OPEN

Wireless signal modulation identification method based on RF I/Q data distribution

Lihui Wang, Zhenjia Chen[✉] & Yonghui Zhang

Electromagnetic spectrum detection is the basis of the next generation wireless communication technology. Wireless signal identification is an important part of electromagnetic spectrum detection and management activities. This paper proposes to extract the distribution features of different modulated signals from the signal I/Q data. A two-dimensional gradient matrix is used to describe the characteristics of the signal classification. The minimum gradient cumulative distance (GCD) estimate between the sample and the model is used as the decision criterion for the signal classification. According to the result of the confusion matrix, the weight of the model is adjusted. Experiments show that the recognition rate of the modulated signal mentioned in this paper can reach 82.75%. The I/Q data sample was extracted under actual engineering conditions involving random noise, and the recognition rate dropped to approximately 79%. Based on the initial model gradient matrix, a reasonable algorithm is set to adjust the weight of the model, which can effectively improve the recognition rate of the modulated signal.

Electromagnetic spectrum resources are non-renewable resources. The use of spectrum resources by various countries is an important national development strategy. The electromagnetic spectrum detection system mainly realizes the perception and visualization of the spatial spectrum distribution. This is the basis for the safe and rational use of spectrum resources from the physical layer. In the spectrum detection system, most researchers pay attention to the spatial spectrum distribution prediction algorithm, spectrum occupancy calculation, and spectrum interference detection algorithm¹. The study of interference maps helps manage spectrum resource utilization, facilitates the rapid location of interference sources, and develops solutions^{2,3}. However, these spectrum detection methods only estimate the energy distribution of the spatial spectrum and do not estimate other signal characteristics. As the number of wireless devices increases, analyzing electromagnetic spectrum detection from the perspective of source feature extraction can more effectively improve the electromagnetic spectrum resource management and control capabilities. For example, the location, tracking and monitoring of unregistered signals. Describe the spectral resources based on the signal source parameters that can greatly reduce the capacity of the spectrum database⁴. Therefore, the research of signal detection methods directly affects the accuracy of spectrum detection results. Researchers propose an algorithm for signal identification problems specific to MIMO systems⁵. Among them, the researchers conducted in-depth research on space-time block code identification and MIMO modulation identification. The MIMO modulation identification methods are mainly divided into: the maximum likelihood function method of the received signal, and the specific modulation features of the received signal. The data processed by the system is obtained after the Fast Fourier transform (FFT) block of the I/Q data at the Radio Frequency (RF) end.

Blind detection is a necessary condition for the spectrum detection system. Matching detection of a specific modulation method increases the complexity of the detection system. With the digitization of wireless signals, we can achieve signal source identification through data analysis methods. RF I/Q data contains all the information of the wireless signal source. We extract the phase information at different receiving antennas from the I/Q data to achieve signal source direction finding⁶. The software defined radio (SDR) module collects I/Q data from the source. FFT is performed on the I/Q data to obtain the energy distribution of the signal at each frequency. Real-time phase information can also be extracted from the I/Q data. By extracting real-time phase information from the I/Q data, we implemented low signal to noise ratio (SNR) signal detection⁷. The data after FFT ignores many features. Therefore⁸, establishes a convolutional neural network (CNN) based on three features of I/Q, Amplitude/Phase and FFT to realize the recognition of signal modulation. But the authors use the I/Q data

School of Information and Communication Engineering, Hainan University, Haikou 570228, China. ✉email: zjchen@hainanu.edu.cn

Figure 1. Wireless signal transmission and reception. At the transmitting end, the software radio module transmits the modulated signal through the *Sink* to transmit I/Q data to the antenna end. At the receiving end, the software radio module obtains the I/Q data of the antenna end from the *Source*.

Figure 2. The distribution of I/Q data components of radio signals in environments with different SNR (one I/Q data sample). The higher the signal strength, most of the components of the I/Q data are concentrated in the positions of the four vertices of the coordinates. The weaker the signal strength, the I/Q component gradually approaches the origin.

before the RF end after the modulation module as shown in Fig. 1. Different modulation signals can be easily distinguished according to the I/Q distribution. However, it is not suitable for practical engineering applications.

The main contribution of this paper is to study the application of I/Q distribution characteristics in modulation signal recognition. The received I/Q data contain the noise of the environment, and its I/Q distribution will change accordingly. The I/Q distribution of signals with different modulation modes has different distribution features. Compared with Amplitude/Phase and the output FFT, I/Q data can better reflect the features of the signal. We extract the I/Q data of the wireless signal source directly from the source module at the receiving end. The I/Q data is the signal after the frequency spectrum is moved to the baseband. It contains the original information of environmental noise. Neural networks can be applied to classification, but learning time is too long, it is difficult to explain the reasoning process and reasoning basis. If the model adds a new category, it will need to be re-learned and will affect the recognition accuracy of the old category. The amount of I/Q data sample is large, and the time it takes to learn with a convolutional neural network is greatly increased. We propose to implement wireless signal modulation recognition based on I/Q component probability distribution.

I/Q component distribution of different modulation signal source

The SDR collects wireless signals in the form of I/Q data. Due to a large amount of I/Q data, the frequency distribution of the signal can be easily analyzed by frequency domain analysis after FFT. The time-frequency domain spectrum distribution characteristics can classify the signal modulation. However, the energy detection method usually has a high probability of missed detection (P_m) at low SNR. In fact, the I/Q data obtained by the RF receiver will be mixed with noise, and its I/Q distribution is distributed in different forms of “ring”. We propose to describe the signal features by extracting the I/Q distribution of different modulated signals. The I/Q data contains all the features of the baseband signal. I/Q data appears in complex form. It can be expressed as

$$S(n) = I + jQ, \quad I, Q \in [-1, 1]. \quad (1)$$

In this paper, signal samples with different energy intensities are collected. The I/Q component distribution of different SNR signals under the condition of no environmental interference is shown in Fig. 2. I represents the abscissa, Q is the ordinate, and the histogram represents the real part and the distribution component. When the signal strength is strong, most of the I/Q components are distributed near the four vertices of the interval, and there are fewer I/Q components near the origin, as shown in Fig. 2a. With the signal strength decreases, the I/Q signal gradually converges to the origin, the number of I/Q components decreases, and the real and imaginary parts are approximately normally distributed, as shown in Fig. 2b. However, the number of I/Q components of the noise sample is close to zero, as shown in Fig. 2c.

Figure 3. Weak signal detection by I/Q data weight extremes distribution method. The X-axis coordinate is expressed as E_W of an I/Q sample. The E_W of the noise is concentrated on a relatively high value.

The amount of I/Q for each sample of SDR acquisition is primarily determined by the ADC/DAC performance of the SDR. The sampling number of the software radio module used in this article is 8 bit, and the sampling rate is 20 Msps. The I/Q distribution is described by the real part and imaginary part of the complex number, and the Z-axis coordinates describe the weight of the I/Q component. In our detection system, the amount of I/Q for one samples is $n \in [1, 1, 131, 072]$. The component number of the sample is defined as cp . The weight of an I/Q component is defined as $C_{ij}^w = \text{Count}(C_i^{\text{real}} = I, C_j^{\text{imag}} = Q)$. In the case of low SNR, 10,000 weak signal samples and 10,000 noise samples are collected, and the $E_W = \max(C_{ij}^w)$ distribution of the samples is shown in Fig. 3. As the SNR decreases, the I/Q component approaches the origin, and the weight of the largest I/Q component can reach $C_{ij}^w \rightarrow 100,000$. When the boundary between noise and signal is set to $E_W = 80,000$, the detection probability of weak signals can reach more than 95%. As the SNR increases, the I/Q component is concentrated toward the four coordinate boundary points $(-1, -1), (-1, 1), (1, -1), (1, 1)$. Moreover, under the same SNR, the I/Q distribution in the complex coordinates of different modulated signals is also different. After the I/Q data is statistically calculated in the same real part and imaginary part, I/Q component distribution can be expressed as

$$\{C_i^{\text{real}}, C_j^{\text{imag}}, C_{ij}^w, j = 1, \dots, cp\}, \quad (2)$$

where $C_i^{\text{real}} = I$ and $C_j^{\text{imag}} = Q$. The positions of C_i^{real} and C_j^{imag} in the coordinates reflect the spatial distribution of the I/Q signal.

Modulation identification method base on I/Q distribution

With the same SNR, the I/Q component distribution of the wireless signal has a similar pattern. We propose a model that combines the gradients of the I/Q distribution to describe the different modulation signals. As shown in Fig. 4a, it is the distribution of I/Q samples of a complete FSK signal. The I/Q sample distribution of the modulated signal after propagation in free space will show the characteristic of “ring”. According to the gradient distance between the sample and the model, the model with the smallest distance is the target category. In order to reduce the amount of data, we take the data of $I, Q \in [0, 1]$. The data volume of the model is one quarter of one I/Q data sample, as shown in Fig. 4b. The real part I of each signal $S(n)$ is used as the X-axis coordinate, and the imaginary part Q is used as the Y-axis. The number of multiple $S(n)$ is the value of the Z-axis coordinate, and finally reflects the I/Q sample distribution of the modulated signal. The modulation signal identification method proposed in this paper is mainly related to the SNR of the signal. The lower the SNR of the signal, the more concentrated the distribution of the I/Q component is at the zero point. The I/Q samples used in this article are all the SNR of the modulated signal is 40 dB.

If the real part and imaginary part of I/Q are accurate to 0.001 ($I, Q \in [0.000, 0.999]$), the matrix of 1000×1000 represents the I/Q component distribution. The two-dimensional I/Q distribution matrix can be modified to

(a) Complete three-dimensional distribution of I/Q samples

(b) A quarter of the three-dimensional distribution of the I/Q sample

Figure 4. I/Q distribution of FSK signal (40 dB). I/Q components are mainly distributed in $I, Q \in [-0.4, +0.4]$.

$$C^{w 1000 \times 1000} = \begin{bmatrix} C_{0,0}^w & C_{0,1}^w & \cdots & C_{0,j}^w & \cdots & C_{0,999}^w \\ C_{1,0}^w & C_{1,1}^w & \cdots & C_{1,j}^w & \cdots & C_{1,999}^w \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{i,0}^w & C_{i,1}^w & \cdots & C_{i,j}^w & \cdots & C_{i,999}^w \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{999,0}^w & C_{999,1}^w & \cdots & C_{999,j}^w & \cdots & C_{999,999}^w \end{bmatrix} \quad (3)$$

where $C_{i,j}^w$ can describe the value of the I/Q component in different SNR environments. When it is a noise sample, then $C_{i,j}^w > 80,000$. When it is a strong signal, there is $C_{i,j}^w \rightarrow 1$ except for (1, 1). So $C_{i,j}^w \in [1, 80,000]$. $C^{w 1000 \times 1000}$ can represent I/Q data samples in different SNR environments.

We propose to describe different modulation signals by extracting a two-dimensional gradient matrix from the two-dimensional I/Q distribution matrix. The two-dimensional gradient matrix is obtained by calculating the gradient of adjacent $C_{i,j}^w$ in the two-dimensional I/Q distribution matrix. Two-dimensional gradient matrix as the basic model for adjusting signal recognition. It can be expressed as

$$G_{i,j}^w = \begin{cases} C_{i+1,j}^w - C_{i,j}^w \\ C_{i,j+1}^w - C_{i,j}^w \end{cases} \quad (4)$$

where the first equation is the column gradient, and the second equation is the row gradient. The two-dimensional gradient matrix model of this paper is calculated by column gradient. The two-dimensional gradient matrix formula of the column gradient is shown in (5).

$$G^{w 1000 \times 1000} = \begin{bmatrix} C_{1,0}^w - C_{0,0}^w & \cdots & C_{1,j}^w - C_{0,j}^w & \cdots & C_{1,999}^w - C_{0,999}^w \\ C_{2,0}^w - C_{1,0}^w & \cdots & C_{2,j}^w - C_{1,j}^w & \cdots & C_{2,999}^w - C_{1,999}^w \\ \vdots & & \vdots & & \vdots \\ C_{i+1,0}^w - C_{i,0}^w & \cdots & C_{i+1,j}^w - C_{i,j}^w & \cdots & C_{i+1,999}^w - C_{i,999}^w \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \quad (5)$$

For the same modulated signal, the two-dimensional I/Q distribution matrix between samples will have a small change, and the corresponding two-dimensional gradient matrix will also have a difference. In the case of the same SNR, we propose GCD to identify the modulated signal. It can be expressed as

$$\begin{aligned} D_{GCD} &= \sum |G_{label}^w - g^w| \\ &= \sum_{i=0, j=0}^{x_{max}-1, y_{max}-1} |G_{i,j}^w - g_{i,j}^w|. \end{aligned} \quad (6)$$

In practical applications, random noise has an influence on the recognition rate of the modulated signal. We propose model weight parameters (W) to adjust the gradient distance of each category and improve the overall recognition rate. (6) can be modified as

$$D'_{GCD} = \sum |W_{label} \times G_{label}^w - g^w|. \quad (7)$$

The value of W depends on the modulation signal identification error detection result. The similarity of the models of the individual categories is higher, and the GCD is lower than the actual value due to random noise, resulting in a error detection. The initialization model is a single sample gradient matrix, and the self-feedback model weight parameter adjustment can effectively improve the overall recognition rate.

Results and analysis

The signal acquisition module is HackRF. Its ADC/DAC sampling rate can reach 8 bits, that is, the range of I/Q component distribution is 1–256. A single transmitting antenna transmits signals, and a single receiving antenna collects I/Q data at a distance of 20 cm. The transmit power of the signal source is directly measured as -70.07 dBm with the feeder. The SNR of samples is about 40 dB. Based on the open source energy detection project (*soapy_power*), we modify the script to calculate the weight of the I/Q sample data to obtain a two-dimensional I/Q distribution matrix. The size of a single initial two-dimensional gradient matrix model is only 8.0 MB⁹. It can be deployed directly to the spectrum detection system. Confusion matrix for the wireless signal modulation identification as shown in Fig. 5. There are a total of eight types of modulated signals in this paper¹⁰, which are collected and saved as I/Q data formats by the SDR module. There are 100 test samples for each type of signal (800 samples in total). The recognition rate of the modulated signal can reach 76.13%.

In current newer research, the modulated signal I/Q data sample after modulation module and before RF module is used⁸. The sample in this article is the I/Q data sample from the RF module. The signal transmission

Figure 5. Confusion matrix for the wireless signal modulation identification (Before adjusting parameters). In the confusion matrix, each column represents the predicted category, and each row represents the true belonging category of the data.

Figure 6. Confusion matrix for the wireless signal modulation identification (after adjusting parameters).

and reception module introduces random noise. Even so, the recognition rate of the GCD method can also approach the recognition rate of the⁸ method. I/Q distributions are all distributed in a “ring” shape. OFDM, BPSK and 16QAM have similar I/Q distributions. Since most samples of OFDM and BPSK were misidentified as 16QAM, the recognition rate of OFDM and BPSK is low. According to the confusion matrix, it can be found that the probability of being misjudged as 16QAM and 64QAM is high, and the overall recognition rate can be improved by increasing their GCD coefficients. In order to improve the recognition rate, the model weight is adjusted. Suppose the initial model weight vector is $[1, 1, 1, 1, 1, 1, 1, 1]$, and the model weight step value is 0.01. As shown in Fig. 5, the samples of OFDM and BPSK are mostly misclassified as 16QAM and 64QAM. Therefore, it is preferred to increase the model weights of 16QAM and 64QAM step by step to increase the relative threshold of GCD and the recognition rate of OFDM and BPSK. In the adjustment process, the model weight of a single category model is increased to ensure that the overall recognition improves until the overall recognition rate converges. The model weight vector after recognition rate convergence is $D'_{GCD} = [1, 1, 1.05, 1, 1, 1, 1.05, 1.05]$. The recognition rate has increased to 82.75%. Confusion matrix for the wireless signal modulation identification after adjusting parameters as shown in Fig. 6. After adjusting the model parameters, the recognition rate of 16QAM is reduced, and the recognition rates of OFDM and BPSK are improved.

We mainly adjust the weight manually, so that the weight of the modulation signal with higher similarity is reduced, and the weight of the adjustment signal with less obvious characteristics is increased. The I/Q data sample used in this paper is measured data and environmental noise has been introduced. In different SNR environments, the I/Q distribution of the modulation signal does not change much, but the weights need to be adjusted slightly. Finally, we conducted engineering tests. Each sample is randomly selected to calculate the recognition rate. After a large sample size experiment, the recognition rate of actual engineering tests converges to 79%. As shown in Fig. 7. In the future, we will study the dynamic model weight adjustment method based on error detection results to improve the recognition rate.

Figure 7. Classification accuracy in a large sample size experiment. The modulation signal of each experiment is randomly selected.

Conclusion

We propose a modulation signal identification based on the signal I/Q distribution and the two-dimensional gradient matrix by estimating the gradient cumulative distance. The I/Q distribution matrix is extracted with a single I/Q data sample and the gradient matrix is calculated. The gradient distance of the sample describes the degree of similarity of the model. Then, the recognition rate of the modulated signal is increased by the model weight. Under the influence of random noise, the measured results show that the proposed method can guarantee a higher recognition rate.

Received: 18 June 2021; Accepted: 14 October 2021

Published online: 01 November 2021

References

- Höyhty, M. *et al.* Spectrum occupancy measurements: A survey and use of interference maps. *IEEE Commun. Surv. Tutor.* **18**(4), 2386–2414 (2016).
- Naranjo, J. D., Ravanshid, A., Viering, I., Halffman, R., & Bauch G. Interference map estimation using spatial interpolation of MDT reports in cognitive radio networks. In *IEEE Wireless Communications and Networking Conference (WCNC)* 1496–1501 (2016).
- Grimaldi, S., Mahmood, A., Hassan, S. A. *et al.* Autonomous Interference Mapping for Industrial Internet of Things Networks Over Unlicensed Bands: Identifying Cross-Technology Interference. *IEEE Ind. Electron. Mag.* **15**(1), 67–78 (2020).
- Chen, Z. J. & Zhang, Y. H. The application of distributed database on spectrum big data. In *Cloud Computing and Security—4th International Conference, ICCS 2018*, vol. 11064, 212–222 (2018).
- Eldemerdash, Y. A., Dobre, O. A. & öner, Mengüç. Signal identification for multiple-antenna wireless systems: Achievements and challenges. *IEEE Commun. Surv. Tutor.* **18**(3), 1524–1551 (2016).
- Chen, Z. J. & Zhang, Y. H. Monostatic multi-source direction finding based on IQ radio frequency data. *AEU Int. J. Electron. Commun.* **97**, 137–148 (2018).
- Chen, Z. J. & Zhang, Y. H. Providing spectrum information service using TV white space via distributed detection system. *IEEE Trans. Veh. Technol.* **68**(8), 7655–7667 (2019).
- Kulin, M., Kazaz, T., Moerman, I. & Poorter, E. D. End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications. *IEEE Access* **6**, 18484–18501 (2018).
- Gradient Matrix of Signal Models [Online]. <https://github.com/zjchen/Signal-Models> (2019).
- I/Q Samples and Signal Identification [Online]. <https://github.com/zjchen/Signal-Identification> (2019).

Acknowledgements

This work was supported in part by the Natural Science Foundation Innovation Research Team Project of Hainan Province 620CXTD435, by the National Natural Science Foundation of China under Grant 61961012.

Author contributions

We propose a modulation signal identification based on the signal I/Q distribution and the two-dimensional gradient matrix by estimating the gradient cumulative distance. The I/Q distribution matrix is extracted with single I/Q data sample and the gradient matrix is calculated. The minimum gradient cumulative distance (GCD) estimate between the sample and the model is used as the decision criterion for the signal classification. L.W. and Z.C. wrote the main manuscript text. Y.Z. participated in the experiment content. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021