A recognition algorithm for BPSK/QPSK signals based on generalized Pareto distribution

Yang Li[1,2], Hu Guobing[1], Hu Xuelong[2]，Xu Xiaoyang[1]
1. Department of Electronic Information Engineering, Jinling Institute of Technology, Nanjing 211169,China;
2. School of Information Engineering, Yangzhou University, Yangzhou 225127, China
Email: yangli691@jit.edu.cn, huxuelong@126.com

Abstract— *In this paper, an algorithm for recognizing BPSK/QPSK signal is proposed. Extreme value theory is introduced, and an exceedance sequence is built from the modified spectrum obtained by the squaring operation of the observed signal, and the problem of recognizing the BPSK/QPSK signal is transformed to the problem of whether the distribution of the exceedance sequence approximately follows its standard generalized Pareto distribution or not. Finally, the Kolmogorov–Smirnov test is applied to perform the goodness-of-fit test of the generalized Pareto distribution of the exceedance sequence to complete the task of BPSK/QPSK signal recognition. Computer simulation results show that the recognition performance is good when the signal-to-noise ratio is moderate and does not require prior information such as noise variance.*

Key words— *BPSK/QPSK signal; signal recognition; extreme value theory; generalized Pareto distribution; Kolmogorov–Smirnov test*

Ⅰ INTRODUCTION

Modulation recognition of signals has a wide range of applications in digital receivers used in radar electronic warfare or software-defined radio[1-9]. It is used as an intermediate stage between signal modulation and demodulation. Modulation recognition involves observing the signal, determining the modulation mode, and estimating the corresponding parameters. Considering the modulation recognition problem, several scholars have conducted research from different perspectives. In [10], signal noise is estimated by using the Wigner–Ville distribution, and the features are extracted by employing time–frequency analysis; finally, parameters are computed by using the fractional Fourier transform to achieve signal recognition. This technique is applied in various systems, such as electronic support，electronic attack，and multiple-input multiple-output systems. The recognition performance achieved with this technique is satisfactory under low signal-to-noise ratio (SNR) conditions. In [11], features are extracted by using Wigner and Choi–Williams time–frequency distributions, and features that are not useful are rejected by an information-theoretic feature selection algorithm. The technique is applied in spectrum management and supervision of cognitive radio and radar systems, and the recognition performance is satisfactory when the SNR is moderate. Another similar approach using time–frequency analysis based on noncoherent integration of the short-time Fourier transform is presented in [12] and multiple signals are recognized effectively by the proposed algorithm. In [13], to address the modulation signal blind recognition problem under low SNR conditions, an energy focusing efficiency feature is defined to identify the modulation type of the signals on the basis of analysis of the condition for the single sinusoid generated from six types of intrapulse modulation signals. The algorithm is effective for the recognition of the common intrapulse modulation signals under low SNR conditions.

In this paper, we introduce the extreme value theory (EVT) theory, and present an algorithm for BPSK/QPSK signal modulation recognition based on generalized Pareto distribution (GPD). First, the amplitude spectrum of squaring operation of the observed BPSK/QPSK signal is transformed into the modified spectrum, and a new sequence is obtained by the square of the modified spectrum. We set a threshold, and the random sequence is generated from the peak over the threshold model, and is named as exceedance sequence in this paper. Finally, the Kolmogorov–Smirnov (KS) test is applied to test the goodness-of-fit between the exceedance sequence distribution and the GPD. A fit indicates that the signal is BPSK; otherwise, the signal is QPSK.

Ⅱ SIGNAL MODEL

The observed discrete-time signal model can be expressed as

$$x(n)=s(n)+w(n)=A\exp(j[\phi(n)])+w(n),0\le n\le N-1 \quad (1)$$

where $s(n)$ is the signal, A and $\phi(n)$ are the amplitude and the phase function of the signal respectively, and N is the length of the observed samples. $w(n)$ denotes the white Gaussian noise, which is uncorrelated to the signal $s(n)$. As for phase modulation, the phase function $\phi(n)$ implies a different modulation mode.

Here, the phase function is defined as

978-1-5090-5035-2/17/$31.00 ©2017 IEEE

$$\phi(n) = 2\pi f_0 n\Delta t + \pi d_i(n) + \theta \quad (2)$$

where f_0 is the signal carrier frequency, Δt is the sampling interval, and θ is the initial phase. $d_i(n)$ denotes the modulation mode of the signal; when $i=0$, the signal is BPSK, and when $i=1$, the signal is QPSK. Therefore, the problem of recognizing BPSK/QPSK signal is summed up as follows: under the H_0 hypothesis, the signal is BPSK; under the H_1 hypothesis, the signal is QPSK.

Ⅲ ALGORITHM DESCRIPTION

A. Modified spectrum

Let the observed BPSK/QPSK signal be $x(n)$; $y(n)$ is given as $y(n)=x(n)\hat{}2$. The amplitude spectrum of the discrete Fourier transformation of $y(n)$ is given by $Y(k)=|DFT(y(n))|$. The position of the maximum value of $Y(k)$ is recorded as k_0, whose left and right N spectral lines are turned to 0. Now, the modified spectrum can be calculated by using the expression

$$Z(k)=\begin{cases}0, & k\in(k_0-N,k_0+N)\\ Y(k), & k\notin(k_0-N,k_0+N)\end{cases} \quad (3)$$

It is obvious that under the H_0 hypothesis, $y(n)$ is a sine wave, and its amplitude spectrum $Y(k)$ has one peak spectrum line, the value of which is the largest among $Y(k)$. Therefore, the modified spectrum $Z(k)$ is the actual noise spectrum, the values of which are all small. Meanwhile, under the H_1 hypothesis, $y(n)$ is BPSK, and its amplitude spectrum $Y(k)$ has several peak spectrum lines, the values of which are larger, so the modified spectrum $Z(k)$ also has several peak spectrum lines.

The new sequence $P(k)$ is defined as $P(k)=Z(k)\hat{}2$. According to the analysis above, under the H_0 hypothesis, $P(k)$ is the noise spectrum; under the H_1 hypothesis, $P(k)$ still has some peak lines.

B. Exceedance distribution of modified spectrum

Theorem: Assuming that X_1, X_2, X_3, … are independent and identically distributed (IID) random variables whose distribution function is $F(x)$, let $M_n=\max\{X_1,X_2,...,X_n\}$, if normalized series $\{a_n>0\}$, $\{b_n\}$ are existed, as for large enough n, there is the conclusion

$$P_r(M_n \le a_n x + b_n) \approx H(x;\mu.\sigma.\varepsilon) \quad (4)$$

where $H(x;\mu.\sigma.\varepsilon)$ is a generalized extreme value (GEV) distribution, because for a large enough threshold u, under the condition $X>u$, the distribution of $X-u$ is approximately GPD. It is given as

$$G\left(y,\bar{\sigma},\varepsilon\right)=1-(1+\varepsilon y/\bar{\sigma})^{-1/\varepsilon} \quad (5)$$

Proof: The theorem above is proved in paper [14].

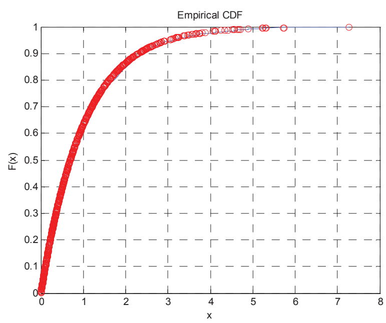
According to the central limit theorem[15], under the H_0 hypothesis, the sequence $P(k)$ is the noise spectrum, which is an IID exponential distribution. Then the limit distribution of the maximum value of $P(k)$ ($\Gamma=\max(P_1,P_2,...,P_n)$) denoted by $F_\Gamma(P)$ is a GEV distribution. By setting the threshold λ_0, the exceedance sequence is given by

$$G=\begin{cases}P(k)-\lambda_0 & ,P(k)>\lambda_0\\ 0 & ,else\end{cases} \quad (6)$$

$G=\{g_1,g_2,...,g_L\}$ has L greater values in total. The limit distribution of the sequence G follows GPD.

However, under the H_1 hypothesis, the new sequence $P(k)$ not only has peak spectrum lines, but also contains noise spectrum lines. The part of the peak spectrum follows noncentral chi-square distribution, while the part of the noise spectrum follows exponential distribution. Therefore, $P(k)$ is not IID random variables sequence, and it does not satisfy the condition of GEV distribution. It can be readily seen that the exceedance sequence $G=\{g_1,g_2,...,g_L\}$ does not follow GPD under the H_1 hypothesis.

Fig.1 shows that the distributions of the exceedance sequences obtained by BPSK and QPSK signals and their GPD under the same simulation condition. In Fig.1, the blue line and the red circle are the GPD and the distribution of the exceedance sequence, respectively. It can be clearly seen that under the H_0 hypothesis, the exceedance sequence approximately follows GPD, whereas under the H_1 hypothesis, it does not follow GPD. Hence, the BPSK/QPSK signal is recognized by taking advantage of the features as mentioned above.



(a) BPSK