A Novel Classification Method Based on Adaboost for Electromagnetic Emission

Jing Nie, Shunchuan Yang, Qiang Ren, and Donglin Su

Institute of Electromagnetic Compatibility Technology
Beihang University, Beijing, 100191, China
niejing0209@163.com, scyang@buaa.edu.cn, qiangren@buaa.edu.cn, sdl@buaa.edu.cn

Abstract ─ Abundant characteristics information of equipment or systems could be obtained from electromagnetic emission data. In this paper, those typical characteristics, like harmonics, damped oscillations, of electromagnetic emission are classified via the adaptive boosting (Adaboost) algorithm and they are validated through measurement results. Based on the "basic emission waveform theory", three types of the basic fundamental elements, characteristics-harmonic, narrowband and envelope-of complex emission in frequency domain, are considered in our proposed method. By taking weights combination patterns to effectively improve the classification performance of a single classifier, quite high classification accuracy could be achieved by Adaboost algorithm in our simulations. In our study, 100% precision classification accuracy of three types of characteristics could be obtained using Adaboost with 13 decision tree weak-classifiers. Compared with other classification methods, the Adaboost algorithm with decision tree weak-classifier used to classify typical characteristics of electromagnetic emission is the most accurate. At the same time, it is very effective to process the measured data. Only through the classification of multiple emission signals can identification and positioning of electromagnetic interference sources further.

Index Terms ─ Adaboost, classification, electromagnetic emission characteristics, classification probability, signal component.

I. INTRODUCTION

The electromagnetic emission is the key parameter to evaluate the performance of equipment and systems. Once the emission characteristics of the system is obtained through numerical simulations or meansurements, useful information to indicate its performance could be extracted [1]. In [2], the "basic emission waveform theory" was proposed to identify the basic emission sources of complex systems [2]. Various emissions are characterized with and decomposed into four basic elements including square wave, sine wave, damped oscillation, and spike wave according to physical sources, which are interpreted as the emission waveforms of different types of circuit elements. Then, based on such decomposition, the performance could be evaluated. Therefore, the emission characteristics of any equipment can be obtained by analyzing its electromagnetic emission. However, to the best of the authors' knowledges, how to identify those fundamental elements are not yet addressed. We have reported our premiliary results in [2]. In this paper, comprehensive studies are carried out and compared with measurements.

The location of the interference source has always been a common concern of researchers. Therefore, when we extract the characteristics of the collected electromagnetic emission signals, can we determine which emission source does the characteristics correspond to? The classification model in this paper is to identify and classify the characteristics so as to further correspond to the emission source. Various classfication approaches have been extensively investigated [3], [4], [5], [6]. In recent years, the machine learning (ML) based methods have been paid special attention to, since MLs could extract hidden pattern and characteristics embedding in a large number of data [4]. Therefore, they are widely used in data mining, image processing, medical diagnosis and etc. [7]. Many machine learning methods are used in signal processing and cybernetics [8], [9]. However, to the authors' knowledge, they are seldom used to analyze the emission and their characteristics in electromagnetic compatibility field. In this paper, we explore the possibility of using the ML method to analyze the electromagnetic emission characteristics through the machine learning algorithm, more specifically, with Adaboost.

The paper is organized as follows. In Section II, four basic waveforms proposed the "basis emission waveform theory" are summarized and the new classification method is illustrated in detail. In Section III, numerical cases are carried out to validate the proposed method. In Section IV, its effectinveness is further confirmed by the good agreement with experiment data. At last, we draw

Submitted On: October 14, 2018
Accepted On: January 26, 2019

some conclusions in Section V.

II. BASIC ELECTROMAGNETIC EMISSION CHARACTERISTICS AND CLASSIFICATION

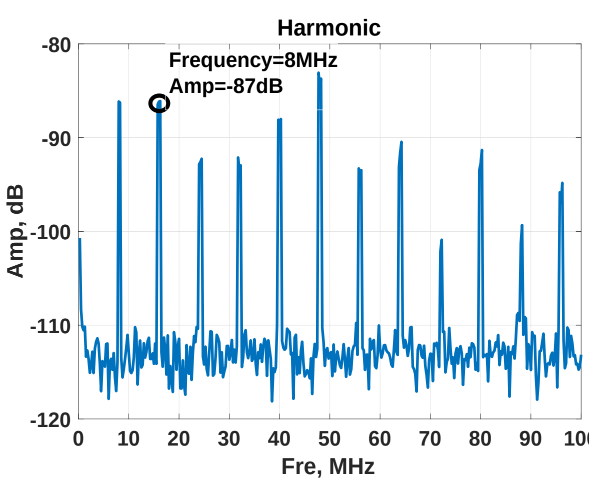
A. Electromagnetic emission characteristics from basic emission waveform theory

As indicated by “basic emission waveform theory”, no matter how complex the electromagnetic emission is, the number and type of basic emission sources are limited, which are generally in some limited forms. By categorizing numerious circuits and their electromagnetic emission, the basic waveforms are cast into four types in the time domain: square waves, sine waves, damped oscillations, and spike waves [2,10].

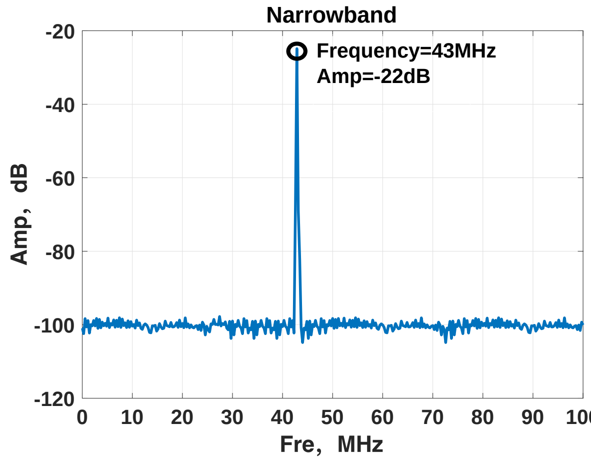
Based on the above-mentioned emission sources, the relationship between the basic emission waveforms in time domain and the typical circuit characteristics in frequency domain are further clarified in this paper. First of all, square wave (clock signal) is the principal function signals in the digital circuit, which always possesses the characteristic of harmonics or an overall lift in the frequency domain. A typical waveform from the measured data of a crystal oscillator module is shown in Fig. 1 (a).

An analog circuit module is generally used for amplification, mixing, detection, and other related purpose. The overall data should have sine signals, which often own the characteristic of narrowband or even single frequency in the frequency domain, as shown in Fig. 1 (b), which is generated by the sine signal with random noise from the signal generator.

In addition, impedance mismatch leads to damped oscillations in the circuit. The damped oscillation signals present the characteristic of envelope in frequency domain, as shown in Fig. 2, which is generated by the damped oscillation signal with random noise from signal generator.



(a)



(b)

Fig. 1. (a) Harmonics characteristic, and (b) narrowband characteristic in frequency domain.

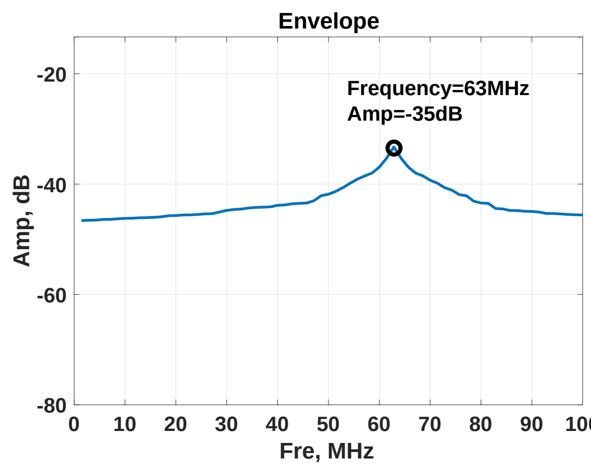


Fig. 2. Envelope characteristic in the frequency domain.

B. Feature recognition and classification based on Adaboost algorithm

The Adaboost algorithm is an iterative procedure that combines many weak-classifiers to the powerful Bayesian classifier $C(x)$. Starting with the unweighted training sample data, the Adaboost constructs a weak-classifier to produce class labels. If a training data point is misclassified, the weight of that is named as training [11]. In addition, original Adaboost algorithm was further extended to the multi-class case in [12]. In this paper, we select this method to analyze the emission characteristics. The pseudo code of the algorithm is listed as follows.

1) Initialize the observation weights:

$$\omega_i = 1/n, i = 1,2,3 \dots, n. \tag{1}$$

2) For $m = 1$ to M:
   a) Fit a classifier $T^{(m)}(x)$ to the training data using weights ω_i.
   b) Compute:

$$err^{(m)} = \sum_{i=1}^{n} \omega_i \cdot II\left(c_i \neq T^{(m)}(x_i)\right) / \sum_{i=1}^{n} \omega_i. \quad (2)$$

c) Compute:

$$\alpha^{(m)} = log\frac{1-err^{(m)}}{err^{(m)}} + log(K-1). \quad (3)$$

d) Set,

$$\omega_i \leftarrow \omega_i \cdot exp\left(\alpha^{(m)} \cdot II\left(c_i \neq T^{(m)}(x_i)\right)\right), \quad i = 1,2,\dots,n. \quad (4)$$

e) Re-normalize ω_i.

3) Output:

$$C(x) = argmax_k \sum_{m=1}^{M} \alpha^{(m)} \cdot II\left(T^{(m)}(x_i) = k\right). \quad (5)$$

Different weak-classifier often generate different classification effects. In our research, decision tree and decision stump are used as weak-classifiers to study a small group of samples and nonlinear data of electromagnetic emission.

III. THE VERIFICATION OF ELECTROMAGNETIC EMISSION CHARACTERISTICS CLASSIFICATION

In this part, 300 groups of typical frequency characteristics data and their labels, which are the ideal waveforms from signal generator, have been used as the training data set, respectively. Harmonics, narrowband and envelope are considered and each has 100 groups.

Corresponding to Part II B, Equ. (1) n = 300, initial sample data weights are $1/300$, m is the number of weak classifiers, K = 3 is data types. Equation (3) ensures the weight update direction is greater than 0. For the m-th classifier being trained, according to Equ. (4), the weight of misclassified data is increased, while the weight of correctly classified data is reduced. The redistributed weight data is used to train the next classifier. Finally, the combination training result of weak classifiers with the highest probability of correct classification is obtained from Equ. (5). While, 110 groups of data in three types and their corresponding labels, which are similar to real emission, are chosen as testing data.

The classification accuracy:

$$A(h,y) = \frac{1}{n}\sum_{i=1}^{n} 1\{h(x^{(i)}) = y^{(i)}\}. \quad (6)$$

For training data (x,y), $y = 1,2,3,\dots d$, where x is the electromagnetic emission characteristics data, y is the type of such data, namely label, d is the number of labels for classification and $h(x)$ is the classification label generated by training and prediction of Adaboost, n is the total number of sample x. So the training error is $E(h,y) = 1 - A(h,y)$.

The decision tree classifier assigns the parameter $\gamma_j, j = 1,2,\dots,J$. to several mutually disjoint regions (attributes J), which can be expressed as:

$$F(x,\Theta) = \sum_{j=1}^{J} \gamma_j I(x \in R_j), \quad (7)$$

and

$$\Theta = argmin \sum_{j=1}^{J} \sum_{x \in R_j} E(y_j, \gamma_j). \quad (8)$$

Θ in Equ. (7) and Equ. (8) represents the parameter that minimizes the empirical risk, namely, the one that minimizes the training error of one classifier. And the objective function is:

$$F(x) = arg\min_{1 \le y \le d} E(h,y) = \frac{1}{n}\sum_{i=1}^{n} 1\{h(x^{(i)}) \neq y^{(i)}\}. \quad (9)$$

It is worth mentioning that different weak classifiers have different results. Decision stump weak classifier makes a single decision on characteristics data, that is, only one split is needed for judgment. Decision tree weak classifier judges characteristics from multiple angles, such as amplitude and frequency interval, and needs to be split into multiple disjoint spaces for judgment. Based on the nonlinearity and amplitude-frequency characteristics of electromagnetic emission data, the two kind weak-classifers are chosen to analyze. When using different numbers of decision tree weak-classifiers, the accuracy is increasing gradually and stabilizing at 100%, which implies that the classification results match the actual classification of the testing data completely. The optimal number of decision tree weak-classifiers in this algorithm is 13, as shown in Fig. 3 (a).

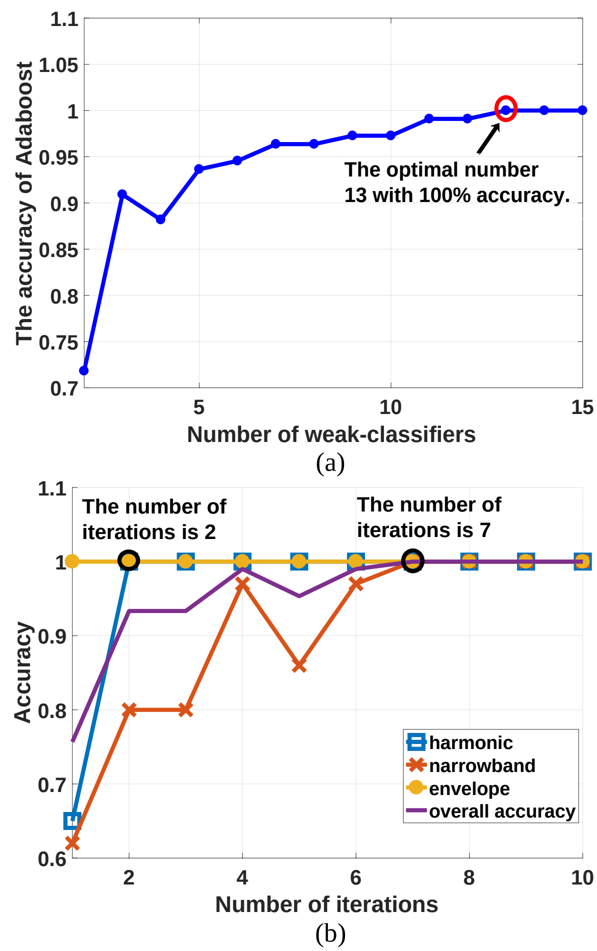


Fig. 3. (a) The accuracy with the number of weak-classifiers, and (b) the accuracy of three varieties characteristics with the number of iterations.

When using the same number of the decision tree weak-classifiers, the accuracy changes with the iteration number. It’s clear from Fig. 3 (b) that the envelope characteristic shows the fastest identification, then followed by the harmonic and narrowband. The Table 1 shows the accuracy of different algorithms, which proves Adaboost-DecisionTree to be optimal.

Table 1: Accuracy of different algorithms

Alogritm	Single Classifier	13 Classifiers
SVM	0.8369	--
Adabost-Decision Stump	0.5636	0.6273
Adaboost-DecisionTree	0.8000	1.0000

IV. VERIFICATION

In this section, the author conducted radiation emission tests on three types of electronic equipments in a 10-meter semi- anechoic chamber, which are intercom, computer host and shielding device, then applied the Adaboost algorithm to classify signals under different conditions, including single type signal classification and mixed types signal classification.

(a)

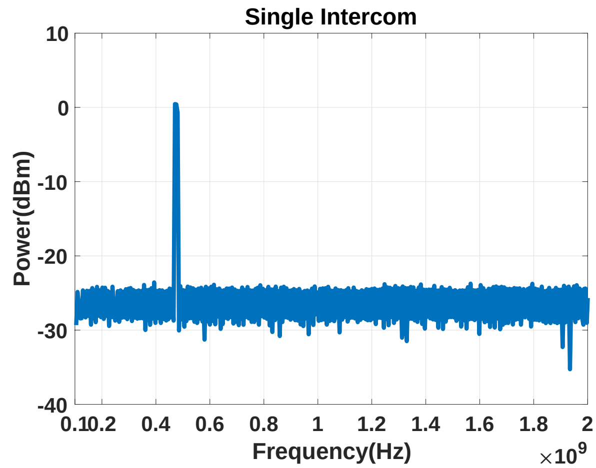
(b)

(c)

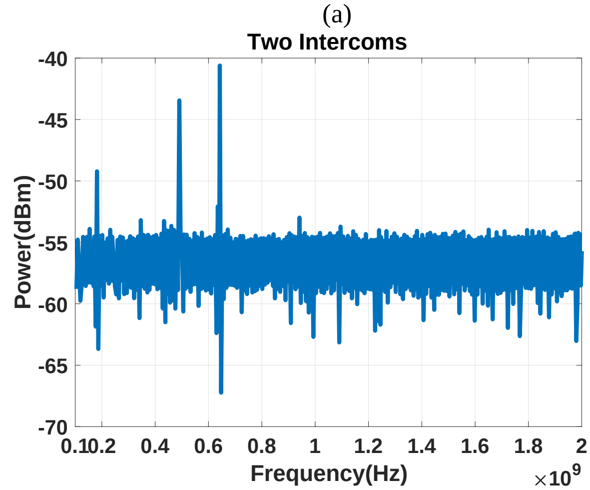
Fig. 4. (a) The testing equipment of Intercom, (b) the testing equipment of Computer Host, and (c) the testing equipment of Shielding Device.

A. Single Type Signal Classification

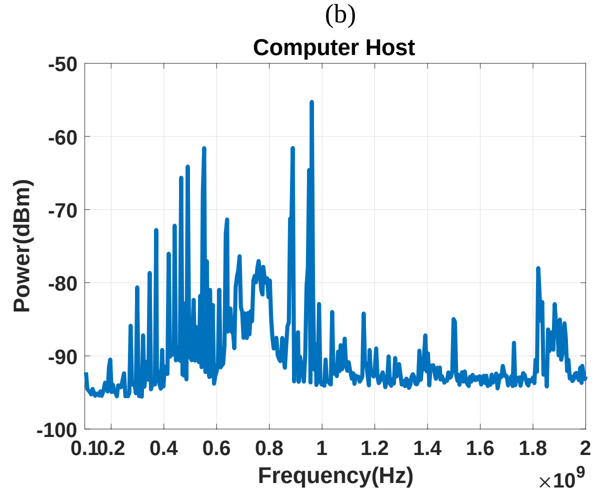
The testing equipments in this research are shown in Fig. 4. Among them, the operation frequency of the intercom is 400MHz-470MHz. The shielding device mainly shields the signal of mobile phones, with the main frequency being 1GHz, 2GHz, 3GHz and 4GHz. The radiation emission of single type is shown in Fig. 5. Among them Figs. 5(a) and (b) are the emission data of a single intercom and two intercoms respectively. In the classification, both kinds data are used as training data of intercom.



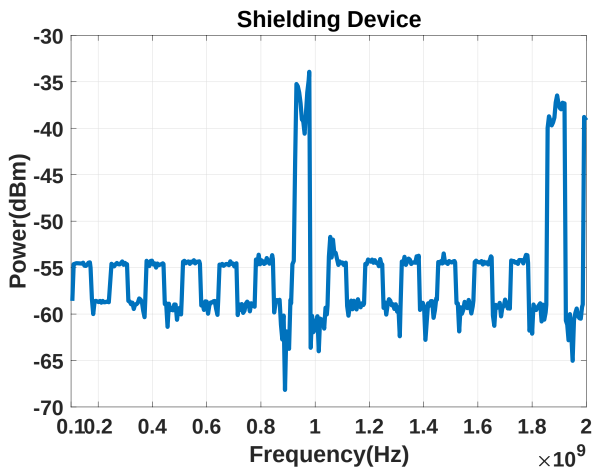
(a)



(b)



(c)



(d)

Fig. 5. (a), (b) The emission of a sigle intercom and two intercoms, (c) the emission of a sigle computer host, and (d) the emission of a sigle shielding device.

The frequency band of the emission is 100MHz-2GHz, and the data number of each type equipment is 90 groups. Therefore, the training data of three types for Adaboost algorithm is 270 groups, containing the emission data and their labels. Based on the training data, there are another 120 groups testing data. It's worth noting that whether the label of the testing data is known determine the classification accuracy of the algorithm. That means when use the label of testing data to predict the classification, the decision tree classifier only needs 1 to classify the 120 groups testing data completely, while the decision stump needs 3. However, when the label of testing data is unknown or we don't use the label while predicting, neither classifier could properly classify the 120 groups testing data. The classification accuracy on the testing data is shown in Table 2.

Table 2: Accuracy of different weak-classifiers

Adabost with Weak-Classifiers	**The Label is Known (with only 1 classifier)**	**The Label is Unknown (with 3 Classifiers)**
Decision Stump	0.68333	0.91675
Decision Tree	1	0.90005

Based on the analysis above, since the specific classification label of emission cannot be obtained usually in the actual complex electromagnetic environment, this paper prefers to use 3 decision stump weak-classifiers to predict the classification of the radiation emission data from the actual equipments without classification labels.

B. Mixed types signal classification

The three types of signals are combined in different forms, including pairwise combination and three types of mixing. The mixed types of testing data are 2,500 groups. At the same time, the training data and the corresponding label used in Adaboost algorithm are still the 90 groups data in part IV.A. The testing site is shown in Fig. 6.

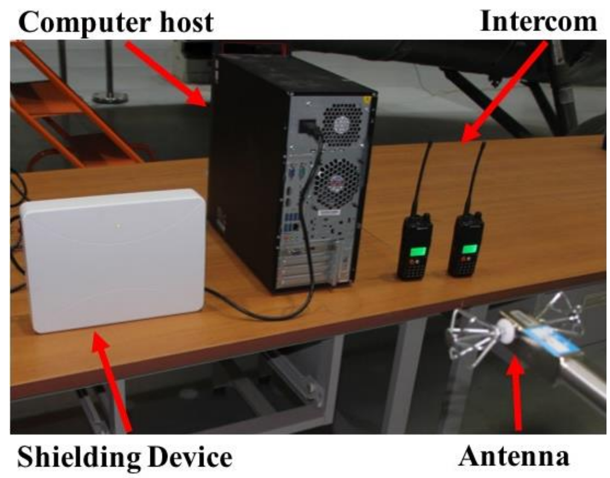


Fig. 6. The testing site.

1) Two Types Signal Mixed

In this part, each two types of radiation emission signals are mixed in proportion. At each proportion, there are 2,500 groups testing data of mixed types signals contain 50 groups for each type, so the classification probability is a statistical result. The quantitative relation curve of classification probability with the proportion of one type signal in the mixed environment is obtained by combining the three types signals from the equipments, i.e., the intercom, the computer host and the shielding device.

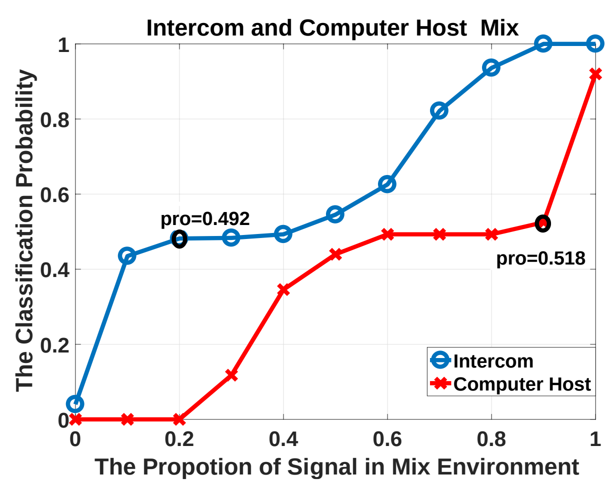


Fig. 7. The quantitative relation curve of classification probability with the proportion of intercom and computer host mixed environment.

By comparing above, it's shown that the Adaboost algorithm has different classification effect on the same type of signal in different combination situations. To be specific, Fig. 7 shows that when the intercome accounts for 20% in the mixed electromagnetic environment and the computer host accounts for 80%, the probability of the algorithm can classify intercom is 0.492. When computer host accounts for 90%, and intercom accounts for 10%, the probability classifying for computer host is only 0.518. It can be seen that the characteristics of the intercom are easier to classify when the two types signals are mixed.

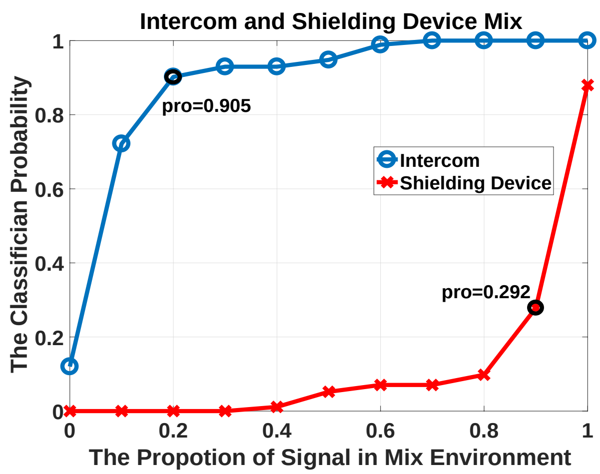


Fig. 8. The quantitative relation curve of classification probability with the proportion of intercom and shielding device mixed environment.

Figure 8 shows the combination of intercom and shielding device signals. When the intercome accounts for 20% in the mixed environment and the computer host accounts for 80%, the probability classifying for intercom is 0.905. When shielding device accounts for 90%, and intercom accounts for 10%, the classification probability of shielding device is 0.292. It can be seen that the characteristics of the intercom are easier to classify than shielding device.

Figure 9 shows the combination of computer host and shielding device signals. When the computer host accounts for 20% in the mixed environment and the shielding device accounts for 80%, the probability of computer host is 0.001. While, the probability of shielding device is 0.840, which shown that the characteristics of the computer host are easier to classify than shielding device.

2) Three Types Signal Mixed

When the three types of signals are mixed, the algorithm classification probability is shown in Fig. 10.

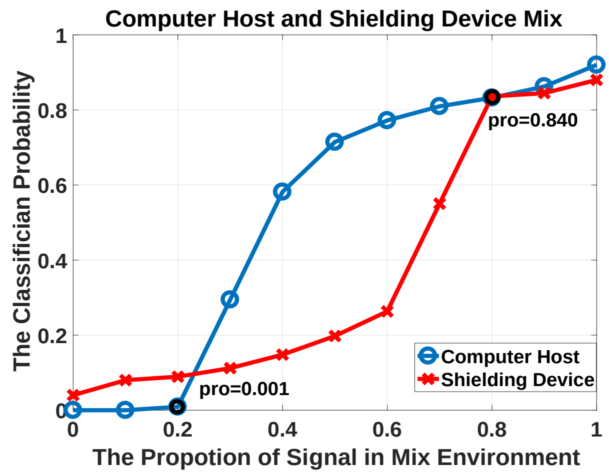
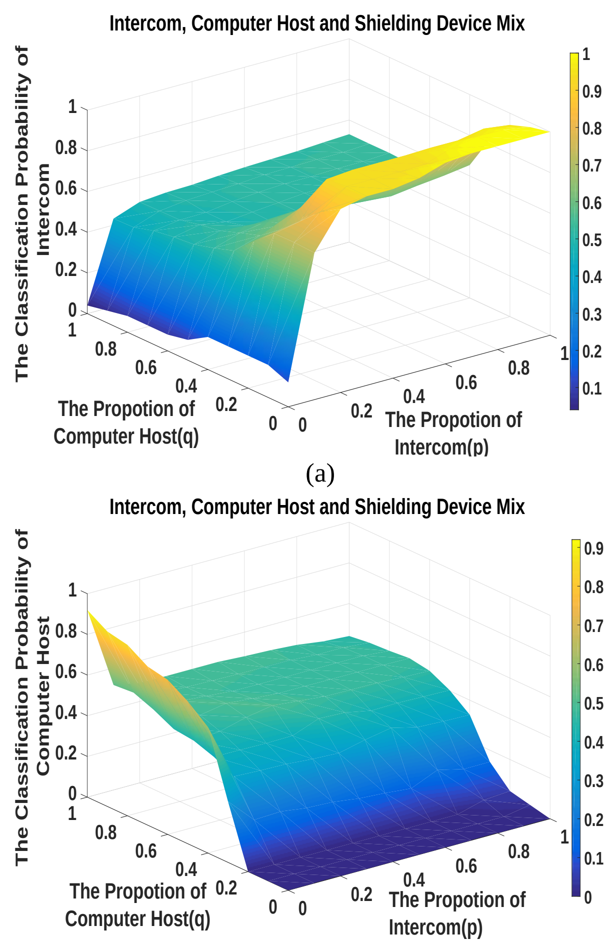
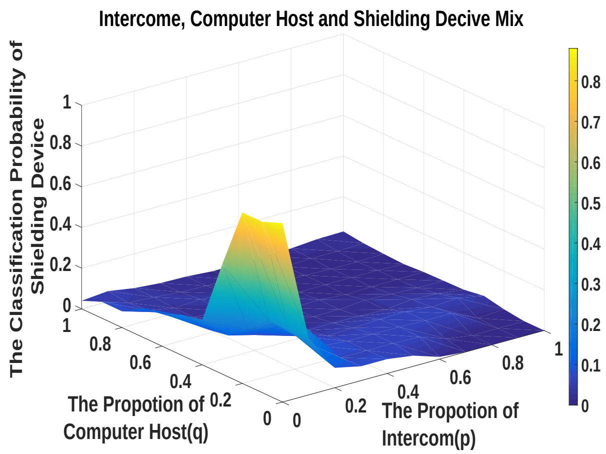


Fig. 9. The quantitative relation curve of classification probability with the proportion of computer host and shielding device mixed environment.



(a)

(b)



(c)

Fig. 10. The quantitative relation curve of classification probability with the proportion of three types mixed electromagnetic environment.

In our study, p is the proportion of intercom signals, q is the proportion of computer host signals, so the proportion of shielding devices is (1-p-q). Figure 10 (a) shows the statistical classification probability of intercom with the change of environment components. It can be seen that when the intercom accounts for about 20%, no matter how the components of computer host and shielding device change, the classification probability for intercom is always no less than 0.5. Figure 10 (b) shows the classification probability of computer host. When the computer host accounts for about 60%, no matter how the other two types signals change, the probability can be kept above 0.45. Figure 10 (c) is the classification probability of the shielding device. Only when the intercom signal accounts for less than 30% and the computer host signal is less than 60% could the probability of the shielding device be around 0.4.

V. CONCLUSION

Based on the "basic emission waveform theory", harmonic, narrowband and envelope are extracted as electromagnetic emission characteristics of different electromagnetic interference sources in the frequency domain. With carefully definition in Part III, the accuracy of Adaboost algorithm reaches 100% with 13 decision tree weak-classifiers. Our experiments shows that the number of weak-classifiers is fixed, the envelope characteristic is the first one to be classified, then followed by the harmonic and narrowband. In validation part, the electronic equipment intercom, computer host and shielding device are chosen to develop radiation emission test and emission data analysis. According to the basis analysis results of the single type signal classification, Adaboost algorithm with decision stump weak-classifier is selected to study on the classification in types mixed electromagnetic environment, the quantitative results of classification probability are obtained under different types signals components environment. It can be seen that the intercom is the closest to the narrowband signal of the basic emission characteristics and is also the most easily classified of the three types of signals.

That is to say, in the environment where multiple electronic devices work together, identification and classification of different devices and their components could be realized through our method. Only by accurate classification of electromagnetic emission and their characteristics can identification and positioning of electromagnetic interference sources be further realized.

ACKNOWLEDGMENT

This work was supported by the project 61427803 from the National Natural Science Foundation of China (NSFC).

REFERENCES

[1] D. L. Su, S. G. Xie, and F. Dai, *The Theory and Methods of Quantification Design on System-level Electromagnetic Compatibility*. National Defence Industry Press, pp. 53-58, 2015.

[2] D. L. Su, S. G. Xie, and A. X. Chen, "Basic emission waveform theory: A novel interpretation and source identification method for electro-magnetic emission of complex systems," *IEEE Transactions on Electromagnetic Compability*, vol. 99, pp. 1-10, 2017.

[3] D. Shi and Y. Gao, "Electromagnetic radiation source identification based on spatial character-istics by using support vector machines," *Appl. Comput. Electromagn. Soc. J.*, vol. 32, no. 2, pp. 120-127, 2017.

[4] J. Y. Wu, S Cui, and D. L. Su, "A support vector machine method based on spectrum data feature analysis," *Presented at Asia-Pacific Electromagnetic Week Conference on Electromagnetic Compatibility*, Xi'an, China, 2017.

[5] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5-32, 2001.

[6] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *J. Jpn. Soc. Artif. Intell.*, vol. 14, no. 5, pp. 771-780, 1999.

[7] F. M. Takbir Hossain, Md. I. Hossain, and S. Nawshin, "Machine learning based class level prediction of restaurant reviews," *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, pp.420-423, 2017.

[8] S. Q. Wu and H. Nagahashi, "Parameterized AdaBoost: Introducing a parameter to speed up the training of real AdaBoost," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 687-691, 2014.

[9] P. B. Zhang and Z. X. Yang, "A novel AdaBoost

framework with robust threshold and structural optimization," *IEEE Transactions on Cyberrnetics*, vol. 48, no. 1, pp. 64-76, 2018.

[10] C. Xiao and T. Zhao, "Identification method of EMI sources based on measured single-channel signal and its application in aviation secondary power source design," *IEEE Trans. Electromagn. Compat.*, vol. 59, no. 2, pp. 439-446, 2017.

[11] Y. X. Wang. "Research of Prediction Method of Sudden Cardiac Arrest Based on ECG Signals," Jilin University, June 2017.

[12] J. Zhu, S. Rosset, H. Zou, and T. Hastie, "Multi-class Adaboost," *Statistics and Its Interface*, vol. 2, no. 3, pp. 349-360, 2009.