)/2}D[N(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2})]=\frac{K^2\sigma_n^4}{T^2L} \tag{A3}$$

According to the autocorrelation of Gaussian white noise, for any two values of *l*, that is $\forall l\neq l'$, there is the formula as follow

$$E[N(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2})N(k+l'+\frac{\alpha}{2})N^*(k+l'-\frac{\alpha}{2})]=0 \tag{A4}$$

The mathematical expectation of cyclic spectrum $S(k)$ under H_1 is given by

$$\begin{aligned}E_i(S)&=\frac{1}{TL}\sum_{l=-(L-1)/2}^{(L-1)/2}E\left\{[S(k+l+\frac{\alpha}{2})+N(k+l+\frac{\alpha}{2})]\cdot[S^*(k+l-\frac{\alpha}{2})+N^*(k+l-\frac{\alpha}{2})]\right\}\\&=\frac{1}{TL}\sum_{l=-(L-1)/2}^{(L-1)/2}[S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2})]\end{aligned} \tag{A5}$$

The variance of cyclic spectrum $S(k)$ is deduced from Equation(A6).

$$\begin{aligned}D_i(S)&=\frac{1}{T^2L^2}E\left\{[\sum_{l=-(L-1)/2}^{(L-1)/2}A_1(l)]^2+2[\sum_{l_1=-(L-1)/2}^{(L-1)/2}A_1(l_1)\sum_{l_2=-(L-1)/2}^{(L-1)/2}A_2(l_2)]\right\}-\frac{1}{T^2L^2}[E_i(S)]^2+\frac{K^2\sigma_n^4}{T^2L}\\&=\frac{1}{T^2L^2}D[\sum_{l=-(L-1)/2}^{(L-1)/2}S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2})]+\frac{1}{T^2L^2}E\{2[\sum_{l_1\neq -(L-1)/2}^{(L-1)/2}A_1(l_1)\sum_{l_1\neq l_2 -(L-1)/2}^{(L-1)/2}A_1(l_2)]+\sum_{l_1=l_2-(L-1)/2}^{(L-1)/2}[A_1(l_1)]^2\\&\quad+2[\sum_{l_1=-(L-1)/2}^{(L-1)/2}A_1(l_1)\sum_{l_2=-(L-1)/2}^{(L-1)/2}A_2(l_2)]\}+\frac{K^2\sigma_n^4}{T^2L}\end{aligned} \tag{A6}$$

$$A_1(l)=S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2})+S(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2})+S^*(k+l-\frac{\alpha}{2})N(k+l+\frac{\alpha}{2})$$

(A7)

$$A_2(l)=N(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2}) \tag{A8}$$

$$A_{11}(l)=S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2}) \quad \text{(A9)}$$

$$A_{12}(l)=S(k+l+\frac{\alpha}{2})N^*(k+l-\frac{\alpha}{2})+S^*(k+l-\frac{\alpha}{2})N(k+l+\frac{\alpha}{2}) \quad \text{(A10)}$$

$$\sigma_0^2=\frac{K^2\sigma_n^4}{T^2L} \quad \text{(A11)}$$

$$\mu=E_i(S)=\frac{1}{TL}\sum_{l=-(L-1)/2}^{(L-1)/2}[S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2})] \quad \text{(A12)}$$

$$\sigma_s^2=\frac{1}{T^2L^2}D[\sum_{l=-(L-1)/2}^{(L-1)/2}S(k+l+\frac{\alpha}{2})S^*(k+l-\frac{\alpha}{2})] \quad \text{(A13)}$$

$$\sigma_{sn}^2=\frac{1}{T^2L^2}E\left\{2[\sum_{l_{11}=-(L-1)/2}^{(L-1)/2}A_{11}(l_{11})\sum_{l_{12}=-(L-1)/2}^{(L-1)/2}A_{12}(l_{12})]+\sum_{l_{12}=-(L-1)/2}^{(L-1)/2}[A_{12}(l_{12})]^2+2[\sum_{l_1=-(L-1)/2}^{(L-1)/2}A_1(l_1)\sum_{l_2=-(L-1)/2}^{(L-1)/2}A_2(l_2)]\right\} \quad \text{(A14)}$$

References

[1] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: A survey", IEEE Journal of Selected Topics in Signal Processing, vol. 5, no. 1, **(2011)**.

[2] H. Sun, A. Nallanathan and C. X. Wang, "Wideband spectrum sensing for cognitive radio networks: a survey", IEEE Journal of Wireless Communications, vol. 20, no. 2, **(2013)**.

[3] Y. C. Liang, K. C. Chen and G. Y. Li, "Cognitive Radio Networking and Communications: An Overview", IEEE Transactions on vehicular technology, vol. 60, no. 7, **(2011)**.

[4] W. Xin, W. Jinkuan and L. Zhigang, "Spectrum sensing algorithm based on random forest in cognitive network", Chinese Journal of Scientific Instrument, vol. 34, no. 11, **(2013)**.

[5] Y. Zeng and Y. Liang, "Maximum-minimum eigenvalue detection for cognitive radio", Proc of IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications(PIMRC 2007) , **(2007)**.

[6] P. D. Sutton, K. E. Nolan and L. E. Doyle, "Cyclostationary signatures in practical cognitive radio applications", IEEE Journal of Selected Areas in Communications, vol. 26, no. 1, **(2008)**.

[7] Y. J. Tang, Q. Y. Zhang and W. Lin, "Artificial neural network based spectrum sensing method for cognitive radio", Proc of the 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM '10), Chengdu, **(2010)**.

[8] H. Hu, J. Song and Y. Wang, "Signal classification based on spectral correlation analysis and SVM in cognitive radio", Proc of the 22st International Conference on Advanced Information Networking and Applications (AINA '08), **(2008)**.

[9] L. Breiman, "Random forests", Machine Learning, vol. 45, **(2001)**.

[10] A. Wang, G. Wan and Z. Cheng, "Incremental Learning Extremely Random Forest Classifier for Online Learning", Journal of Software, vol. 22, no. 9, **(2011)**.