IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 26, 2017, accepted August 13, 2017, date of publication August 29, 2017, date of current version October 12, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2746140

Robust Automatic Modulation Classification Under Varying Noise Conditions

ZHILU WU[1], SIYANG ZHOU[1], ZHENDONG YIN[1], (Member, IEEE), BO MA[1,2], AND ZHUTIAN YANG[1], (Member, IEEE)

[1]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China
[2]China Academy of Space Technology, Beijing 100081, China

Corresponding author: Zhendong Yin (yinzhendong@hit.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61571167, Grant 61471142, Grant 61102084, and Grant 61601145.

ABSTRACT Automatic modulation classification (AMC) plays a key role in non-cooperative communication systems. Feature-based (FB) methods have been widely studied in particular. Most existing FB methods are deployed at a fixed SNR level, and the pre-trained classifiers may no longer be effective when the SNR level changes. The classifiers may also need to be re-trained to be suitable for the varying channel environment. To address these problems, a robust AMC method under varying noise conditions is proposed in this paper. The method attempts to select noise-insensitive features from a large feature set to ensure that the trained classifiers will be robust to SNR variations. First, a feature set consisting of 25 types of features is extracted, and 4 features that are insensitive to noise are chosen through a feature selection method based on rough set theory. The generalizability of an SVM classifier trained on the 4 chosen features is evaluated based on numerical results. The classification accuracy remains reasonable when the SNR varies between 5 and 20 dB, indicating that the proposed method can be deployed under varying noise conditions.

INDEX TERMS Automatic modulation classification, feature extraction, feature selection, noise robustness.

I. INTRODUCTION

Automatic modulation classification (AMC) is a core technique in non-cooperative communication systems and has been widely studied in recent years. AMC has been shown to be of outstanding value in many applications, both civil and military. The purpose of AMC is to identify the type of modulation of a received signal. In general, AMC algorithms can be divided into two categories: likelihood-based (LB) and feature-based (FB) methods. LB methods theoretically yield optimal solutions, but they suffer from high complexity. They also require prior knowledge of the statistical information of the received signal, which is usually unavailable in practice. Meanwhile, although FB methods usually produce sub-optimal solutions, they are much easier to implement and do not depend on prior information. Moreover, a well-designed FB method can produce sub-optimal solutions that are very close to the optimal one. As a result, FB methods are most commonly investigated and applied. FB methods usually consist of 2 steps: First, features are extracted from the received signal, most of which represent statistical information related to either the original received signal or its transform. Then, suitable classifiers are trained to classify different modulation types. Thus, the two most important aspects that affect the performance of FB methods are feature extraction and the classifier.

Regarding feature extraction, Nandi and Azzouz use instantaneous features for the classification of both analog and digital signals in [1], which is the most representative work in the field of AMC. Instantaneous features are extracted from the instantaneous amplitude, instantaneous frequency and instantaneous phase of the signal, or a combination thereof. Cyclostationary features [2], [3] are based on the spectral correlation density (SCD) function of the received signal. Wavelet features [4], [5] can be obtained using the discrete wavelet transform, through which a multi-scale decomposition of the signal can be generated. High-order cumulant (HOC) features [6], [7] represent high-order statistical information of the received signal. Cyclostationary features and HOC features have been proven to be insensitive to noise, as additive Gaussian white noise (AWGN) can be highly suppressed in these features [8]. There are also other kinds of features, such as fractal features [9], which are occasionally adopted for radar signals.

However, the extraction of unnecessary features may lead to redundant information that is not useful for classification but increases the complexity of the training process. Redundant features may also cause the "curse of dimensionality" [10] to arise. Feature selection, which refers to choosing a subset of the original feature set by eliminating redundant features without affecting the classification performance, is applied to overcome the problem of high dimensionality. Current feature selection methods usually focus on choosing different subsets from the original feature set for different SNR levels. The most representative research has been performed by Wong and Nandi [11]. In this work, a genetic algorithm (GA) was applied, with an artificial neural network (ANN) as the classifier. In [12], Avci presented a hybrid algorithm that combines a support vector machine (SVM) approach with a GA.

Regarding classifiers, the most widely used linear classifiers in the early years were decision trees [13]. Linear classifiers are simple to implement; however, they cannot properly handle linearly inseparable features. Two of the most popular non-linear classifiers are ANN [14] and SVM [15] classifiers. An ANN classifier, because of its nature as a gradient descent method, can easily fall into local optima. Considering the advantages of SVM classifiers in terms of limited sample learning and generalization capabilities, such classifiers have become the most popular type applied to AMC problems in recent years.

In this paper, we focus on the problem that existing FB methods are trained and tested at a constant SNR level and lack of generalization ability. To address this problem, a noise-robust AMC method that can operate under varying noise conditions is proposed. The key to a noise-robust AMC algorithm is to find a feature set that is insensitive to noise to ensure that the trained classifier will be robust to SNR variations. In other words, the separating hyperplane of the feature set should remain nearly unchanged as the SNR varies. Initially, an original feature set that contains instantaneous features, HOC features, cyclostationary features and wavelet features is extracted. Subsequently, a feature selection method based on rough set theory is applied to choose noise-insensitive features from this original feature set. Thus, a noise-robust feature set is obtained that is used to trained an SVM classifier. The classification accuracy remains reasonable as the SNR varies between 5 dB and 20 dB, indicating that the proposed method can be deployed under varying noise conditions.

The remainder of this paper is organized as follows: Section II describes the basic model of an AMC system. The signal model and the extracted feature set are introduced in this section, followed by a detailed description of the presented algorithm, including feature selection and classification, in Section III. The performance of the proposed scheme is evaluated through simulation in Section IV. Finally, the paper is concluded in Section V.

II. SYSTEM MODEL

AMC is a process that takes place at the receiver, between signal detection and demodulation. The signal can be either at the base band or intermediate frequency. A general AMC system model is illustrated in Fig. 1.

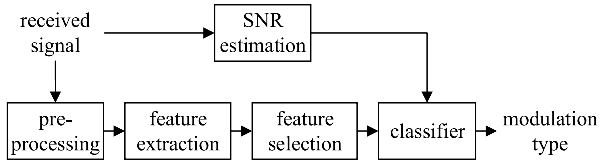

FIGURE 1. Block diagram of an AMC system model.

Pre-processing usually refers to sampling and quantization, followed by feature extraction and selection. Corresponding classifiers that are trained on the results of SNR estimation are applied to classify the modulation types of received signals.

A. SIGNAL MODEL

In this paper, the considered signals are corrupted by additive Gaussian white noise (AWGN). Thus, a digitally modulated signal can be represented as follows:

$$r(t) = s(t) + n(t) \tag{1}$$

where $s(t)$ is expressed as

$$s(t) = [A_m \sum_n a_n g(t - nT_s)] cos(2\pi(f_c + f_m)t + \varphi_0 + \varphi_m) \tag{2}$$

In (2), A_m, a_n, T_s, f_c, f_m, φ_0, and φ_m denote the modulation amplitude, symbol sequence, symbol period, carrier frequency, modulation frequency, initial phase, and modulation phase, respectively, and $g(t)$ is a function expressed as

$$g(t) = \begin{cases} 1 & \text{if } 1 \leqslant t \leqslant T_s \\ 0 & \text{other} \end{cases} \tag{3}$$

The modulation types considered in this paper are M-ASK, M-FSK, M-PSK ($M = 2, 4, 8$) and 16-QAM. M-ASK, M-FSK, and M-PSK signals can be expressed as shown in (2). However, a 16-QAM signal is slightly different; it can be represented as shown in (4):

$$\begin{aligned} s(t) = {} & [A_m \sum_n a_n g(t - nT_s)] cos(2\pi f_c t + \varphi_0) \\ & + [A_m \sum_n b_n g(t - nT_s)] sin(2\pi f_c t + \varphi_0) \end{aligned} \tag{4}$$

where $a_n, b_n \in [(2m - 1 - \sqrt{M})]$, $m = 1, 2, 3, \ldots, \sqrt{M}$, and the two carriers are individually modulated by a_n and b_n.

B. EXTRACTION OF THE ORIGINAL FEATURE SET

Before we can choose a feature set for our robust AMC algorithm, an original feature set consisting of instantaneous features, HOC features, cyclostationary features and wavelet features must first be extracted.

1) INSTANTANEOUS FEATURES

There are three important parameters in instantaneous feature extraction: the instantaneous amplitude, instantaneous phase and instantaneous frequency, which are represented by a_n,

ϕ_{NL}, and f_N, respectively. They can be easily obtained in many ways, e.g., through the Hilbert transform [8]. There are eight instantaneous features considered in this paper, most of which were proposed by Nandi and Azzouz in [16], are described as follows:

- Envelope variation:

$$m_A = \sigma^2/\mu^2 \tag{5}$$

where σ and μ are the variance and mean, respectively, of a_n.

- Standard deviation of the amplitude envelope:

$$E = \sqrt{\frac{1}{N_s - 1}\sum_{i=1}^{N_s}(a_n(i) - \frac{1}{N_s}\sum_{i=1}^{N_s} a_n(i))^2} \tag{6}$$

where N_s is the number of sampling points of the received signal.

- Maximum value of the power spectral density of the normalized-centered instantaneous amplitude:

$$\gamma_{max} = \frac{MAX\{|DFT[a_{cn}(i)^2]|\}}{N_s} \tag{7}$$

where $a_{cn} = A_n(i) - 1, A_n(i) = a_n(i)/m_a, m_a = 1/N_s \cdot \sum_i^{N_s} a_n(i)$.

- Standard deviation of the absolute value of the normalized-centered instantaneous amplitude:

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

- Standard deviation of the absolute value of the centered non-linear components of the instantaneous phase:

$$\sigma_{ap} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} \Phi_{NL}^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t}|\Phi_{NL}(i)|)^2} \tag{9}$$

- Standard deviation of the centered non-linear components of the direct instantaneous phase in non-weak segments:

$$\sigma_{dp} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} \Phi_{NL}^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t}\Phi_{NL}(i))^2} \tag{10}$$

where C is the number of samples in $\Phi_{NL}(i)$ for which $a_n(i) > a_t$, with a_t being a threshold used to eliminate noise-sensitive phases.

- Standard deviation of the absolute value of the normalized-centered instantaneous frequency:

$$\sigma_{af} = \sqrt{\frac{1}{C}(\sum_{a_n(i)>a_t} f_N^2(i)) - (\frac{1}{C}\sum_{a_n(i)>a_t} f_N(i))^2} \tag{11}$$

- Maximum value of the normalized spectrum:

$$P_{max} = \max\{\frac{|DFT(r(t))|}{\sum_{i=1}^{N}|DFT(r(t))(i)|}\} \tag{12}$$

2) HOC FEATURES

A received signal with noise can be represented in complex form as follows:

$$r(t) = A\sum_n a_n g(t - nT_s)exp[j(\omega_c t + \theta_c)] + n(t) \tag{13}$$

The sequence obtained after downconversion and sampling is represented as

$$r_k = Ae^{j\theta}a_k + n_k \quad k = 1, 2, \ldots, N \tag{14}$$

where N is the length of r_k. Let M_{km} be the mixed moment, which can be defined as $M_{km} = E[r^{k-m}(r^*)^m]$; then, cumulants of various orders are defined as follows [8]:

$$C_{20} = cum(r, r) = M_{20} \tag{15}$$

$$C_{21} = cum(r, r^*) = M_{21} \tag{16}$$

$$C_{40} = cum(r, r, r, r) = M_{40} - 3M_{20}^2 \tag{17}$$

$$C_{41} = cum(r, r, r, r^*) = M_{41} - 3M_{21}M_{20} \tag{18}$$

$$C_{42} = cum(r, r, r^*, r^*) = M_{42} - |M_{20}|^2 - 2M21^2 \tag{19}$$

$$\begin{aligned} C_{60} &= cum(r, r, r, r, r, r) \\ &= M_{60} - 15M_{40}M_{20} + 30M_{20}^3 \end{aligned} \tag{20}$$

$$\begin{aligned} C_{63} &= cum(r^*, r^*, r^*, r, r, r) \\ &= M_{63} - 9M_{41}M_{21} - 6M_{21}^3 \end{aligned} \tag{21}$$

In practice, the length of a received signal is limited, meaning that the HOC features must be estimated as follows:

$$\hat{C}_{20} = \hat{M}_{20} = \frac{1}{N}\sum_{k=1}^{N} r_k^2 \tag{22}$$

$$\hat{C}_{21} = \hat{M}_{21} = \frac{1}{N}\sum_{k=1}^{N} |r_k|^2 \tag{23}$$

$$\hat{C}_{40} = \hat{M}_{40} - 3\hat{M}_{20}^2 = \frac{1}{N}\sum_{k=1}^{N} r_k^4 - 3\hat{M}_{20}^2 \tag{24}$$

$$\hat{C}_{41} = \hat{M}_{41} - 3\hat{M}_{20}\hat{M}_{21} = \frac{1}{N}\sum_{k=1}^{N} r_k^3 r_k^* - 3\hat{M}_{20}\hat{M}_{21} \tag{25}$$

$$\begin{aligned} \hat{C}_{42} &= \hat{M}_{42} - |\hat{M}_{20}|^2 - 2\hat{M}_{21}^2 \\ &= \frac{1}{N}\sum_{k=1}^{N} |r_k|^4 - |\hat{M}_{20}|^2 - 2\hat{M}_{21}^2 \end{aligned} \tag{26}$$

$$\begin{aligned} \hat{C}_{60} &= \hat{M}_{60} - 15\hat{M}_{40}\hat{M}_{20} + 30\hat{M}_{20}^3 \\ &= \frac{1}{N}\sum_{k=1}^{N} r_k^6 - 15\hat{M}_{40}\hat{M}_{20} + 30\hat{M}_{20}^3 \end{aligned} \tag{27}$$

$$\begin{aligned} \hat{C}_{63} &= \hat{M}_{63} - 9\hat{M}_{41}\hat{M}_{21} - 6\hat{M}_{21}^3 \\ &= \frac{1}{N}\sum_{k=1}^{N} |r_k|^6 - 9\hat{M}_{41}\hat{M}_{21} - 6\hat{M}_{21}^3 \end{aligned} \tag{28}$$

The final seven HOC features extracted for classification are represented as follows:

$$\begin{aligned} d_1 &= |C_{40}|/|C_{42}| \\ d_2 &= |C_{41}|/|C_{42}| \\ d_3 &= |C_{42}|/|C_{21}|^2 \\ d_4 &= |C_{60}|/|C_{21}|^3 \\ d_5 &= |C_{63}|/|C_{21}|^3 \\ d_6 &= |C_{60}|^2/|C_{42}|^3 \\ d_7 &= |C_{63}|^2/|C_{42}|^3 \end{aligned} \tag{29}$$

3) WAVELET FEATURES

The continuous wavelet transform of a signal $r(n)$ is defined as follows:

$$C(j,k) = \sum_{n\in Z} r(n)\psi_{j,k}(t) \tag{30}$$

where $\psi_{j,k}(t)$ is the wavelet function, $C(j,k)$ is called the wavelet coefficient, j is the scale variable, and k is the translation variable. The Daubechies-3 wavelet is chosen as the mother wavelet function. Then, the original signal is represented as follows:

$$r(t) = \sum_{j\in Z}\sum_{k\in Z} C(j,k)\psi_{j,k}(t) \tag{31}$$

We apply the algorithm proposed in [17] to obtain different levels of decompositions of the received signal and thus to extract the low-frequency component at each level. In this paper, five levels of signal decomposition are performed. Let A and D represent the low-frequency component and the high-frequency component, respectively. The tree structure of the decomposition process is shown in Fig. 2.

Let d_m be the high-frequency-component coefficient of the mth-level decomposition; then, the signal energy of the mth-level decomposition is defined as

$$E_m = \sum_n d_m(n)^2 \tag{32}$$

In total, five levels of decomposition are applied to the received signal, and thus, E_1, E_2, E_3, E_4, and E_5 are calculated in accordance with (32). The results are considered as extracted features for use in classification.

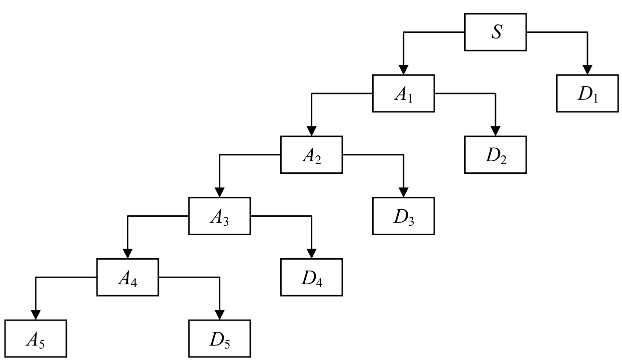

FIGURE 2. Tree structure of multi-scale wavelet decomposition.

4) CYCLOSTATIONARY FEATURES

Under the assumption that the received signal $r(t)$ is a cyclostationary signal, the mean and autocorrelation of the signal vary periodically with time T: $M_r(t+T) = M_r(t)$ and $R_r(t+T, u+T) = R_r(t,u)$ for all t and u. The cyclic autocorrelation function can be expressed as $R_x(t+\tau/2, t-\tau/2)$, which is also periodic in t with period T. It can be expressed as a Fourier series as follows:

$$R_r(t+\tau/2, t-\tau/2) = \sum_\alpha R_r^\alpha(\tau)e^{j2\pi\alpha t} \tag{33}$$

where $\alpha = m/T_0, (m \in \mathbb{Z})$ is the cyclic frequency. The Fourier coefficient of (33), which is also defined as the cyclic autocorrelation function, can be represented by

$$\begin{aligned} R_r^\alpha(\tau) &= \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} R_r(t+\tau/2, t-\tau/2)e^{-j2\pi\alpha t}dt \\ &= \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} r(t+\tau/2)r^*(t-\tau/2)e^{-j2\pi\alpha t}dt \end{aligned} \tag{34}$$

The Fourier transform of (34) at cyclic frequency α is called the spectral correlation density (SCD) function, which is expressed as

$$S_r^\alpha(f) = \int_{-\infty}^{\infty} R_r^\alpha(\tau)e^{-j2\pi f\tau}d\tau \tag{35}$$

The length of the signal at the receiver is considered to be infinite; however, the length of a received signal is limited in practice. Here, we use the smoothed cyclic periodogram method [18], which is an approach for estimating the spectral correlation from the received signal. The cyclic periodogram of a signal $x(t)$ is defined as follows:

$$S_{rT}^\alpha = \frac{1}{T}X_T(t, f+\alpha/2)X_T^*(t, f-\alpha/2) \tag{36}$$

where X_T is the time-variant Fourier transform, which is defined as

$$X_T(t,f) = \int_{t-T/2}^{t+T/2} r(u)e^{-j2\pi fu}du \tag{37}$$

The SCD function that is estimated using the frequency-smoothed cyclic periodogram is represented as

$$S_{rT}^\alpha(t,f)_{\Delta f} = \frac{1}{\Delta f}\int_{f-\Delta f/2}^{f-\Delta f/2} S_{rT}^\alpha(t,v)dv \tag{38}$$

The SCD function can also be inversely expressed as

$$S_r^\alpha(f) = \lim_{\Delta f\to\infty}\lim_{T\to\infty} S_{rT}^\alpha(t,f)_{\Delta f} \tag{39}$$

In most cyclostationary methods, the highest values of the SCD for a given α are taken by a function called *profile*(α) [2] and are directly classified by an ANN with no additional extracted features. However, there are still some other features that are valuable for classification and are worth extracting. Based on $S_x^\alpha(f)$ as obtained above, in total, 5 features are extracted:

- $$R_1 = 1/\sigma_1^2, \quad \sigma_1^2 = \mathrm{var}[S_r^\alpha(f=0)] \tag{40}$$
- $$R_2 = 1/\sigma_2^2, \quad \sigma_1^2 = \mathrm{var}[S_r^\alpha(f=f_c)] \tag{41}$$
- $$R_3 = 1/\mu, \quad \mu = \mathrm{mean}[S_r^\alpha(f=f_0)] \tag{42}$$
- $$\beta = |\frac{\max\{S_r^\alpha(0)\}}{\max\{S_r^0(f)\}}| \tag{43}$$
- $$P = \int_0^{f_c} |S_r^{2f_c}(f)^2|df \tag{44}$$

III. PROPOSED METHOD

In this section, we describe the framework of our proposed algorithm. The method attempts to select noise-insensitive features from the original feature set obtained as described in section II and to use them to train an SVM classifier. The trained SVM should still be suitable for classification when the results of SNR estimation are inaccurate.

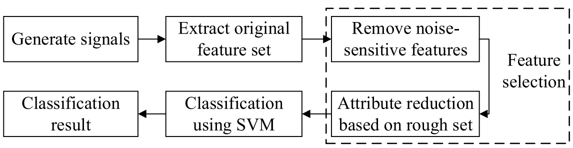

FIGURE 3. Process of the proposed AMC algorithm.

The process of the proposed algorithm is illustrated in Fig. 3. The model of the received signal and the original feature set have already been introduced in Section II. Feature selection is a two-step process, the purpose of which is to identify noise-insensitive features with which to form a robust feature set. The normalized-centered variances of the features, which directly reflect the fluctuations of different features, are chosen as the parameters for evaluating whether a feature is sensitive to noise. First, the normalized-centered variance of each feature is calculated, based on which we remove noise-sensitive features from the feature set. Then, attribute reduction based on rough set theory is applied to remove possible redundant features. The feature set we obtain after this two-step feature selection process is used to train the SVM classifier. Finally, the generalization ability of the trained SVM is evaluated based on numerical results.

A. ROUGH SET THEORY

Rough set theory is widely used in data analysis. It enables the removal of redundant information from a given data set through attribute reduction while preserving the original classification accuracy and reducing the dimension of the feature set at the same time. In rough set theory, an information system is represented by $S = \{U, R, V, F\}$. $U = \cup x_r$ is defined as the universe, which is a finite and non-empty set that contains all objects. R is a finite set consisting of attributes and can be defined as $C \cup D$, where C and D are the condition attributes and decision attributes, respectively. $V = \cup v_r$ is defined as the set of attribute values. f represents the information function mapping U and R to V as follows: $f : U \times R \rightarrow V$. U can be divided by R to produce equivalent classes of different attributes, expressed as $U/R = \cup E_i$. If there are two different elements $u, v \in U$ in the same equivalent class E_i for R, they are said to be indistinguishable, expressed as $ind(R)$. If $ind(R - r) = ind(R)$, then r is unnecessary to R; otherwise, r is necessary to R.

Suppose that in an information system S, there exists a subset X of U ($X \subseteq U$) and $P \subseteq R$; then, the lower approximation and upper approximation of X can be defined as follows:

$$\underline{R}X = \{Y \in U/R : Y \subseteq X\} \tag{45}$$
$$\overline{R}X = \{Y \in U/R : Y \cap X \neq \phi\} \tag{46}$$

where $\underline{R}X$ represents the set that can certainly be merged into X, or the positive region, also denoted by $POS(X)$. By contrast, $\overline{R}X$ is the set that can possibly be merged into X. If $\underline{R}X = \overline{R}X$, then X is a precise representation of U; otherwise, X is a rough set. The most important parameter in attribute reduction is the dependent extent. Given $P, Q \subseteq U$, the dependent extent of P to Q is defined as

$$\gamma_Q(P) = POS_Q(P)/|U| \tag{47}$$

where $\gamma_Q(P)$ can be seen as the positive region of Q in $U/ind(P)$. P is completely dependent on Q if $\gamma_Q = 1$. Furthermore, for a newly added attribute a in Q, the attribute importance of a for $U/ind(P)$ is defined as

$$SGF(a, Q, P) = \gamma_Q(p) - \gamma_{Q-\{a\}}(p) \tag{48}$$

If $SGF(a, Q, P) > SGF(b, Q, P)$, then attribute a is more important than b for $U/ind(P)$.

B. SUPPORT VECTOR MACHINE

Consider a supervised classification problem in which the training set $(\boldsymbol{x}_i, y_i)$, with $\boldsymbol{x}_i \in \Re^d$ and $y_i \in \{1, -1\}$, can be classified with respect to the hyperplane defined by $\boldsymbol{w} \cdot \boldsymbol{x} + b = 0$. This binary classification problem can be formulated as

$$y_i\,(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geqslant 1 \tag{49}$$

The margin between hyperplanes is $\frac{2}{\|\boldsymbol{w}\|}$. Maximizing $\frac{2}{\|\boldsymbol{w}\|}$ is equivalent to minimizing $\frac{1}{2}\|\boldsymbol{w}\|^2$. However, the margin $\frac{2}{\|\boldsymbol{w}\|}$ allows no tolerance for misclassification, which is often not feasible in practice. By introducing slack variables $(\xi_1, \xi_2, \ldots, \xi_i)$, (49) can be modified as follows:

$$y_i\,(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \geqslant 1 - \xi_i, \quad i = 1, 2, \ldots N \tag{50}$$

Thus, the basic SVM model is a programming problem described as follows:

$$\min_{\boldsymbol{w},b,\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geqslant 1-\xi_i$$
$$\xi_i \geqslant 0, \quad i=1,2,\ldots,N \tag{51}$$

C is a penalty parameter used to control the error. The dual problem corresponding to (51) can be obtained by applying Lagrange multipliers $\alpha_i \geqslant 0$. This leads to another programming problem, expressed as

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)$$
$$\text{subject to } \sum_{i=1}^{N}\alpha_i y_i = 0,$$
$$0 \leqslant \alpha_i \leqslant C, \quad i=1,2,\ldots,N \tag{52}$$

which is the linear SVM model. For non-linear problems, a kernel function $K(\boldsymbol{x}_i,\boldsymbol{x}_j)$ is applied to map linearly inseparable data into a high-dimensional space. This paper considers the RBF kernel function, which takes the following form:

$$K(\boldsymbol{x}_i,\boldsymbol{x}_j) = \exp\left\{-\gamma\left\|\boldsymbol{x}_i-\boldsymbol{x}_j\right\|^2\right\} \tag{53}$$

By replacing $\boldsymbol{x}_i\cdot\boldsymbol{x}_j$ with (53), solutions for non-linear data can be obtained. The decision function, which determines the result of classification, can be written as follows:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N}\alpha_i y_i K(\boldsymbol{x}_i,\boldsymbol{x}_j)+b\right) \tag{54}$$

The performance of an SVM classifier is determined by the selection of the parameters C and γ.

C. FEATURE SELECTION

In essence, training a classifier usually means determining a hyperplane for classification. The reason why trained classifiers can no longer function correctly when the SNR changes is that some values of the extracted feature changes as the SNR varies. This will cause a pre-trained hyperplane to become no longer feasible for classification. This is also why HOC features are usually suitable for noise-robust AMC algorithms. To achieve the purpose of finding noise-robust features, we first need to calculate the variances of all features. We define the normalized variance of a feature F as shown in (55):

$$\text{var}(F) = \frac{1}{K}\sum_{i=1}^{10}\sum_{k=1}^{K}[F_i(k)-E(F_i)]^2 \tag{55}$$

where K is the number of training data and $F_i(k)$ represents the normalized center value of $f_i(k)$, which is expressed as

$$F_i(k) = \frac{f_i(k)}{\frac{1}{K}\sum_{i=1}^{K}f_i(k)} \tag{56}$$

TABLE 1. Normalized variances of extracted features.

Feature	Var	Feature	Var	Feature	Var
m_A	5.25	d_1	0.16	E_4	13.06
E	2.18	d_2	0.01	E_5	13.54
γ_{max}	0.95	d_3	1.04	R_1	0.01
σ_{aa}	1.59	d_4	1.21	R_2	0.02
σ_{ap}	1.15	d_5	1.54	R_3	0.51
σ_{dp}	1.32	d_6	0.06	β	0.01
σ_{af}	1.28	d_7	1.2	P	0.20
P_{max}	1.22	E_1	0.19		
E_2	3.59	E_3	6.63		

where $f_i(k)$ is the feature value of the kth sample of the ith modulation type.

The normalized variances of all features can be calculated based on (55), with the results shown in Table 1.

Features with excessively large variances need to be removed, whereas those with small variances should be kept. We apply the K-means clustering method to divide the features into the following clusters:

- $\{d_1, d_2, d_6, E_1, R_1, R_2, R_3, \beta, P\}$
- $\{E, \gamma_{max}, \sigma_{aa}, \sigma_{ap}, \sigma_{dp}, \sigma_{af}, P_{max}, d_3, d_4, d_5, d_7\}$
- $\{E_4, E_5\}$
- $\{m_A, E_3\}$
- $\{E_2\}$

The first cluster, $\{d_1, d_2, d_6, E_1, R_1, R_2, R_3, \beta, P\}$, which has the minimum average var(F) among all clusters, is retained as the set of noise-robust features. The other clusters are removed from the feature set. Although the chosen feature set is already a robust one, it still needs further reduction because different features may carry the same information about the received signal, which would lead to information redundancy.

This reduction can be achieved through attribute reduction based on rough set theory. In AMC, the extracted features and modulation types are regarded as the condition attributes F and decision attributes D, respectively, based on which a decision table is first established. The features are then normalized and discredited. The attribute importance of a feature F_n is calculated from the discredited data as follows:

$$\gamma_D(F_n) = \gamma_F - \gamma_{F-F_n} \tag{57}$$

A feature is eliminated from the feature set if its attribute importance is very close to zero. Let S denote the decision table; the details of the attribute reduction process are shown in Table 2.

The knowledge rules for the features can be obtained based on the results of attribute reduction. The resulting rules are shown in Table 3. Finally, 4 features remain in the feature set, the values of which are illustrated in Fig. 4. The 4-dimensional feature set obtained after feature selection is referred to as the noise-robust feature set.

TABLE 2. Process of attribute reduction.

1. begin
2. initialize input = $S = (U, F \cup D, V, f)$, Core(F) = ϕ
3. for F_n in F
4. calculate $POS_{F-F_n}(D)$
5. if $POS_{F-F_n}(D) \neq POS_F(D)$
6. Core(F)=Core(F)$\cup F_n$
7. end if
8. end for
9. R = Core(F)
10. calculate $\gamma_R(D)$
11. if $\gamma_R(D) = \varepsilon$
12. to 22
13. else
14. to 16
15. end if
16. for F_n in $F - R$
17. calculate $SGF(F_n, F, D)$
18. select the attribute F_i with maximum $SGF(F_i, F, D)$
19. $R = R \cup F_n$
20. end for
21. to 11
22. output = R
23. end

TABLE 3. Knowledge rules for the extracted features.

	d_1	d_2	d_6	R_1
2-ASK	$[0.8, \infty)$	$[0.8, \infty)$	[2.59,19.81)	[2915,5491)
4-ASK	$[0.8, \infty)$	$[0.8, \infty)$	[19.81,23.52)	[2915,5491)
8-ASK	$[0.8, \infty)$	$[0.8, \infty)$	[23.52,29.12)	[2915,5491)
2-FSK	$[0, 0.2)$	$[0, 0.2)$	[0,2.59)	[1954,2915)
4-FSK	$[0, 0.2)$	$[0, 0.2)$	[0,2.59)	[885,1954)
8-FSK	$[0, 0.2)$	$[0, 0.2)$	[0,2.59)	[0,885)
2-PSK	$[0.8, \infty)$	$[0.8, \infty)$	$[29.12, \infty)$	$[5491, \infty)$
4-PSK	$[0.8, \infty)$	$[0, 0.2)$	[0,2.59)	$[5491, \infty)$
8-PSK	$[0, 0.2)$	$[0, 0.2)$	[0,2.59)	$[5491, \infty)$
16-QAM	$[0.8, \infty)$	$[0, 0.2)$	[0,2.59)	[2915,5491)

The noise-robust feature set consists of 3 HOC features and 1 cyclostationary feature. It can be observed from Fig. 4 that the values of the 4 features remain almost unchanged at all SNR levels. The distances between classes are also significant, thereby satisfying the demands of classification. The contributions of each feature to the classification task can be described as follows:

$$ind(d_1) = \{\{\text{2-ASK, 4-ASK, 8-ASK, 2-PSK, 4-PSK, 16-QAM}\},\{\text{2-FSK, 4-FSK, 8-FSK, 8-PSK}\}\}$$

$$ind(d_2) = \{\{\text{2-ASK, 4-ASK, 8-ASK, 2-PSK}\}, \{\text{2-FSK, 4-FSK, 8-FSK, 4-PSK, 8-PSK, 16-QAM}\}\}$$

$$ind(d_6) = \{\{\text{2-ASK}\}, \{\text{4-ASK}\}, \{\text{8-ASK}\}, \{\text{2-PSK}\}, \{\text{2-FSK, 4-FSK, 8-FSK, 4-PSK, 8-PSK, 16-QAM}\}\}$$

$$ind(R_1) = \{\{\text{2-ASK, 4-ASK, 8-ASK, 16-QAM}\}, \{\text{2-FSK}\}, \{\text{4-FSK}\}, \{\text{8-FSK}\}, \{\text{2-PSK, 4-PSK, 8-PSK}\}\}$$

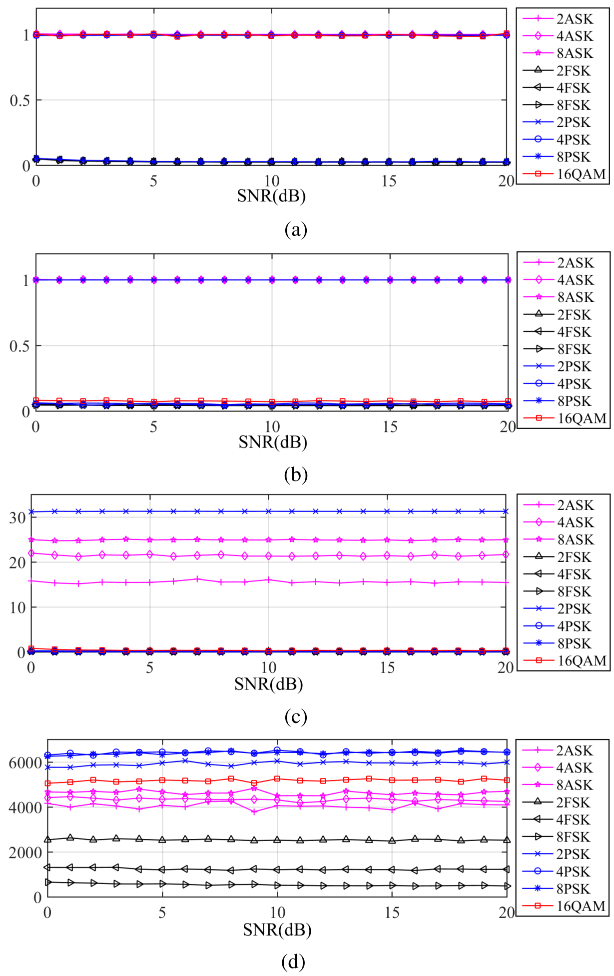

FIGURE 4. Values of the 4 selected features under SNR levels from 0 to 20 dB. (a) d_1. (b) d_2. (c) d_6. (d) R_1.

TABLE 4. Modulation parameters.

Parameter	Symbol	Value
Code rate	f_d	2 $Mbps$
Carrier frequency	f_c	70 MHz
Sampling frequency	f_s	200 MHz
Frequency interval	$f_\triangle$	1 MHz

IV. PERFORMANCE EVALUATION

Here, the classification accuracies of the original feature set and the noise-robust feature set under SNRs of 0∼20 dB are evaluated. We generated 400 samples at each SNR level, that is, 8400 samples in total. The other parameters are listed in Table 4.

The classification accuracy of the original feature set with an SVM classifier is evaluated first. Classifiers trained at different fixed SNRs in the range of 0∼20 dB were tested at all SNR levels from 0 to 20 dB to evaluate the classification accuracy and generalization ability of the original feature set. The results are illustrated in Fig. 5.

The performance of the original feature set reaches an extremely high accuracy rate at the SNR for which the classi-

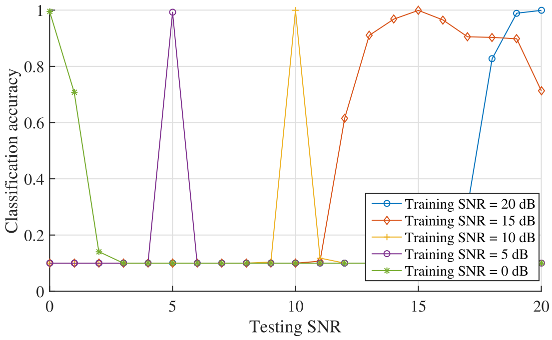

FIGURE 5. Classification accuracy of the original feature set.

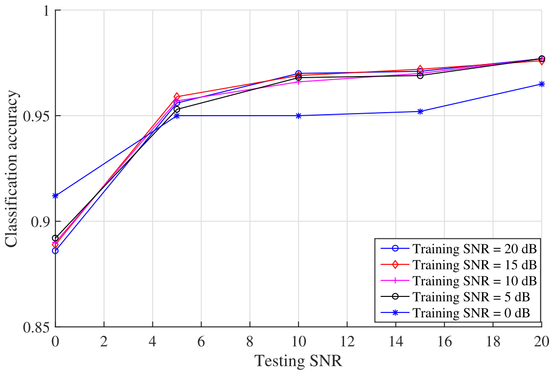

FIGURE 6. Classification accuracy of the noise-robust feature set.

fier was trained. However, the lower is the SNR under which the classifier was trained, the worse is the generalization ability of that classifier. For a training SNR lower than 10 dB, the classifier is no longer effective even for a testing SNR that is only 1 dB different from the training SNR.

Next, the classification accuracy of the selected feature set is evaluated. The selected feature set is $\{d_1, d_2, d_6, R_1\}$ and is tested in the same way. The results are shown in Fig. 6. The classification accuracy at the training SNR is decreased by approximately 2.5% compared with that in Fig. 5, but the accuracy reaches 95% as soon as the SNR of the test set increases to 5 dB. The classification accuracy is above 97% when the testing SNR is at least 10 dB. At the cost of a decreased classification accuracy at the training SNR, excellent generalization ability is achieved. However, the accuracy rate at an SNR level of 0 dB is obviously worse than that at any other SNR. To find out the reason, the detailed data on the classification accuracies for each modulation type at training SNRs of 20 dB, 10 dB and 0 dB are shown in Tables 5, 6 and 7, respectively.

We find that the classification accuracies are all very close to 100% except for M-ASK signals, especially 4-ASK signals. When M-ASK signals are not considered, the classification accuracies for all modulation types under all SNR levels are very close to 100%. Only the accuracy for 8-FSK signals is slightly decreased (96.6% and 94.3%) at 0 dB. The average classification accuracy for M-FSK, M-PSK and 16-QAM signals is greater than 99% for SNR levels from 5 to 20 dB.

TABLE 5. Classification accuracies for each class (training SNR=20 dB).

SNR / class	0 dB	5 dB	10 dB	15 dB	20 dB
2-ASK	0.824	0.918	0.928	0.916	0.914
4-ASK	0.526	0.839	0.869	0.856	0.858
8-ASK	0.658	0.921	0.947	0.942	0.949
2-FSK	1.000	1.000	1.000	1.000	1.000
4-FSK	0.999	0.999	1.000	1.000	1.000
8-FSK	0.943	0.986	0.991	0.997	0.996
2-PSK	0.991	1.000	1.000	1.000	1.000
4-PSK	0.977	0.993	0.999	0.997	0.998
8-PSK	1.000	1.000	1.000	1.000	1.000
16-QAM	0.996	0.995	1.000	0.999	1.000

TABLE 6. Classification accuracies for each class (training SNR=10 dB).

SNR / class	0 dB	5 dB	10 dB	15 dB	20 dB
2-ASK	0.837	0.932	0.948	0.940	0.934
4-ASK	0.478	0.834	0.857	0.845	0.851
8-ASK	0.653	0.925	0.947	0.943	0.950
2-FSK	1.000	1.000	1.000	1.000	1.000
4-FSK	0.999	1.000	1.000	1.000	1.000
8-FSK	0.966	0.991	0.992	0.997	0.998
2-PSK	0.998	1.000	1.000	1.000	1.000
4-PSK	0.990	0.998	1.000	0.997	0.999
8-PSK	1.000	1.000	1.000	1.000	1.000
16-QAM	0.995	0.993	0.998	0.998	0.999

TABLE 7. Classification accuracies for each class (training SNR=0 dB).

SNR / class	0 dB	5 dB	10 dB	15 dB	20 dB
2-ASK	0.840	0.928	0.910	0.940	0.920
4-ASK	0.443	0.673	0.650	0.600	0.750
8-ASK	0.852	0.913	0.950	0.960	0.980
2-FSK	1.000	1.000	1.000	1.000	1.000
4-FSK	1.000	1.000	1.000	1.000	1.000
8-FSK	1.000	1.000	1.000	1.000	1.000
2-PSK	1.000	1.000	1.000	1.000	1.000
4-PSK	1.000	1.000	1.000	1.000	1.000
8-PSK	1.000	1.000	1.000	1.000	1.000
16-QAM	1.000	1.000	0.990	1.000	1.000

By contrast, the results for ASK signals are not reasonable, especially when the SNR is near 0 dB. Because we consider the AWGN channel model, the noise is added directly to the signal amplitude, which makes it difficult to extract noise-robust features for ASK signals. The classification of ASK signals mainly depends on R_1 and d_6, and the distances between classes are small. The value for 4-ASK varies between those for 2-ASK and 8-ASK, which makes the performance worse, especially when the SNR is low. Therefore, it will be necessary to continue seeking better features for M-ASK classification, especially for cases in which the SNR is low.

V. CONCLUSION

This paper mainly focuses on addressing the problem that existing FB AMC algorithms are deployed for a specific fixed SNR and lack generalization ability. An AMC method that can operate under varying noise conditions is proposed to solve this problem. An original feature set containing

25 types of features is initially established, from which we choose noise-insensitive features to form a robust feature set. Features are preliminarily selected based on the normalized variances of all features. Features with excessively large variances are eliminated from the feature set. Then, the feature set is simplified through attribute reduction based on rough set theory. We finally obtain a robust feature set consisting of only 4 features. Subsequently, the robust feature set is used to train an SVM classifier. Numerical results demonstrate that an SVM classifier trained on the robust feature set at a single SNR level can successfully classify signals at SNRs of 5$\sim$20 dB. The average accuracy rate of the algorithm is higher than 95% for all signals at 5$\sim$20 dB. The average classification accuracy for M-FSK, M-PSK and 16-QAM signals can reach 99%. The generalization ability is excellent, demonstrating that our method is robust to SNR variations; therefore, it can be deployed under varying noise conditions.

REFERENCES

[1] A. K. Nandi and E. E. Azzouz, "Automatic analogue modulation recognition," *Signal Process.*, vol. 46, no. 2, pp. 211–222, Oct. 1995.

[2] U. Satija, M. S. Manikandan, and B. Ramkumar, "Performance study of cyclostationary based digital modulation classification schemes," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–5.

[3] S. Vukotić and D. Vučić, "Cyclic feature-based joint detection and clasiffication of OFDM/QAM, OFDM/OQAM and SC-FDMA signals," in *Proc. 24th Telecommun. Forum (TELFOR)*, Nov. 2016, pp. 1–4.

[4] J. Liu and Q. Luo, "A novel modulation classification algorithm based on daubechies5 wavelet and fractional fourier transform in cognitive radio," in *Proc. IEEE 14th Int. Conf. Commun. Technol. (ICCT)*, Nov. 2012, pp. 115–120.

[5] K. Hassan, I. Dayoub, W. Hamouda, and M. Berbineau, "Automatic modulation recognition using wavelet transform and neural network," in *Proc. 9th Int. Conf. Intell. Transp. Syst. Telecommun.*, Oct. 2009, pp. 234–238.

[6] D. Das, A. Anand, P. K. Bora, and R. Bhattacharjee, "Cumulant based automatic modulation classification of QPSK, OQPSK, Π/4-QPSK and 8-PSK in MIMO environment," in *Proc. Int. Conf. Signal Process. Commun. (SPCOM)*, Jun. 2016, pp. 1–5.

[7] C. M. Spooner, "Classification of co-channel communication signals using cyclic cumulants," in *Proc. Conf. Rec. 29th Asilomar Conf. Signals, Syst. Comput.*, Oct. 1995, pp. 531–536.

[8] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. 1st Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.

[9] J. Han, M.-H. He, Z.-B. Zhu, and J. Wang, "Sorting unknown radar emitter signal based on the complexity characteristics," *J. Electron. Inf. Technol*, vol. 31, no. 11, pp. 2552–2556, 2009.

[10] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. Int. Workshop Artif. Intell. Stat.*, 2003, pp. 1–8.

[11] M. L. D. Wong and A. K. Nandi, "Automatic digital modulation recognition using artificial neural network and genetic algorithm," *Signal Process.*, vol. 84, no. 2, pp. 351–365, Feb. 2004.

[12] E. Avci, "Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm–support vector machines: HGASVM," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1391–1402, Mar. 2009.

[13] C. Louis and P. Sehier, "Automatic modulation recognition with a hierarchical neural network," in *Proc. Conf. Rec. IEEE Military Commun. Conf. (MILCOM)*, Oct. 1994, pp. 713–717.

[14] A. K. Nandi and E. E. Azzouz, "Modulation recognition using artificial neural networks," *Signal Process.*, vol. 56, no. 2, pp. 165–175, 1997.

[15] Z. Yin, "Research of communication signal modulation scheme recognition based on one-class SVM Bayesian algorithm," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2009, pp. 1–4.

[16] E. E. Azzouz and A. K. Nandi, "Automatic identification of digital modulation types," *Signal Process.*, vol. 47, no. 1, pp. 55–69, Nov. 1995.

[17] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 7, pp. 710–732, Jul. 1992.

[18] R. S. Roberts, W. A. Brown, and H. H. Loomis, "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Process. Mag.*, vol. 8, no. 2, pp. 38–49, Apr. 1991.

ZHILU WU is currently a Professor with the School of Electronics Information Engineering, Harbin Institute of Technology. His research interests include space information acquisition and processing, formation flying satellite control, cognitive radio, and software radio.

SIYANG ZHOU received the M.S. degree from the Harbin Institute of Technology in 2014, where he is currently pursuing the Ph.D. degree with the School of Electronics Information Engineering. His current research interests include deep learning, modulation classification, and power allocation.

ZHENDONG YIN (M'15) received the Ph.D. degree from HIT in 2008. He is currently an Associate Professor with the School of Electronics Information Engineering, Harbin Institute of Technology. His current research interests include UWB wireless communication, formation flying satellites communication, and relay satellite system.

BO MA received the M.Sc. degree in electronics and communication engineering from the School of Information Science and Technology, Tsinghua University, Beijing, China, in 2007. He is currently pursuing the Ph.D. degree with the Harbin Institute of Technology. He is currently with the China Academy of Space Technology. His research interests include satellite communication, wireless relaying, and cognitive radio.

ZHUTIAN YANG (M'13) received the Ph.D. degree from Harbin Institute of Technology (HIT) in 2013. He was an Academic Visitor with King's College London, in 2015. He is currently an Associate Professor with the School of Electronics Information Engineering, HIT. His current research interests include machine learning, signal processing, and smart city communications.